# British Society for the Philosophy of Science

# Annual Conference Abstracts

University of Exeter

6-8 July 2022





# Contents

Plenary Talks4
Cooper, Rachel4
Morgan, Mary5
Individual Papers6
Ackermann, Matthias6
Antonio, José and Escobar, Pérez7
Aybar, Raphael9
Bakshi, Kabir11
Bielinska, Marta13
Bräutigam, Maren15
Britten-Neish, George17
Brown, Rachael19
Calderón, Francisco
Coraci, Davide and Cevolani, Gustavo22
Curtis-Trudel, Andre24
Devanesan, Arjun25
Dewar, Neil26
Emmerson, Nicholas28
Erasmus, Adrian30
Franklin, Alexander32
Furman, Katherine34
Gebharter, Alexander and Osimani, Barbara36
Geddes, Alexander
Guo, Bixin
Hilligardt, Hannah41
Jaksland, Rasmus and Salimkhani, Kian43
Jäntgen, Ina45
Lakhany, Farhan47
Landström, Karl49
Lari, Teemu51
Letrud, Kåre and Johnsen, Svein Åge Kjøs53
Lopez, Luis G55

	Makins, Nicholas	57
	Malik, Uzma	59
	Matarese, Vera	61
	Maziarz, Mariusz and Serpico, Davide	63
	McCoy, Ryan	65
	Meyer, Russell	67
	Michelini, Matteo	69
	Muller, Sean	71
	Ney, Milan	73
	Nissan-Rozen, Ittay and Liron, Amir	75
	Pain, Ross	77
	Piotrowska, Monika	79
	Sassoon, Judyth	81
	Segarra, Adrià	83
	Sequeiros, Sofia Blanco	85
	Skaf, Rawad El and Palacios, Patricia	87
	Spagnesi, Lorenzo	88
	Steinbrink, Lukas Christian	90
	Suárez, Javier and Stencel, Adrian	92
	Sung, Richard D. and Holman, Bennett	94
	Sus, Adán	96
	Taşdan, Ufuk, Roberts, Bryan W., and Thebault, Karim P. Y	98
	Thompson, Morgan	100
	Tulodziecki, Dana, Porterfield, David Marshall, and Kruse, Colin P.S.	102
	Wells, Aaron	104
	Whiteley, Cecily	106
	Żuradzki, Tomasz	108
S	ymposia	110
	Comparative Cognitive Archaeology	110
	Feminist Philosophy of Biology Beyond Gender	112
	Exploring the senses of exploration in experimental neuroscience	114
	The Rise of Basal Cognition and its Relevance for the Mind and Brain Sciences	118
	On the use of racial categories in medicine across geographic and national contexts	120
	Understanding Causal Inference	123
	Aetiology of the Replication Crisis	125

Quantum Probability Beyond Quantum Measurement	.128
Reliability in science: how is it established, and how is it lost?	.134
Assisted Discoveries, Understanding and the Aims of Science	.137
New Mechanism and Alternative Causal Concepts in Biology: Processes and Pathways	.140
Function in the light of frequency-dependent selection	142

# **Plenary Talks**

## Cooper, Rachel

Disorder in the Philosophy of Medicine

A major research project in the philosophy of medicine has sought to use conceptual analysis to consider what makes a condition count as a disorder, or as pathological (as opposed to being a benign variation, or a moral failing, or a social or educational problem). It's now commonly accepted that this research programme has run into difficulties. Problems have emerged for a number of reasons. One issue is that our concept of disorder changes over time (partly because philosophers and others have analysed, and shaped, our thinking in this area). This causes problems for projects of descriptive conceptual analysis. Conceptual change means that a criterion that was necessary for a condition to be a disorder at one time may cease to be necessary a relatively short time later. I consider how best we might overcome the methodological problems that emerge when philosophers seek to do conceptual work on 'disorder' and other such disorderly concepts.

### Morgan, Mary

### Narrative in Science

Narrative is ubiquitous inside the sciences. While it might be hidden, evident only from its traces, it can be found regularly in scientists' accounts both of their research, and of the natural, human and social worlds they study. Investigating the functions of narrative, it becomes clear that narrative-making provides scientists a means of making sense of the phenomena in their field, that narrative provides a means of representing that knowledge, and that narrative may even provide the site for scientific reasoning. Narrative emerges as a 'general purpose technology', used in many different forms in different sites of science, enabling scientists to figure out and express their scientific knowledge claims. Understanding scientists' use of narrative as a sense-making technology suggests that narrative functions as a bridge between the interventionist practices of science and the knowledge gained from those practices.

# **Individual Papers**

## Ackermann, Matthias

# Towards causal counterfactual storylines---or: should you be a Bayesian about climate information?

Climate change made North America's deadly heatwave 150 times more likely" (Schiermeier, 2021). Statements like these stem from a field of climate research that investigates relations between climate drivers and observations of the earth system's responses, known as detection & attribution. Broadly speaking, the field stands in the tradition of what has become known as the 'probability-based approach' (IPCC AR6, 2021), an approach that effectively reduces uncertainty in thermodynamic aspects of climate, while mostly disregarding uncertainty regarding atmospheric circulation aspects. In doing so, it gains predictive reliability at the expense of informativeness. Put differently, focussing on model-based probabilistic projections means to fail to reasonably characterise regional climate information. This is acknowledged in the IPCC's AR6, where it says that a limitation of the probability-based approach includes "the failure to address physically plausible, but low probability high-impact scenarios." Against this background, I want to propose a complementary perspective to the standard framework that provides a tool for regionalscale information and that maintains relevant climate risk information while emphasising a distinction into epistemic and aleatoric uncertainty (Shepherd, 2019). More specifically, I aim to show how causal counterfactual storylines, that is "physically self-consistent unfoldings of past events, or of plausible future events or pathways" (Shepherd, 2019) formulated within the framework of causal modelling (Pearl, 2009) may prove to provide a complementary perspective to the interpretation of statistical analyses using the example of North America's 2021 heatwave. After briefly introducing the storyline approach and causal networks, I present a causal storyline of North America's heatwave based on Philip et al.'s (2021) rapid attribution analysis of the event. Subsequently, I show how their statistical analysis may be complemented with (a toy model of) causal attribution (following Hannart et al., 2016)---a simple conceptual framework for how to plausibly learn causal effects from data. It aims to demonstrate how to potentially causally complement attribution analyses by building on their statistical findings. In a nutshell, causal counterfactual storylines transform statements such as 'climate change made the heatwave 150 times more likely' to statements like these: 'anthropogenic climate forcing seems to be extremely likely to be a necessary and likely to be a sufficient cause for the occurrence of at least one heatwave in North America since the industrial era'. I conclude with a corollary of formulating eventbased storylines within the framework of causal models; namely that causal networks are (among others) most likely unable to provide us with a stance on the debate of what counts as internal and external variability (related to the general issue of where to draw the earth system's boundaries). However, because the relations between relevant climate factors are depicted clearly, one is able to visualise (e.g.) the human impact on the climate system and, consequently, on internal variability (Katzav and Parker, 2018). Secondly, we may be able to make uncertainty in changes in natural variability due to human forcing transparent in terms of treating aleatoric as-if-it-were epistemic uncertainty.

# Antonio, José and Escobar, Pérez

### Deflating mathematical difference making in biology: hexagons, honeycomb cells and grid cells

Recently, the issue of whether there are mathematical explanations outside mathematics has received substantial attention. Consider the following example. One can try to fit a hard plastic square peg inside a round hole the diameter of which is as long as one of the square's sides, but it will not fit. There is a physical and a mathematical component in the explanation of why the square peg does not fit. Physically, the properties of the square peg are such that when we try to fit it inside the round hole, the particles of the corner collide against the outside of the carved hole. Mathematically, the Pythagorean Theorem specifies that the diagonal of the square is longer than the diameter of the round hole, and hence, it is impossible for the square to fit in a circle the diameter of which is equal in length to one of the square's sides.

There are some accounts of mathematical explanatoriness outside mathematics. For instance, Baker (2005) argues that distinctively mathematical explanations are characterized by the indispensability of mathematical objects. Lyon and Colyvan (2008) appeal to the notion of inference to the best explanation and consider that mathematical explanatoriness relies on whether mathematics constitute the best explanation available. Lange (2013) argues that some explanations are distinctively mathematical because mathematical necessity, and not causes, has most of the explanatory burden (hence, mathematical explanations are of a modal kind). Baron, Colyvan and Ripley (2017) bring to this debate recent ideas from counterfactual reasoning and difference making. They suggest that there is a mathematical explanation outside mathematics when, should a given mathematical fact be otherwise, the fact outside mathematics would be different as well.

I will challenge the idea that counterfactual changes in pure mathematics ought to translate to changes in biology. I will discuss the case of the hexagonal shape of honeycomb cells and argue that it is not feasible to translate geometrical facts into biological shapes even from a neo-Aristotelian perspective (which acknowledges a relationship between abstract mathematical facts and empirical facts). According to my arguments, if pure mathematical fact about geometry were different, biological shapes would not be necessarily different. If pure mathematical facts about geometry do not make a difference on biological shapes, then, a fortiriori, the rest of accounts do not hold either: mathematical facts are not indispensable in the explanation of the biological shape, they do not constitute the best available explanation, and do not determine biological shapes by mathematical necessity.

However, I will also argue that mathematics does make an epistemic difference acting as heuristics. To this end, I will discuss another case of "hexagons" in biology hitherto untreated by the philosophical literature on mathematical explanations: the hexagonal periodicity of grid cell activity. Because grid cells were discovered in 2004 but their hexagonal periodicity was observed only after 2005, one can isolate the impact of such an epistemic difference in the explanations of grid cell activity. To do this, I will reconstruct the "mathematical" explanation featuring the intuitions of biologists/neuroscientists.

Baker, A. (2005). Are there Genuine Mathematical Explanations of Physical Phenomena?. Mind 114, 223–238.

Baron, S., Colyvan, M., & Ripley, D. (2017). How mathematics can make a difference. Philosophers' imprint, 1-19.

Lange, M. (2013). What makes a scientific explanation distinctively mathematical?. The British Journal for the Philosophy of Science 64(3), 485-511.

Lyon, A. & Colyvan, M. (2008). The Explanatory Power of Phase Spaces. Philosophia Mathematica 16, 1–17.

# Aybar, Raphael

### Hypothetical and Targetless Modeling in Cognitive Science: The cases of the Action Observation Network Model and the Agent-Environment Coupled System model

Modeling is a popular strategy in the scientific exploration of the mind. Scientists construct information processing models for predicting behaviors at the neural and phenomenal levels or for depicting hypothetical mechanisms that possibly explain brain behavior. To model the brain/mind, scientists must first describe target systems by relying on analogies between cognitive and computer systems, such as neural networks or dynamical systems. These analogies serve to characterize cognitive behavior as functional tasks that, due to their abstract description, can be realized by natural and artificial cognitive agents. Importantly, modelers furnish these functional descriptions with algorithmic structures that specify how cognitive tasks are achieved by following specific action rules.

Though the need for modeling is often recognized in the cognitive science literature, little attention has been paid to distinguishing the classes of cognitive science models. As a result, it is often unclear whether scientists are dealing with general or specific models of cognition or whether their models have target systems or not. This neglect may lead to subsuming the types of understanding different models provide to empty truisms, such as the idea that models provide understanding by representing. My presentation contributes to the philosophy of cognitive science by exploring how hypothetical and targetless models of cognition deliver scientific understanding. I will argue that hypothetical models serve to make predictions about the behavior of specific target systems, and targetless models serve the purpose of theory development, understood as the generation of new model descriptions and structures for modeling specific cognitive systems.

To this end, in my presentation, I will analyze the construction processes of the Action Observation Network (AON) model (Urgen and Saygin 2018) and the Agent-Environment Coupled System (AECS) model (Bruineberg et al. 2018). Based on Weisberg's (2013) account on model construction, I will explore the processes of model description, attribution of model structure, and model construal of these two models. The AON model has a specific target, namely a network of activations in specific brain regions triggered by the perception of action of biological systems. The AON model hypothesizes that a predictive mechanism explains the communication of the nodes of this network by using, apart from fMRI data, a model description, and structure taken from the predictive processing framework. Using model description and structure enables scientists to formulate falsifiable predictions concerning the N400 effect amplitude in mismatch conditions (e.g., in the perception of biological appearance and mechanical motion). The AECS model, on the other hand, is a targetless agent-based model, which offers a proof in principle that free energy equations can be used for modeling niche construction. The purpose of the AECS model is not the generation of scientific hypotheses and predictions about specific target systems but the development of new model descriptions and model structures that can be used to investigate hypothetical mechanisms and specific cognitive phenomena. The examination of these models will demonstrate that the understanding that models of cognition provide is heterogeneous and partly depends on their construction processes.

Bruineberg, Jelle, Erik Rietveld, Thomas Parr, Leendert van Maanen, and Karl J. Friston. 2018. "Free-Energy Minimization in Joint Agent-Environment Systems: A Niche Construction Perspective." Journal of Theoretical Biology 455: 161–78. https://doi.org/10.1016/j.jtbi.2018.07.002.

Urgen, Burcu, and Ayse Saygin. 2020. "Predictive Processing Account of Action Perception: Evidence from Effective Connectivity in the Action Observation Network." Cortex 128: 132– 42. https://doi.org/10.1016/j.cortex.2020.03.014.

Weisberg, Michael. 2013. Simulation and Similarity: Using Models to Understand the World. New York: Oxford University Press.

### Bakshi, Kabir

### Against Indirect Confirmation

Say that a proposition E is indirect evidence for a hypothesis H just in case E is evidence for H and E lies outside the intended domain of H. Call the confirmation of H due to E indirect confirmation. Is indirect confirmation possible? Many think that it is (i.a. Dawid 2013, Dardashti, Thébault, and Winsberg 2017).

A very weak principle that the proponents of indirect confirmation assent to is:

Indirect Confirmation (IC) A proposition E can confirm a scientific hypothesis H even if E lies outside the intended domain of H.

In a series of recent works, Richard Dawid and his collaborators Stephan Hartmann and Jan Sprenger (i.a. Dawid, Hartmann, and Sprenger 2015) have argued for expanding ambit of indirect confirmation (henceforth DHS). One of DHS's primary motivations in calling for this expansion is to make sense of confirmation in some branches of modern physics, esp. string theory.

My aim in this paper is to argue against (IC). I do this by critically evaluating DHS's recent influential argument for (IC): the No Alternatives Argument (NAA). The NAA is the claim that a hypothesis is confirmed by the fact that despite considerable efforts scientists have been unable to find any alternatives to H. Other critiques of NAA have focussed on arguing against NAA on the basis that it is, in some sense, unsatisfactory (Menon 2019, Rovelli 2019). I show that that DHS's argument for NAA is incorrect because it (i) overgenerates (ii) based on incorrect assumptions.

Roughly, the problem of overgeneration is that DHS's argument for (IC) makes confirmation cheap. DHS use the theory of Bayesian networks to prove the validity of NAA under some assumptions. But an important thing to keep in mind is that the variables in Bayesian network do not intrinsically stand for a specific proposition. In a further step each variable is endowed with an interpretation. I show that without changing the Bayesian network proffered by DHS but changing the interpretation of a variable we can show that H is confirmed by the fact that H is a dominant view in the relevant scientific community. This I think is an implausible result.

Roughly, the problem of incorrect assumptions is that DHS illegitimately use an assumption in their proof. A central assumption in DHS's proof is that the probability of a hypothesis being empirically adequate does not increase as the number of alternatives to the hypothesis increase (A4 in DHS 2015). I show that there are infinitely many counterexamples where (i) decreasing the number of alternatives to a hypothesis decreases the probability that the hypothesis is empirically adequate and (ii) increasing the number of alternatives to the hypothesis increases the probability that the hypothesis is empirically adequate. This I suggest is a fatal blow to DHS's program.

Connecting my critique of the NAA with (IC), I argue that we must dispense with indirect confirmation. I show that either (IC) is incorrect or if (IC) is correct, it collapses into direct confirmation. Hence, the take home message of this paper is that there is no such thing as indirect confirmation and indirect evidence.

Dardashti, Radin, Stephan Hartmann, Karim Thébault, and Eric Winsberg (2019). Hawking Radiation and Analogue Experiments: A Bayesian Analysis. In: Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 67, 1–11.

Dawid, R. (2013). String Theory and the Scientific Method. Cambridge University Press.

Dawid, Richard, Stephan Hartmann, and Jan Sprenger (2015). The No Alternatives Argument. In: British Journal for the Philosophy of Science, 66.1, 213–234.

Menon, T. (2019). On the Viability of the No Alternatives Argument. In: Studies in History and Philosophy of Science Part A, 76, 69–75.

Rovelli, C. (2019). "The Dangers of Non-Empirical Confirmation". In: Why Trust a Theory? Epistemology of Modern Physics. Ed. by Radin Dardashti, Richard Dawid, and Karim Thebault. Cambridge University Press

### Bielinska, Marta

### Testing spacetime orientability

Although the debate about the epistemology of geometry has a long pedigree, the question of whether or not spacetime is orientable has been sorely missing from this literature. In the vast majority of discussions it is simply taken from granted. This is not surprising: time orientability is, for example, an underlying premise of such robust theories as causality and quantum gravity, and their utility makes it convenient simply to assume it. We feel reassured in this belief due to the apparent incongruence of such objects as hands, and therefore we are not interested in the experimental confirmation of spacetime orientability. However, it turns out that our empirical evidence do not rule out the possibility that spacetime is, in fact, non-orientable. During my presentation, I address these issues and suggest possible ways of testing spacetime orientability.

I begin with an observation that there is some ambiguity in understanding spacetime orientability in the mathematical and scientific literature. In particular, I distinguish two kinds of this notion. One of them, which I call manifold orientability, is defined for all differentiable manifolds, whereas the other one, time and space orientability, is defined only for (3+1)-Lorentzian manifolds. I provide rigorous definitions of these notions and show that, in general, they are mutually independent. Moreover, I define respective kinds of spacetime orientations.

Having prepared formal setup for the further discussion, I argue that, on the basis of the available experimental data, it is impossible to determine whether our spacetime manifold is orientable in any of the senses listed above. In order to show this, I discuss several physical facts, and explicitly demonstrate how one can use them to argue in favor of various kinds of orientability and orientation. The examples that I discuss include: existence and the role of spinors in quantum field theory (manifold orientability), causality in general relativity (time orientability), parity violation in the Standard Model evidenced by the Wu experiment (space orientation), and the arrow of time (time orientation). Then I argue that each of these arguments can be rejected: in the case of quantum field theory – because spinors can be defined on non-parallelizable manifolds (see Pitts 2012); in the case of causality –because it can be violated in some circumstances; and in the case of parity violation and the arrow of time – because they are confined to some spacetime region (see Hawking and Ellis 1973).

Finally, since it seems that so far there is no empirical data that would rule out the possibility that our spacetime is non-orientable, I suggest an experiment specifically designed to survey time and space orientability, and discuss its limitations. This has been only done so far by Hadley (2002); I review his article and point out possible ways of improving his project. I conclude that it is physically impossible to experimentally confirm that our spacetime manifold is orientable and, therefore, one should not rule out the possibility that it is non-orientable.

Geroch, R. and Horowitz, G. (1979). Global structure of spacetimes. In S. Hawking and W. Israel (Eds.) General Relativity: An Einstein Centenary Survey. Cambridge: Cambridge University Press, pp. 212–293

Hadley, M. (2002). "The orientability of spacetime". In Classical and Quantum Gravity 19, pp. 4565-4572.

Hawking, S. and Ellis, G. (1973). The Large Scale Structure of Space-Time. Cambridge: Cambridge University Press

Pitts, B. (2012). "The nontriviality of trivial general covariance: How electrons restrict 'time' coordinates, spinors (almost) fit into tensor calculus, and 7 16 of a tetrad is surplus structure". In Studies in History and Philosophy of Modern Physics 43, pp. 1-12.

### Bräutigam, Maren

### On the qualitative identity of similar fermions: orthodoxy versus heterodoxy

There are two different kinds of views on the qualitative identity of similar fermions, the orthodox, and the heterodox kind. According to one orthodox view (see French and Redhead (1988)), similar fermions are qualitatively identical in all of the allowed states, so that Leibniz's principle of the identity of indiscernibles (PII) is violated. As several authors have pointed out (see, e.g., Bigaj (2022), Caulton (2018), Dieks and Lubberdink (2020)), this view relies on the semantical standard interpretation that the indices of the Hilbertspace formalism refer directly to particles (direct factorism). In contrast to that, heterodox views rely on either descriptive anti-factorism, i.e. the view that new labels with descriptive content can be introduced from outside the formalism (see Dieks and Lubberdink (2020) and Friebe (2014)), or descriptive factorism, i.e. the view that indices of the formalism refer to particles via properties (see Leegwater and Muller (2020)), to establish the qualitative distinctness of similar fermions in at least some (if not all) of the allowed states, thereby validating PII.

In this talk, I argue that it is not yet sufficiently clear what these two heterodox approaches amount to, both ontologically and semantically speaking. I introduce two distinctions, one bearing on ontology, the other on semantics. The first distinction concerns two different interpretations of PII, which I call the grounding interpretation and the accompanying interpretation. According to the grounding interpretation, PII is supposed to work as a principle of individuation, saying that whenever two objects are numerically distinct, their numerical distinctness is grounded in their being qualitatively distinct. According to the accompanying interpretation, however, PII is not supposed to work as a principle of individuation itself (i.e. objects are individuated independently from qualitative distinctness), but as an additional principle which ensures that numerical distinctness is always accompanied by qualitative distinctness. As I will argue, these different interpretations of PII come with different ontological commitments. I will then investigate in which sense orthodoxy denies and heterodoxy affirms that PII is validated for similar fermions. This helps to clarify how the latter is opposing the former ontologically.

The second distinction aims at determining how heterodoxy and orthodoxy differ semantically. At first glance, one might think that both descriptive anti-factorism and descriptive factorism imply descriptive reference and thereby deviate from the standard semantics (which involves directly referring names). By drawing on the distinction between semantics and metasemantics (see Michaelson and Reimar, (2019), sec. 2.2), however, I argue that the semantics of proper names answers the question in which way a name refers, or what the meaning of a name consists in, while the metasemantics of proper names answers the question name refers to the object it refers to, or how the reference of a name gets fixed. By showing that 1. descriptive reference fixing does not imply descriptive reference, and 2. that both descriptive antifactorism and descriptive factorism are concerned with descriptive reference fixing, I argue that both heterodox positions establish that they differ from orthodoxy metasemantically, but do not establish that they differ from it semantically. This is an important point in an overall evaluation and comparison of orthodox and heterodox positions.

Bigaj, T. (2022). Identity and Indiscernibility in Quantum Mechanics, Palgrave Macmillan.

Caulton, A. (2014). Qualitative Individuation in Permutation-Invariant Quantum Mechanics, arXiv:1409.0247v1 [quant-ph].

Dieks, D. and Lubberdink, A. (2020). Identical Particles as Distinghuishable Objects, General Journal of Philosophy, https://doi.org/10.1007/s10838-020-09510-w.

French, S. and Redhead, M. (1988). Quantum Physics and the Identity of Indiscernibles, The British Journal for the Philosophy of Science 39 (2), 233-246.

Friebe, C. (2014). Individuality, distinguishability, and (non-)entanglement: A defense of Leibniz's principle, Studies in History and Philosophy of Modern Physics 48 (A), 89-98.

Leegwater, G. and Muller, F.A. (2020). The Case Against Factorism. On the Labels of -Factor ⊗ Hilbert-Spaces of Similar Particles in Quantum Mechanics, Journal for General Philosophy of Sciene, https://doi.org/10.1007/s10838-020-09514-6.

Michaelson, E. and Reimer, M. (2019). Reference, The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2019/entries/reference.

# Britten-Neish, George

### Isomorphic vs. non-Isomorphic Styles of Subpersonal Explanation: Migraine Aura Phenomenology

The notion of subpersonal explanation plays a central role in many naturalising projects in philosophy of mind. A standard objection to attempts to explain personal-level phenomena in subpersonal terms is that they problematically reduce their target phenomenon; they 'explain away' rather than explain. In this paper, I argue that this criticism is misdirected. I distinguish between subpersonal explanations that respect or fail to respect a content isomorphism principle. I discuss a successful isomorphic subpersonal explanation of the phenomenon of visual migraine aura, and contrast it with more controversial cases in the philosophical literature. I suggest that here, isomorphic explanation often introduces inappropriate constraints, and non-isomorphic explanations are preferable. The problem

Subpersonal explanation is best understood as a kind of content-ascriptive mechanistic explanation that targets psychological phenomena. It works by integrating a way of analysing psychological phenomena (functional analysis) and a general principle of mechanistic explanation (structural decomposition). Psychological phenomena are characterised as a series of operations on mental contents, and those operations are assigned to component parts of neurophysiological mechanisms. These can now be characterised as neurocomputational mechanisms, their functioning interpretable as computation over contents which are involved in realising the relevant, personal-level, psychological phenomenon.

One way would be by putting mechanistic components and elements of their personal-level explanatory targets into one-one correspondence:

Content isomorphism: For each component of a psychological phenomenon,  $\Psi$ n, there is a corresponding neurophysiological mechanisms component, Mn, so that  $\Psi$ n and Mn have the same content

The relevant sense of 'isomorphism' here is the intuitive notion of a structure preserving mapping across domains, familiar from the philosophical literature. The domains in question are those mapped out by a task analysis (at the personal level) and a functional decomposition (at the subpersonal level) in the way described above. A contentually isomorphic explanation maps contents ascribed to different elements in the functional analysis to components of neurophysiological mechanisms identified by structural decomposition.

Understanding subpersonal explanation in this way works for explaining some psychological phenomena. For example the phenomenology of scintillating scotoma during migraine aura can be quite fully explained in terms of isomorphic mappings from retinotopic maps in the early visual cortex to characteristic disruptions in visual processing, describable by task-analysis. Whether visual phenomenology in general is amenable to this approach depends on whether task-analysis of ordinary perception can be mapped to the brain in the same way. Recent work on natural scene perception involving retroperceptual effects and sensorimotor affordances and suggest not.

Opponents of reductionism should therefore reformulate their criticisms: from a general objection to subpersonal explanation, towards a focus on putative explanations of particular phenomena. However, criticism of subpersonal explanation also frequently conflates it with an ancillary project. Fully reductionist subpersonal explanations depend on a naturalistic theory of content. Much of what is really at issue between proponents and critics of explanatory reduction turns on the availability of such a theory.

### References:

Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. Philosophical perspectives, 26, 1-18.

Goadsby, P. J., Holland, P. R., Martins-Oliveira, M., Hoffmann, J., Schankin, C., & Akerman, S. (2017). Pathophysiology of migraine: a disorder of sensory processing. Physiological reviews.

Hurley, S. (2003). Action, the unity of consciousness, and vehicle externalism. The unity of consciousness: Binding, integration, and dissociation, 78-91.

Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. Cortex, 98, 262-268.

### Brown, Rachael

# Gradualism as a constraint on theorising in comparative and evolutionary psychology

In From Signal to Symbol (2021, p. x) Ron Planer and Kim Sterelny argue that any "adequate" theory of language evolution "must identify a plausible trajectory from greatapelike communicative abilities to those of modern humans where each step along the way is small, cumulative and adaptive (or at least not maladaptive: there might be some role for drift)". They are not alone in invoking such a constraint. Gradualism is cited as an important assumption amongst those concerned with the evolution of cognition and the nature of animal minds going right back to Darwin's mental continuity thesis (Darwin 18XX/Descent of Man).

The gradualist assumption is typically invoked by scholars wishing to push back against the anthropocentric allure of human uniqueness; the idea being that the postulation of the evolution of entirely novel cognitive capacities in our lineage alone is evolutionarily implausible. Indeed, Planer and Sterelny call the capacities such theories postulate "miracles" (2021, p. 213). In this paper, I explore the evolutionary justification for the gradualist assumption in comparative and evolutionary psychology given the understanding of phenotypic evolution offered by contemporary evolutionary developmental biology (evodevo). In particular, I ask, are evolutionary trajectories made up of "small, cumulative and adaptive" steps indeed more evolutionarily plausible than those that postulate entirely novel cognitive capacities within lineages? If so, why?

One reason one might question the gradualist assumption (or at least suggest it needs to be applied with more care) comes from evidence that, although genetic evolution is typically gradual, microevolutionary change of this type is not always associated with gradual change at the phenotypic level (Moczek 2008). As understanding of the relationship between genes and phenotypes in development has grown, so too has an appreciation of the important role played by neutral evolution and developmental mechanisms in evolution. At least from the perspective of evo-devo, the existence of such mechanisms undermines any bald gradualist assumption based on the gradualism of micro-evolution. Even if genetic evolution is gradual, one cannot assume that phenotypic evolution will be.

Another reason to question gradualism lies in the nature of the supply of variation. Gradualism is often justified on the belief that it is much more likely for large random phenotypic changes to be deleterious than small ones, and thus, most large phenotypic shifts will fail to persist and propagate in populations (Calcott 2011). Again, however, work in evo-devo on plasticity and other mechanisms of adaptation suggests that there are ways that developmental systems have evolved to make large adaptive shifts in phenotype possible (Moczek 2008). This also undermines any bald gradualist assumption.

In this paper, I look at whether this new knowledge from evo-devo undermines the gradualist assumption in comparative and evolutionary psychology as invoked by scholars such as Planer and Sterelny (2021). Ultimately, I offer a novel account of gradualism as a constraint on theorising in comparative and evolutionary psychology which better reflects contemporary evolutionary developmental biology.

Calcott B (2011) Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings. Biol Philos 26:281–293. https://doi.org/10.1007/s10539-010-9202-x

Moczek AP (2008) On the origins of novelty in development and evolution. BioEssays 30:432–447. https://doi.org/10.1002/bies.20754

Planer RJ & Sterelny, K (2021) From signal to symbol: The evolution of language. The MIT Press

## Calderón, Francisco

### The Causal Axioms of Algebraic Quantum Field Theory: A Diagnostic

This paper examines the axioms of algebraic quantum field theory (AQFT) that aim to characterize the theory as one that implements relativistic causation. I suggest that the spectrum condition (SC), microcausality (MC), and primitive causality (PC) axioms, taken individually, fail to fulfill this goal. This claim goes against Jeremy Butterfield's claim that SC is the "most direct expression of the prohibition of spacelike processes" (2007, Reconsidering Relativistic Causality, International Studies in the Philosophy of Science, 21:3, 295-328) and to John Earman and Giovanni Valente's (2014, Relativistic Causality in Algebraic Quantum Field Theory, International Studies in the Philosophy of Science, 28:1, 1-48) argument for proposing PC instead. I argue in favor of what Hofer-Szabó and Vecsernyés (2018, Locality and Causality Principles. In: Quantum Theory and Local Causality. SpringerBriefs in Philosophy. Springer, Cham.) call the "local primitive causality" (LPC) condition. This condition is implied by the first three axioms and other considerations, so I will incidentally prove that we should not worry about them being redundant or of the theory being unparsimonious.

My methodology is that of a diagnostic: if we analyze what it means for an axiom of AQFT to be (i) an axiom of a field theory that has inputs from (ii) special relativity and (iii) quantum mechanics, I identify the features that each axiom contributes to a notion of relativistic causation in AQFT. For the notion of "relativistic causation," my guiding principle is the prohibition of no-superluminal signaling. However, other aspects of "causation" (as physicists use the term, not philosophers) as a form of deterministic local action are relevant desiderata for deeming each axiom fit or unfit to be considered a "causal axiom."

After analyzing each axiom, I will conclude (i) that SC fails to characterize relativistic causation since it is not in virtue of this condition that quantum fields exhibiting no superluminal signaling avoid this problem; (ii) that MC will import some of the interpretive difficulties of the measurement problem of standard quantum mechanics, thus obscuring the claim that MC clearly expresses the condition of no-superluminal signaling; and (iii) that the determinism embodied by PC is not enough to characterize relativistic causation.

However, each axiom specifies some aspect of relativistic causation. SC makes the metric structure of Minkowski spacetime salient. MC emphasizes locality and, to a certain extent, the limit of the influence that one quantum system can have on another one. Finally, PC puts determinism at the center stage. Given the difficulties I mentioned before, we need a condition that captures each axiom's advantages to have an appropriate characterization of relativistic causation in AQFT. LPC is such a condition. This claim follows immediately from the groundwork I lay for each axiom and from a construction from Haag and Schroer (1962, Postulates of Quantum Field Theory

J. Math. Phys. 3, 248) that makes explicit that the SC, MC, and PC, taken together, imply LPC. Assuming LPC has become a widespread move in more technical literature in AQFT. Still, its motivations are rarely stated, its role as a causal axiom is left uninterpreted, and its interdependence with the other causal axioms is ignored.

## Coraci, Davide and Cevolani, Gustavo

### Probability, Confirmation, and Reverse Inference in Cognitive Neuroscience

Reverse inference (RI) is a crucial inferential strategy widely employed in cognitive neuroscience to derive conclusions about the engagement of cognitive functions from patterns of brain activation. Despite its central role, in recent years RI faced increasing skepticism, especially after the influential critical analysis advanced in Poldrack (2006). Here, we propose Bayesian confirmation theory (e.g., Sprenger and Hartmann 2019) as a new way to advance the debate. Experiments based on functional Magnetic Resonance Imaging (fMRI) allow researchers to observe patterns of brain activation in subjects performing a given task. In "forward" inference, one assumes that some cognitive process Cog is recruited during a task and tries to individuate the most consistent activation Act representing its neural correlate. A "reverse" inference, instead, consists in concluding, based on such experimental evidence, the likely engagement of Cog from observing Act in a new experiment. While widely employed, RI is controversial for several reasons (Poldrack 2006). First, it is logically invalid, reflecting the fallacy of "affirming the consequent". Second, it is affected by the problem of (lack of) selectivity of the neural regions and the absence of a clear-cut cognitive ontology. Third, its concrete applications can be criticized on both methodological and experimental grounds (Wager et al. 2016). Despite recent attempts to assess the status of RI, the debate is still ongoing among both neuroscientists and philosophers (Costa et al. 2021; Calzavarini and Cevolani 2022).

In this paper, we start from Poldrack's Bayesian analysis, which formalizes RI by computing the probability p(Cog|Act) of Cog conditional on Act as based on meta-analytic data extracted from the experimental literature. We then evaluate neuroscientists' current practice of using Bayes Factors (BF) to quantify the selectivity and hence the reliability of RI (Poldrack 2006; Costa et al. 2021). Finally, we advance three main claims, using simple case studies to illustrate each of them. First, that the well-known (if often overlooked) distinction between posterior probability and confirmation is instrumental in clarifying current research practices (e.g., by noting that BF can be construed as a measure of confirmation). Second, that assuming "flat" prior probabilities, as consistently done in the neuroscience literature, undermines the probability/confirmation distinction and makes the Bayesian analysis of RI conceptually confused. Third, that under the flat prior assumption RI, being completely driven by the likelihood p(Act|Cog), actually collapses on forward inference, thus blurring the distinction between these two kinds of inferences. If we are correct, the Bayesian analysis of RI needs a serious re-thinking; we conclude with a discussion of the alternatives and some suggestion to make it more robust.

### References:

Calzavarini, F., Cevolani, G. (2022). "Abductive Reasoning in Cognitive Neuroscience: Weak and Strong Reverse Inference". Forthcoming in Synthese.

Costa, T., Manuello, J., Ferraro, M., Liloia, D., Nani, A., Fox, P. T., ..., Cauda, F. (2021). "BACON: A tool for reverse inference in brain activation and alteration". In: Human Brain Mapping 42(11), pp. 3343-3351.

Poldrack, R. A. (2006). "Can cognitive processes be inferred from neuroimaging data?" In: Trends in Cognitive Sciences 10.2, pp. 59–63.

Sprenger, J., Hartmann, S. (2019). Bayesian philosophy of science. Oxford University Press.

Wager, T. D., Atlas, L. Y., Botvinick, M. M., Chang, L. J., Coghill, R. C., Davis, K. D., Iannetti,G. D., Poldrack, R. A., Shackman, A. J., and Yarkoni, T. (2016). "Pain in the ACC?". In: Proceedings of the National Academy of Sciences, 113(18), pp. E2474–E2475.

# Curtis-Trudel, Andre

### Are computational explanations mechanistic explanations?

What is computational explanation? According to a popular recent view, computational explanation is a kind of mechanistic explanation. This paper argues against the mechanistic view of computational explanation on the grounds that it overlooks a broad class of computational explanations of central importance to contemporary computer science. This broad class of explanations, which I call 'limitative' explanations, explain why certain problems cannot be solved computationally, either in principle or under certain constraints on computational resources such as time or space. After introducing the mechanistic view, I survey a representative sample of limitative explanations, argue that they are not mechanistic explanations, and close with some thoughts about the nature of computational explanation more broadly.

### Devanesan, Arjun

### Immunology and pregnancy- a reply to Pradeu

The traditional conception of immune function is that of a system which differentiates the organism's own tissues (the self) from any foreign invaders (nonself). Mammalian (and in particular human) pregnancy presents a serious challenge to the self-nonself theory which has sometimes been called the immunological paradox of pregnancy. Transplants from a one body to another are usually rejected by the recipient, so "how does the pregnant mother contrive to nourish within itself, for many weeks or months, a foetus that is an antigenically foreign body?" (Medawar 1953). While known to be paradoxical, the self-nonself theory still forms the basic heuristic in immunology research and clinical practice.

In section 1 of this paper I present the self-nonself theory of immunology and the immune paradox of pregnancy in cellular and molecular terms. I argue that the what is known about the immunological interactions that occur during pregnancy are sufficient to reject the self-nonself theory as a general model for immune function.

In section 2 I will present Thomas Pradeu's (2012) alternative account of immune function, the continuity theory, which he proposes as a superior unifying or reductive account of immunology. According to the continuity theory, the immune system does not operate on the basis of the distinction between self and nonself and so there is no immune paradox to be solved. On the one hand, discontinuity theory provides valuable insight into the functioning of the immune system and avoids an important paradox. However, for continuity theory to succeed as a unifying account, it must be able to explain all major immune phenomena in terms of discontinuity. To be reductive, all immune behaviour must be explained by a response to discontinuities. I will argue that there are diverse immune phenomena which are integral to physiological pregnancy which cannot be unified by Pradeu's account or reduced to discontinuity.

Firstly, continuity theory fails to establish the important topology and context sensitivity of the immune system. While Pradeu acknowledges that immune function is context sensitive, he provides no good explanation or mechanism for how this occurs. To be fair, it is not clear to anyone how this occurs but I will argue that context sensitivity and immune topology cannot be reduced to continuity or discontinuity.

Secondly, Pradeu's theory preserves the notion that immune function is dichotomous there is either continuity or discontinuity and immune tolerance or rejection respectively. In pregnancy, however, when the immune system is activated during foetal implantation, this leads to tissue destruction and inflammation of the uterine wall, but acceptance rather than rejection of the foetus. In discontinuity theory, tissue destruction entails immune rejection whereas in pregnancy tissue destruction is part of tissue remodelling for acceptance of the foetus. If these two qualitatively different immune phenomena are both triggered by discontinuity and continuity theory cannot explain the difference, it cannot unify them under a single guiding principle.

### Dewar, Neil

### Intrinsic quantum mechanics

Recently, Dewar (forthcoming) has outlined a proposal for characterising the intrinsic structure of physical quantities. However, Dewar's discussion considers only classical physics. This paper considers how to extend Dewar's apparatus to quantum physics. Dewar characterises his work as a variation on Field's program (Field 1980), and Malament (1982) famously argues that extending Field's work to quantum mechanics would prove especially difficult; it is therefore of interest to know if Dewar's approach fares any better than Field's.

First, I show how to characterise a discrete quantity in a fashion analogous to Dewar: whereas the quantities that Dewar considers are obtained as principal homogeneous spaces over the strictly positive real numbers, discrete quantities such as spin may be obtained as principal homogeneous spaces over the strictly positive integers.

Next, let Q be such an intrinsically characterised quantity (whether continuous or discrete): for example, the space of possible positions for a single spinless particle (which is isomorphic to physical space), or of possible spin-magnitudes for a spin-1/2 particle. It is shown that the set of all complex-valued functions on Q constitutes a Hilbert space.

However, so characterised, this Hilbert space has a preferred basis: essentially, the eigenbasis for the quantity Q. So the next question is how different quantities might be associated to the "same" Hilbert space, with a view to then abstracting away from any preferred basis for that Hilbert space. For instance, if Q is physical space, then we know that the Hilbert space associated to the quantity P of momentum (i.e. the set of complex functions on P) should be identifiable with that associated to Q. In other words: what makes it the case that Q and P are complementary quantities?

To answer this question, we must introduce Lagrangian and Hamiltonian mechanics into Dewar's framework; in doing so, I employ (a much more basic version of) some ideas recently discussed in Zapata-Carratala (2019). By enriching Dewar's kinematical algebra of quantities with a Lagrangian dynamics, we are in a position to identify certain pairs of quantities as complementary to one another, and hence to establish a canonical isomorphism between the Hilbert spaces associated to them. For instance, in the example above, we are able to establish a canonical Hilbert-space isomorphism between the set of all complex functions on Q and the set of all complex functions on P: essentially, the Fourier transform. We can then take the abstract Hilbert space (i.e. the Hilbert space considered independently of any preferred basis) as the common structure between these two Hilbert spaces (each of which has a preferred basis); for, the isomorphism in question does not preserve the preferred basis.

This is still only a preliminary investigation: it remains to be seen how to take tensor products of Hilbert spaces in this framework, or how the C\*-algebra formalism might be applied (since "sums" of quantities will not, in general, be well-defined). However, I hope that this represents at least a reasonable starting-point.

Dewar, N. (forthcoming). On Absolute Units. The British Journal for the Philosophy of Science, 715236. https://doi.org/10.1086/715236

Field, H. (1980). Science Without Numbers: A Defence of Nominalism. Blackwell Publishers Ltd.

Malament, D. (1982). Review of Science Without Numbers: A Defense of Nominalism. [Review of Review of Science Without Numbers: A Defense of Nominalism., by H. H. Field]. The Journal of Philosophy, 79(9), 523–534. https://doi.org/10.2307/2026384

Zapata-Carratala, C. (2019). A Landscape of Hamiltonian Phase Spaces: On the foundations and generalizations of one of the most powerful ideas of modern science [PhD thesis, University of Edinburgh]. <u>http://arxiv.org/abs/1910.08469</u>

### Emmerson, Nicholas

### It Ain't That Deep: Metaphysics and the Problem of Progress

In this paper I provide a novel, unifying account of progress across science and metaphysics. Progress, on my view, is made when both scientists and metaphysicians grasp deepening explanations of a target phenomenon. In doing so, I utilize a methodology recently developed by Finnur Dellsén, Insa Lawler & James Norton (2021), who argue that we ought to use science as a "testing ground" for a general account of progress, before then applying the resultant notion to philosophy.

At a first glance metaphysics might seem like a discipline where such a methodology would prove particularly fruitful. As James Ladyman and Don Ross argue, for example, 'to the extent that metaphysics is closely motivated by science, we should expect to make progress in metaphysics iff we can make progress in science' (2007:35). In a recent paper, however, Kerry McKenzie (2020) has argued that metaphysics cannot make progress in same way that science and that, as a result, we have little reason to think that it makes progress at all.

McKenzie (2020) takes scientific progress to be constituted by better approximating the truth; a methodology which cannot be meaningfully applied to metaphysical theses. Crucially, McKenzie suggests that the notion of correspondence, or "retention through change", is a necessary component of approximation in science which has no analogue within metaphysics. Despite this, I believe that a more optimistic attitude can be motivated towards a unified accounts of progress across both science and metaphysics.

My approach builds upon Dellsén, Lawler & Norton's (2021) "testing ground" methodology by combining Dellsén's (2016) understanding-based conception of scientific progress with unifying interventionist analyses of both explanation (Schaffer 2017) and explanatory depth (Hitchcock & Woodward 2003b). On the resulting thesis, scientific and metaphysical progress are constituted, not by better approximating the truth, but by increasing understanding; providing explanations which are deeper than their predecessor(s).

I suggest that correspondence, and thus progress, can be captured by a scientific explanation being invariant under a wider range of interventions than its predecessor, where this range contains those interventions under which the prior theory was invariant. In order to move this depth-based account of progress beyond the 'testing ground' of science, I argue that an analogous notion of explanatory depth is operative in the metaphysical domain.

Finally, I apply this notion of progress to a case-study concerning two rival accounts of the identity and distinctness of concrete objects: the qualitative properties proposal (QPP) and the weak discernibility proposal (WDP). The WDP, I maintain, is progressive with respect to the prior QPP precisely because the former theory provides explanations which remain invariant under a wider range of testing interventions than those provided by the latter. What's more, since the range R\*, under which explanations provided by the WDP are invariant, strictly contains the range R, under which explanations provided by the QPP are invariant, these theories can meaningfully be said to correspond.

Dellsén, F. (2016). Scientific progress: Knowledge versus understanding. Studies in History and Philosophy of Science Part A 56:72-83.

Dellsén, F. Lawler, I. & Norton, J. (2021). Thinking about Progress: From Science to Philosophy. Noûs https://doi.org/10.1111/nous.12383.

Hitchcock, C. & Woodward, J. (2003b). Explanatory Generalizations, Part II: Plumbing Explanatory Depth. Noûs 37 (2):181–199.

McKenzie, K. (2020). A Curse on Both Houses: Naturalistic Versus A Priori Metaphysics and the Problem of Progress. Res Philosophica 97 (1):1-29.

Schaffer, J. (2017). Laws for Metaphysical Explanation. Philosophical Studies 27 (1):303-321.

### Erasmus, Adrian

### A Pragmatic Approach to Disease Classification

Nosology, the part of medicine concerned with disease classification, is crucial to both the practical and theoretical goals of medicine. It not only improves communication among researchers, physicians, and insurers, but also informs our understanding of different diseases, thus contributing to diagnosis, drug development, and therapeutic prediction. Given its significance in many areas of medicine, it is important to analyze and evaluate the principles underlying disease classification. In this paper, I examine three conventional approaches in nosology and argue for a broadly pragmatic approach to disease classification.

In modern medicine, there are three prominent approaches to disease classification: the etiological approach, the symptom-based approach, and the pathophysiological approach. The etiological approach involves classifying diseases according to their distal causes—those causes which are extraneous to the constitutive mechanisms of the disease. The symptom-based approach classifies diseases according to clinical picture or syndrome. Finally, the pathophysiological approach appeals to abnormal biomechanistic structure or function. My first aim is a clarificatory one. I analyze these three approaches to disease classification, outline their distinctive principles, and touch on the advantages and disadvantages of each.

Despite its importance there is relatively little philosophical work on disease classification. Many accounts are limited to psychiatry and the purported flaws of using a symptom-based approach to classifying mental disorders (e.g., Zachar and Kendler 2017). Other work, on physical disease classification, typically examines the different models that fall under the etiological approach (Cooper 2002; Broadbent 2013). Common to these accounts is the assumption that the etiological approach is best. For example, it is typically argued that etiological classifications are more appropriate than symptom-based classifications typically used in psychiatry. And, in discussions regarding etiological classification, much emphasis is placed on whether diseases are best classified using a monocausal or multifactorial model, without considering alternative approaches. Some medical professionals also prefer the etiological approach, maintaining that such classifications demonstrate a better understanding of a disease and that, when we are able to classify diseases in different ways, etiological classifications are more useful than pathophysiological ones, which in turn are more useful than symptom-based classifications (Snider 2003).

To my knowledge, the only philosophical account to consider a different approach is that of Fuller (2018), who argues that diseases are classified pathophysiologically. However, the reality is that physical diseases are often classified using other approaches. My second aim in this paper is to defend a broadly pragmatic approach to disease classification. That is, classificatory choices should depend on the medical goal being pursued. The approach to disease classification I suggest is this: the medical goal in question determines what kind of information about a disease is important, which in turn determines the classificatory approach that should be used. This is because each model has particular advantages that are helpful in certain circumstances. To illustrate this point, I provide an argument for the use of the pathophysiological approach in the context of therapeutic prediction.

Broadbent, A. (2013) Philosophy of epidemiology. New York: Palgrave Macmillan.

Cooper, R. (2002) 'Disease', Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 33(2), pp. 263–282.

Fuller, J. (2018) 'Universal etiology, multifactorial diseases and the constitutive model of disease classification', Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, pp. 8–15.

Snider, G. L. (2003) 'Nosology for our day: its application to chronic obstructive pulmonary disease', American Journal of Respiratory and Critical Care Medicine, 167(5), pp. 678–683.

Zachar, P. and Kendler, K. S. (2017) 'The philosophy of nosology', Annual Review of Clinical Psychology, 13(1), pp. 49–71.

# Franklin, Alexander

### Incoherent? No, Just Decoherent: How Quantum Many Worlds Emerge

In its modern guise the Everett interpretation of quantum mechanics describes an emergent multiverse. The goal of this talk is to provide a perspicuous characterisation of how the multiverse emerges making use of a recent account of ontological emergence. This will enable the rebuttal of recent critiques of the emergent multiverse ontology as incoherent.

I'll first demonstrate that the emergence of a quasi-classical world in Everettian Quantum Mechanics can be described by the same philosophical formalism as is applicable in other, non-quantum contexts, making use of a recent account of ontological emergence; see XXXX. Claims about how these worlds emerge are subject to a critique, developed at length in Dawid and Thébault (2015). The argument is that the components that modern Everettians, such as Wallace (2012) take to be required to resolve various conceptual puzzles result in incoherence and vicious circularity. An upshot of clarity over the details of the emergence of Everettian worlds is that such critiques may be undermined.

In some more detail: it's well known that one Everettian strategy to make sense of probabilistic predictions is to adopt a decision-theoretic framework – this involves analysis of the rational betting strategy of ideal agents faced with world splittings. Dawid and Thébault build on Baker (2007) and Zurek (2003) to argue that such justifications are beset by vicious circularities since probabilistic reasoning is assumed to be required to evidence the claim that observers emerge at all. That's because the decoherence framework for emergence entails that interference terms are very small, and some further argument is needed to justify the claim that such terms are negligible: if that justification relies on the decision-theoretic strategy, we have circularity, because the emergence of observers is required in order to make sense of the decision-theoretic strategy.

I contend that this circularity can be avoided if two appeals to the Born rule are distinguished – while Born measure averaging is ubiquitous, Born rule as probability is only required in special cases. I'll leave controversies over the interpretation of applications of the latter as full-fledged probabilities to another occasion, though this of course well-trod territory. This talk focusses on the Born measure contexts and argues that these are all that's needed for the emergence of the multiverse. Pace Dawid and Thébault, I argue that the negligibility of interference terms as a consequence of decoherence can be independently and non-probabilistically empirically confirmed. Therefore I claim that one can justify discarding very small terms on this basis.

I defend this claim by appealing to the example of the orbit of Hyperion – following Habib, Shizume, and Zurek (1998) and Berry (2001) – in this case application of decoherence theory predicts a determinate orbit for this moon of Saturn and no appeal to full-fledged probabilities is required. Thus, the circularity is avoided.

The conclusion is that the framework as a whole is coherent and poses a respectable metaphysics notwithstanding unsettled questions about the nature of the full-fledged probabilities.

Baker, David J (2007). "Measurement outcomes and probability in Everettian quantum mechanics". In: Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics 38.1, pp. 153–169.

Berry, Michael V (2001). "Chaos and the semiclassical limit of quantum mechanics (is the moon there when somebody looks?)" In: Quantum Mechanics: Scientific perspectives on divine action 41, p. 56.

Dawid, Richard and Karim P. Y. Thébault (2015). "Many worlds: decoherent or incoherent?" In: Synthese 192.5, pp. 1559–1580.

Habib, Salman, Kosuke Shizume, and Wojciech Hubert Zurek (1998). "Decoherence, chaos, and the correspondence principle". In: Physical review letters 80.20, p. 4361.

Wallace, David (2012). The Emergent Multiverse: Quantum Theory according to the Everett Interpretation. Oxford University Press.

Zurek, Wojciech Hubert (May 2003). "Decoherence, einselection, and the quantum origins of the classical". In: Rev. Mod. Phys. 75 (3), pp. 715–775.

### Furman, Katherine

### **Objective Activist Social Research**

Many philosophers of science and social science are concerned about the role of values in research, especially social research. A key concern is whether this undermines objectivity, and thus public trust in research outcomes. Typical responses are that the values should be transparent, that they should be open to public scrutiny, and some argue that the values should be democratically selected (De Melo-Martin & Intemann, 2018; Douglas, 2009; Kitcher, 2011). Social scientists who study the way practitioners engage with threats to public trust argue that prominent researchers engage in "credibility management" strategies (Geiger, 2021), this can involve suppressing personal values in public or presenting values in such a way that the practitioner is perceived as an 'honest broker' – that is, someone who wouldn't allow their personal values to overwhelm their research.

In this context, activist researchers – that is, researchers with explicit commitments to social justice and/or social change – appear to pose a special challenge to objective social research. While the values are transparent, these researchers are open to the criticism that their social commitments bias their research. While their values are transparent, they are not doing the credibility management work described Geiger.

In this paper, I will respond to this concern about activist researchers by noting that there are many ways of being an activist social scientist. One way is to consider oneself as playing a specific role within a broader social movement. This is the role that Sally Haslanger has defended in public interviews (Haslanger, 2021). A social movement requires various members satisfy specific roles – someone has to make the banners, someone has to manage the accounts, someone else has to do the research. From this perspective, just like we would require that the movement's accountant satisfy the norms of 'good' accountancy, so too we would require that activist researchers satisfy the norms of 'good' research. Not anything that aligns with our values would be acceptable.

To defend my position, I will look at the cases of two South African social scientists who were also activists against the apartheid state – Harold Wolpe and Ruth First. Wolpe was an anti-apartheid activist who escaped prison and took refuge in the UK until the end of apartheid, when he returned to South Africa to an academic role and researched how to develop education policy in newly democratic South Africa. This research was from an explicit activist viewpoint – how to improve educational justice – but the stakes were high and the need to get the research "right" was a pressing concern. Doing research that merely aligned with the activist goals would have been unsatisfying. Ruth First is perhaps South Africa's most famous activist, assassinated by the apartheid state in 1982. She was also a social scientist and a Sociology lecturer at Durham University. At the end of her life she was working in Mozambique on the issue of how Mozambique could economically disentangle itself from South Africa. Again, producing research that merely aligned with activist aims would have been insufficient. Paying attention to scientific objectivity in activist research remains important and this talk will look at how this can be done.

De Melo-Martin, I., & Intemann, K. (2018). The Fight Against Doubt: How to Bridge the Gap Between Scientists and the Public. Oxford: Oxford University Press.

Douglas, H. (2009). Science, Policy and the Value-Free Ideal. Pittsburgh, PA: Pittsburgh University Press.

Geiger, B. (2021). Performing Trustworthiness: The 'Credibility Work' of Prominent Sociologists. Sociology, 785-802.

Haslanger, S. (2021, June 21). Doing Justice to the Social. (J. Bacavic, Interviewer)

Kitcher, P. (2011). Science in Democratic Society. Prometheus Books.
## Gebharter, Alexander and Osimani, Barbara

#### Analogical inference Bayesian style 2.0

Recently, Dardashti, Hartmann, Thébault, and Winsberg (2019) proposed a Bayesian model for analogical inference. In this talk we investigate how their model performs when varying the degree of certainty about the similarity between the source and the target system. We show that there are circumstances in which the degree of confirmation for the hypothesis about the target system obtained by collecting evidence from the source system goes down when increasing the degree of certainty about the similarity between the source and the target system. We then develop an alternative Bayesian model for analogical inference and show that the direction of the variation of the degree of confirmation always coincides with the direction of the degree of certainty about the similarity between the two systems in this model.

# Geddes, Alexander

## Biological Individuality and Pluralism

Recent work in the philosophy of biology distinguishes between evolutionary and physiological accounts of biological individuality. Often, accounts of each sort are thought to have some theoretical utility, but to conflict in their verdicts concerning which organic entities and pluralities are to be counted as biological individuals. And so it has become de rigueur to respond to this conflict by endorsing some form of pluralism.

One consequence of this, as typically understood, is the existence of multiple, closely overlapping biological entities corresponding to what we might otherwise have taken to be a single organism. Where you are, for example, there will in fact be both an evolutionary individual and a physiological individual, differing in their composition and nature.

In the present paper, I argue that this response is a mistake, and that the insights of evolutionary and physiological accounts of biological individuality can in principle be harmonised without accepting pluralism of this sort. The key idea is that each sort of account can fruitfully be read as speaking, in the first instance, to distinct sets of questions, with the sense of conflict generated by natural but mistaken assumptions about how answers to these sets of questions must relate.

In §1, I introduce evolutionary and physiological accounts of biological individuality, identify some of the central ways in which they are typically thought to conflict, and raise some difficult questions that face those who endorse pluralism in the face of this apparent conflict.

In §2, I distinguish between four sets of questions that one might ask about organisms, but which are often run together. These concern organismality (what it is to be an organism); organismic realisation (the conditions under which a plurality realises an organism); organismic composition (the conditions under which a plurality composes an organism); and organismic parthood (the conditions under which something is part of an organism.)

In §3, I focus on some specific evolutionary and physiological accounts of biological individuality in order make the case that, whereas the former are primarily concerned with questions of organismality and organismic realisation, the latter are primarily concerned with questions of organismic composition and organismic parthood. And I argue that something akin to a fallacy of composition or division is responsible for the conviction — shared by those on both sides of this debate—that (for example) an evolutionary account's claims concerning organismality must inevitably generate a tension with with a physiological account's claims concerning organismic parthood. Finally, I make the case the the compatiblity of these accounts' central aspects ought not, in fact, to be considered surprising.

I close, in §4, with a brief consideration of the import that this might have for some wider issues concerning biological individuality. These include: the status of the properties of being an 'evolutionary individual' and a 'physiological individual' in general; conceptions of biological individuality as a functional/multiply-realizable property; whether there are degrees of biological individuality; and whether there are weaker forms of pluralism that ought still to be accepted.

#### References:

Clarke, Ellen (2013). "The Multiple Realizability of Biological Individuals", The Journal of Philosophy 110 (8): 413–435.

DiFrisco, James (2019). "Kinds of Biological Individuals: Sortals, Projectibility, and Selection", British Journal for the Philosophy of Science, 70 (3): 845-875.

Godfrey-Smith, Peter (2009). Darwinian Populations and Natural Selection. (New York: Oxford University Press.)

Pradeu, Thomas (2012). The Limits of the Self: Immunology and Biological Identity. (New York: Oxford University Press.)

Sterner, Beckett (2015). "Pathways to pluralism about biological individuality" Biology and Philosophy, 30 (5): 609-628.

## Guo, Bixin

### Two Approaches to Reduction: A Case Study from Statistical Mechanics

The relation between thermodynamics and statistical mechanics is one of the most paradigmatic instances of reduction: When one attempts to develop an account of reduction and needs an example to demonstrate how exactly that account works, the reduction of thermodynamics to statistical mechanics is the canonical case to which one appeals. However, it is in fact questionable whether, and in what sense, thermodynamics can be reduced to statistical mechanics. Worse, it is not even clear what the correct framework or axiomatic foundation of statistical mechanics is. Instead of assuming that we have a clear grasp of the reduction relation between thermodynamics and statistical mechanics and using that as a paradigmatic case to understand reduction, I propose to approach the problem from a different direction: I argue that there are two distinct approaches to understanding reduction—what I call the ontology-first approach and the theory-first approach. Furthermore, I argue for the significance of this distinction by demonstrating that either one or the other approach has been taken as an implicit assumption in, and has in fact shaped, our understanding of what statistical mechanics is. More specifically, I argue that the Boltzmannian framework of statistical mechanics (BSM) assumes and relies on the ontology-first approach, whereas the Gibbsian framework (GSM) should assume the theory-first approach.

I first explicate the distinction between the ontology-first and the theory-first approach. Reduction is a relation. However, there is no univocal understanding of what its relata should be. Sometimes reduction is taken to be a relation between objects at two levels (say, chlorine gas and molecular chlorine), and sometimes a relation between two scientific theories. The ontology-first and the theory-first approaches concern exactly how these two kinds of reduction—ontological and inter-theoretic—are related. According to the ontologyfirst approach, inter-theoretic reduction (if there is any) depends on and follows from ontological reduction. The theory-first approach takes it to be the other way around.

I then briefly introduce the essential elements of BSM and GSM. Roughly, they offer different descriptions of the same physical system at the statistical-mechanical level. In particular, they differ in whether such descriptions should involve probability.

To show that BSM assumes and relies on the ontology-first approach, I analyze how the key Boltzmannian concepts and justifications directly appeal to ontological reduction between thermodynamic and statistical-mechanical systems. Moreover, I argue that the Boltzmannian criticisms of GSM for its use of ensemble, probability, and coarse-graining, and in particular the Boltzmannian insistence that statistical mechanics should be about actual individual systems, assume the ontology-first approach.

Last, I argue that GSM would be immune to those Boltzmannian criticisms if the theory-first approach is adopted. This approach (1) permits the Gibbsian use of probability, (2) provides a way for GSM to justify its choice of coarse-graining, and (3) allows a broader understanding of statistical mechanics than is conceived of by the ontology-first approach, and this broader understanding is more congenial to GSM than to BSM.

#### References:

Bickle, J. (1996). New wave psychophysical reductionism and the methodological caveats. Philosophy and Phenomenological Research 56(1), 57–78.

Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy. The Journal of Philosophy 96(7), 348–373.

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics. In D. Rickles (Ed.), The Ashgate Companion to Contemporary Philosophy of Physics, pp. 99–196. Ashgate.

Goldstein, S. (2019). Individualist and Ensemblist Approaches to the Foundations of Statistical Mechanics. The Monist 102(4), 439–457.

Wallace, D. (2020). The necessity of Gibbsian statistical mechanics. In V. Allori (Ed.), Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature. World Scientific Publishing Co.

# Hilligardt, Hannah

## The democratic legitimacy of particular interests in science

The aim of this talk is to explore whether it is democratically legitimate for scientists to represent particular interests. It is a traditional concern of philosophers of science to explore the epistemic effect of such influences of interests, as they arguably bias researchers' judgements and undermine scientific credibility. Many commentators have shown that this is not necessarily the case and that, epistemically speaking, interests need not negatively affect scientific research. My concern, however, is whether it is politically (that is, democratically) legitimate for scientists to work in the interest of specific groups rather than for the public in toto. What political mandate do scientists have; what mandate should they have?

According to a popular view, research is democratically legitimate if it serves the interest of the public or serves public aims, as Kitcher (2011) argues and others (Douglas, Intemann, Schroeder, Alexandrova) echo. I show how their scepticism of particular interests relates to theories of deliberative democracy. Rolin (2021) argues against this view that, due of the difficulties in legitimately determining what the public interest is, space must be given to scientific/intellectual movements and advocacy scholarship. She holds such research explicitly represents particular interests and has the potential to correct for epistemic and hermeneutical distortions resulting from societal power discrepancies.

I develop Rolin's position from a political perspective, and theories of representation in particular. In principle, an abstract public cannot speak for itself: it always has to be represented. Therefore, I assume that if scientists work in the interest of the public, they can be said to represent it and be assessed accordingly; likewise, if they work in the interest of particular groups, they represent their interests.

Political philosophers like Pitkin (1972), Young (2000), Rosanvallon, and Brown argue that legitimate representation involves mechanisms of authorisation and mechanisms that hold representatives accountable, the latter including channels of public contestation. This is particularly important for minorities and underrepresented groups. In our current political system, advocacy and interest groups are important tools to do so which is why the explicit representation of particular interest can be considered a necessary element of the finding of a common good. This supports Rolin's point that if scientists are to represent the interest of the public, there need to be appropriate channels of contestation. To this end, some research should represent particular interests.

I illustrate my argument at the hand of a recent case study from the Netherlands: a conflict around the health risks posed by the emissions of a steel factory in Wijk aan Zee, owned by the Tata group. In this case, local interest groups contested scientists' formulations of research questions and their weighting of inductive risks. The interest groups commissioned their own research, increasing pressure on public scientists to issue a statement that confirmed a causal relation between the factory's emissions and locals' health complaints.

References:

Kitcher, Philip (2011): Science in a democratic society. Amherst, N.Y.: Prometheus.

Pitkin, Hanna Fenichel (1972): The concept of representation. Berkeley, London: University of California Press.

Rolin, Kristina H. (2021): Objectivity, trust and social responsibility. In Synthese 199 (1-2), pp. 513–533. DOI: 10.1007/s11229-020-02669-1.

Young, Iris Marion (2000): Inclusion and democracy. New York: Oxford University Press.

# Jaksland, Rasmus and Salimkhani, Kian

#### The many problems of spacetime emergence in quantum gravity

That spacetime may be non-fundamental according to theories of quantum gravity is widely debated in the philosophy of quantum gravity as 'the problem of spacetime emergence' (e.g., Huggett and Wüthrich (2013); Le Bihan (2019); Baron (2020); Baron and Le Bihan (2021)). We contest, however, that the umbrella term 'spacetime' is well-defined and that the notion 'the problem of spacetime emergence' picks out a well-defined philosophical problem that can be dealt with in the abstract, i.e., without further specification.

According to the literature, 'the problem of spacetime emergence' either seems to denote (a) some general or overarching philosophical problem associated with all theories of quantum gravity, or (b) some exceptional philosophical problem which stands out among the philosophical problems raised in this context. This paper argues that both versions of the issue are ill-posed.

More concretely, we defend the following five related claims: (1) the notion of spacetime is ambiguous and underdetermines the exact spatiotemporal aspects under consideration; (2) therefore, what is often discussed under the umbrella term 'spacetime emergence' in fact consists of a plethora of distinct and even highly different issues that are tied to the emergence of specific spatiotemporal aspects; (3) since these various specific problems do not seem to share any common core problem, they defy an investigation in the abstract -- in short, there is no general problem of spacetime emergence; (4) there is also none among the specific problems that seems to stand out as an exceptional problem; (5) instead, spacetime emergence talk and talk about the problem of spacetime emergence is simply unspecific talk that harms the debate. We advocate to discontinue this abstract talk of spacetime emergence and propose to cast such debates more specifically in terms of the emergent spatiotemporal aspects.

We start by observing that spacetime emergence talk in physics rarely suffers from the ambiguity that is often found in philosophical discussions. We then argue that the notion of spacetime (even in general relativity) is ambiguous. In trying to clear up the ambiguity, we offer five ways of conceptualising spacetime, each coming with its own understanding of when the (alleged) problem of spacetime emergence arises. We consider each of these understandings more closely with respect to the question of whether any general or exceptional problem arises that might warrant a treatment of it in the abstract. We find that the list of philosophical problems relating to the emergence of spatiotemporal aspects is long, but no argument currently on offer shows that these problems have a common core or that any one problem stands out among them. Thereare many problems of spacetime emergence, and each of them must be considered in its own specificity.

Importantly, our points are also relevant for those who explore the implications of "spacetime emergence in quantum gravity" for general philosophical accounts. Assessing whether, say, Humeanism is compatible with quantum gravity requires specifying what spatiotemporal aspects are taken to be emergent.

Our proposal that also philosophical debates should take place at the level of spatiotemporal aspects serves two purposes. First, it makes manifest that there are many

interrelated, yet distinct, problems associated with 'spacetime emergence'. Second, it indicates that these problems are to be resolved by detailed (philosophical) analysis of the particular emergent features and not through general or abstract discussions of how spacetime as such can emerge from the non-spatiotemporal. Since the various spatiotemporal aspects inflict very different philosophical problems, insisting on specificity is not an exercise in pedantry but key to the resolution of these problems.

## Jäntgen, Ina

# How to measure effect sizes for rational decision-making: the dominance of absolute outcome measures

In empirical studies testing the effectiveness of treatments, the collected trial data is analyzed using outcome measures. These measures describe how the treatment and the outcome relate and are usually seen as measuring the effect size of the treatment. Such outcome measures provide information for choices between treatments. However, in the case of binary variables, two classes of measures – absolute and relative ones – can drastically disagree on a treatment's effect size. For instance, the Heart Protection Study tested the effectiveness of cholesterol-lowering drugs to prevent heart attacks. The study found a relative risk reduction of 12.2 % of death. By contrast, the absolute risk reduction was 1.8 %. Only the former effect size was reported, as is common in biomedical research. Yet, the difference in claimed effect size is striking. Which effect size is more informative for someone aiming to decide on taking the drug?

In this talk, I argue that from a decision-theoretic and an epistemic perspective, absolute measures are equally good or better than relative ones for informing choices across several circumstances. I start by modelling two choice scenarios using decision theory, one involving outcome measures from a single trial and another involving outcome measures from several trials. Based on decision theory, Sprenger & Stegenga (2017) argue that only absolute measures provide sufficient probabilistic information for rationally choosing between treatments. I show that this argument only applies to choices involving outcome measures from a single trial. When instead choosing between treatments tested in distinct trials, absolute measures only have such a decision-guiding advantage if we know that the probabilities of the outcome absent treatment are equal without knowing these probabilities. Otherwise, both classes of measures fare equally well in providing decision-relevant probabilistic information, contrary to Sprenger & Stegenga's (2017) view. Still, from a decision-theoretic perspective, absolute measures dominate relative ones.

Furthermore, absolute measures remain equally good or better than relative ones even for inferences from effect sizes measured in trials to ones relevant for choosing amongst treatments. Using Bayesian epistemology, I model such inferences as involving a causal inference, an extrapolation and a particularization. I argue that neither relative nor absolute measures have an advantage for figuring in causal inferences or extrapolations, challenging an argument by Broadbent (2013). Moreover, Stegenga (2018) suggests that relative measures but not absolute ones contribute to agents committing base rate fallacies in particularization inferences. I show under which conditions relative measures indeed have such a disadvantage, and when they do not. Overall, I conclude that absolute measures epistemically and decision-theoretically dominate relative ones.

This dominance argument challenges the current practice to only report relative measures. At the same time, I show under what conditions both absolute and relative measures provide sufficient probabilistic information for treatment choices, and when only absolute ones do so. Exposing these conditions allows for a more detailed discussion on the merits of different outcome measures for guiding choices. I conclude my talk by deriving principles for reporting effect sizes from these conditions, to be scrutinized in further work. References:

Broadbent, A. (2013). Philosophy of Epidemiology. Palgrave Macmillan. https://doi.org/10.1057/9781137315601

Sprenger, J., & Stegenga, J. (2017). Three Arguments for Absolute Outcome Measures. Philosophy of Science, 84(5), 840–852. https://doi.org/10.1086/693930

Stegenga, J. (2015). Measuring effectiveness. Studies in History and Philosophy of Biological and Biomedical Sciences, 54, 62–71. https://doi.org/10.1016/j.shpsc.2015.06.003

Stegenga, J. (2018). Medical Nihilism. Oxford University Press. https://doi.org/10.1093/oso/9780198747048.001.0001

## Lakhany, Farhan

#### Evolution & Experience

One way of making sense of experiential states is via an evolutionary explanation. To do this, we need to clearly explicate the adaptive role which experiential states play in the organisms which have it. In this works, Peter Godfrey-Smith provides two different ways of understanding experiential states as evolutionary adaptions and calls the two views the "transformation view" and the "latecomer view". The transformation view understands experiential states as a relatively early evolutionary adaptation and the latecomer view understands it as a relatively late evolutionary adaptation. Godfrey-Smith adopts the transformation view, and, in this article, I defend the latecomer view. This paper is broken up into six sections. In the first section, I provide an overview of the transformation view and the latecomer view. In the second section, I provide a more detailed analysis of the latecomer view, highlight its connection to the global workspace theory of consciousness and provide an argument for it that I call the argument from non-necessity. To explicate the global workspace theory of consciousness, I highlight work done by Bernard Baars, Stanislas Dehaene, Lionel Naccache and Jean-Pierre Changeux. In the third section, I provide a response on behalf of Godfrey-Smith to my argument and provide two arguments for why we should prefer the transformation view to the latecomer view. To bolster Godfrey-Smith's position, I turn to Derek Denton's work on "primordial emotions" and the two arguments raised are called the argument against special pleading and the argument from the logic of evolutionary explanation. In the fourth section, I argue against the transformation view by claiming that, unlike the latecomer view, it does not provide an adaptive role for experiential states. In the fifth section, I attempt to provide, on behalf of Godfrey-Smith and the transformation view, an adaptive role for the transformation view but argue that the response fails. To argue against the adaptive role, I turn to work done by JD Rose and others on whether fish can, in fact, feel pain or whether they are simply engaging in nociceptive behavior. In the sixth section, I respond to the two arguments articulated in the third section by showing how they can be overcome. To respond to the argument from special pleading, I focus specifically on non-reflexive behavior and the role it plays in Godfrey-Smith's argument and to respond to the argument from the logic of evolutionary explanation, I focus on how aspects of a global neuronal workspace might gradually evolve over time. Finally, I conclude by discussing the upshot of the arguments presented in the article and how an advocate of the transformation view could respond to the worries presented.

#### **References:**

Dehaene, Stanislas. Consciousness and the Brain Deciphering How the Brain Codes Our Thoughts. Viking, 2014.

Denton, Derek, et al. "The Role of Primordial Emotions in the Evolutionary Origin of Consciousness." Consciousness and Cognition, vol. 18, no. 2, 2009, pp. 500–514., doi:10.1016/j.concog.2008.06.009.

Godfrey-Smith, Peter. Other Minds: the Octopus and the Evolution of Intelligent Life. William Collins, 2018.

Rose, J D, et al. "Can Fish Really Feel Pain?" Fish and Fisheries, vol. 15, no. 1, 2012, pp. 97– 133., doi:10.1111/faf.12010.

## Landström, Karl

## Research funding as domination: Epistemic Freedom and Resistance

In March 2021, UK Research and Innovation (UKRI), a leading research funding body in the UK announced a substantial reduction in its international research development budget following the UK government's cut to its overseas development aid budget (UKRI 2021). The £1.5 billion Global Challenges Research Fund (GCRF) was heavily affected and research projects which had been awarded funding but were yet to start were cancelled and an estimated 800 research projects that were already undergoing had to face the consequences of the funding reduction in the form of either termination or reprofiling. GCRF funding has since been partially restored to some of those projects, but these decisions will undoubtedly have lasting effects on UK researchers, research institutions and their international partners (Imperiale & Phipps 2022).

The interference with projects who had already been awarded funding sparked plenty of debate pertaining to the ethics of research funding and academic freedom, a debate to which this paper aspires to contribute. In this paper I advance the view that it is not the interference itself that is the core issue, but rather the funder's ability to interfere. This view is inspired by neo-republican accounts of freedom as put forth by scholars such as Quentin Skinner (1998) and Philip Pettit (1997). I draw on the work of the decolonial theorist Sabelo Ndlovu-Gatsheni (2018) and the neo-republican epistemology of James Bohman (2012) to argue that certain forms of research funding constitute domination. That is, in such funding schemes the funder has the ability to arbitrarily interfere in the funded research projects, and arbitrarily assign duties and responsibilities to the recipients of funding without there being any meaningful possibility for contestation. Thus, such funding schemes limits the freedom of the recipients of funding from those schemes. In this paper I discuss the Global Challenges Research Fund as an example of such a research funding scheme. I argue for four theses: (I) Certain forms of research funding constitute domination. (II) Such domination risks producing significant negative ethical, epistemic and practical consequences, particularly when it reproduces colonial hierarchies. However, (III) despite the dominating position inhabited by the funder, there is still space and possibility for agency and resistance on the part of the dominated. Lastly, (IV) I argue that despite the agency and resistance on the part of the dominated, these funding schemes nonetheless constitute domination. I conclude by considering some aspects of what a non-dominating funding scheme would entail. Thus, this paper hopefully adds to the growing critical literature on the Global Challenges Research Fund, but also to wider discussions pertaining to the ethics of research funding, research policy and governance, academic freedom and to wider debates pertaining to the decolonisation of knowledge and epistemic justice.

#### References:

Bird, Alexander & Ladyman, James (2013). Free Inquiry: The Haldane Principle and the Significance of Scientific Research. Social Epistemology Review and Reply Collective, 2 (7) 14-22

James Bohman (2012). Domination, Epistemic Injustice and Republican Epistemology. Social Epistemology, 26:2, 175-187

Imperiale, Maria Grazia. & Phipps, Alison (2022). Cuts destroy, hurt, kill: a critical metaphor analysis of the response of UK academics to the UK overseas aid budget funding cuts. Journal of Multicultural Discourses, 1-17

Ndlovu-Gatsheni, Sabelo (2018). Epistemic Freedom in Africa: Deprovincialization & Decolonization. Oxon: Routledge

# Lari, Teemu

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

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

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

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

I argue that the tension arises from conflating the "first-level" scientific discussion of an epistemic community with the "second-level" discussion about that epistemic community and its role in the broader societal context. In the first-level scientific discussion, one must indeed take as given the overall goals and other core commitments of the epistemic community. The norm [B] is appropriate only at this level. However, as [A] requires, there must be forums in which to discuss second-level questions like "what kind of knowledge should a certain discipline strive for?" The second-level discussion requires a revised norm [B\*], according to which the relevance of criticism may be established by appealing to considerations like values or policy needs that may even conflict with the established goals of the community. The second-level discussion also needs to consider the institutional context such as the division of intellectual labor between epistemic communities and the way these communities exert influence in the society. I illustrate this refinement of Longino's account by explaining what values, policy needs, and aspects of the institutional context would be relevant in an appropriate debate between mainstream economists and their feminist critics.

References:

Doss, C. (2021). Diffusion and Dilution: The Power and Perils of Integrating Feminist Perspectives Into Household Economics. Feminist Economics, 27(3), 1–20.

Fourcade, M., Ollion, E., & Algan, Y. (2015). The Superiority of Economists. Journal of Economic Perspectives, 29(1), 89–114.

Hirschman, D., & Berman, E. P. (2014). Do Economists Make Policies? On the Political Effects of Economics. Socio-Economic Review, 12(4), 779–811.

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

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

# Letrud, Kåre and Johnsen, Svein Åge Kjøs

# Do we agree on the essentials? Surveying philosophers' paradigms of pseudoscience

There is allegedly a substantial extensional consensus 'on most particular issues of demarcation' (Hansson, 2021). Philosophers and scientists alike draw the demarcation line at the same place, separating between sciences and pseudosciences in very much the same way (Mahner, 2013, pp. 30-31). This agreement is purportedly the product of a shared 'ready recognition', a tacit skill for spotting a pseudoscience when they see it (Hansson, 2013, p. 61; Pigliucci & Boudry, 2013, p. 2). The extensional consensus justifies the continued philosophical work on pseudoscience (Hansson, 2021), and gives cause for optimism on behalf of the pseudoscience demarcation project: The process is a matter of '...offering philosophical justification to decisions that have already been made.' (Fasce, 2020, p. 165). However, the evidence for the existence of an extensional consensus appears anecdotal.

In this paper, we survey philosophers on what are clear and strong examples of pseudosciences, i.e. paradigmatic pseudosciences. We present philosophers with 40 oft cited pseudosciences and ask them a) whether they would classify them as pseudoscientific, b) whether those classified as pseudoscientific are clearly and unambiguously pseudoscientific, and c) how strongly they hold them to be pseudoscientific.

We have presently received answers from 280 respondents. 169 has a PhD in philosophy 48 has an MA, 8 a BA. With a few exceptions, the rest have PhDs in other fields. A majority of the respondents with a PhD in philosophy classified 35 of the cases on the list as pseudosciences. However, only a minority classified the 35 cases as unambiguous and clear examples of pseudoscience.

The results reveal no agreement among the philosophers asked that these 40 pseudosciences are paradigmatic. We argue that the justification for the pseudoscience demarcation project, as well as the process based on this agreement, may need reassessment.

#### References:

Fasce, A. (2020). Are Pseudosciences Like Seagulls? A Discriminant Metacriterion Facilitates the Solution of the Demarcation Problem. International Studies in the Philosophy of Science, 32(3-4), 155-175. doi:10.1080/02698595.2020.1767891

Hansson, S. O. (2013). Defining Peudoscience and Science. In M. Pigliucci & M. Boudry (Eds.), Philosophy of Pseudoscience. Reconsidering the demarcation problem (pp. 61-78). Chicago: Chicago University Press.

Hansson, S. O. (2021, 20 May). Science and Pseudo-Science. Stanford Encyclopedia of Philosophy. Retrieved from http://plato.stanford.edu/entries/pseudo-science/

Mahner, M. (2013). Science and pseudoscience. How to demarcate after the (alleged) demise of the demarcation problem. In M. Pigliucci & M. Boudry (Eds.), Philosophy of

Pseudoscience. Reconsidering the demarcation problem (pp. 29-44). Chicago: Chicago University Press.

Pigliucci, M., & Boudry, M. (Eds.). (2013). Philosophy of Pseudoscience: Reconsidering the Demarcation Problem. Chicago and London: The University of Chicago Press.

## Lopez, Luis G.

#### Understanding with Deep Neural Networks

The deployment of Machine Learning (ML) models in scientific research has shown that they can make accurate predictions in domains where traditional models or simulations have failed to do so (e.g., AlphaFold 2 and the protein-folding problem). However, science is not just about prediction; it is also about understanding. In this talk, I address the following questions: can ML models provide understanding of phenomena? If so, how? And, more importantly, what is the nature and reach of that understanding? I argue that the answers to these questions depend on whether these models are interpretable or opaque. Here, I follow the distinction made and defended by Rudin et al. (2022). Namely, while an interpretable ML model "obeys a domain-specific set of constraints to allow it to be more easily understood by humans," an opaque ML model is a "formula that is either too complicated for any human to understand, or proprietary" (ibid.). (I am not concerned with proprietary black boxes in this contribution). I show that this distinction has implications not only for understanding the model itself (as it directly follows from Rudin's definitions) but for the understanding of its target phenomenon. To illustrate my point, I focus on Deep Neural Networks (DNNs) – which are the guintessential black box– and compare them with traditional models. In addition to opaque DNNs, I consider two kinds of interpretable ML models: 'disentangled' DNNs and approaches that combine Deep Learning with Symbolic Regression. Through these comparisons, I show that the explanatory work of these models is done by the hypotheses they provide (in the case of interpretable ML models) or by the hypotheses they are in part based on (in the case of traditional models). To make this clearer, I draw on a classification of scientific hypotheses -based on their explanatory depth- made by Bunge (1997). Namely, black box, grey box, and translucent box hypotheses. I argue that while interpretable DNNs can provide black box hypotheses (i.e., those that answer questions of the "what is it?" type), opaque DNNs cannot do so. This contrasts with the built-in grey and translucent box hypotheses (i.e., those that answer questions of the "how does it work?" type) of semi-phenomenological and mechanistic models, respectively. Nonetheless, I show how interpretable DNNs can aid both semiphenomenological and mechanistic explanations. In addition, I discuss to what extent one can extract black box hypotheses from opaque DNNs with post hoc XAI methods (not to be confused with interpretable ML). Overall, this contribution shows that low levels of 'link uncertainty,' as defined by Sullivan (2019), and transparency at the highest-level of implementation (i.e., structural transparency according to Creel, 2020) may be sufficient for prediction but not for understanding; for the latter, we also need functional transparency (as defined by Chirimuuta, 2020, and Creel, 2020). Finally, I argue that the account of (objectual) understanding defended by Dellsén (2020) –i.e., understanding as dependency modelling- is the account of scientific understanding that better accommodates the kind of understanding provided by interpretable ML models.

#### References:

Bunge, M. (1997). Mechanism and explanation. Philosophy of the Social Sciences, 27(4), 410–465.

Chirimuuta, M. (2020). Prediction versus understanding in computationally enhanced neuroscience. Synthese, 1–24.

Creel, K. A. (2020). Transparency in complex computational systems. Philosophy of Science, 87(4), 568-589.

Dellsén, F. (2020). Beyond Explanation: Understanding as Dependency Modelling. The British Journal for the Philosophy of Science, 71(4), 1261–1286.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistics Surveys, 16, 1-85.

Sullivan, E. (2019). Understanding from machine learning models. The British Journal for the Philosophy of Science. (https://doi.org/10.1093/bjps/axz035).

## Makins, Nicholas

#### Doctors, patient, and risk attitudes

A lively topic of debate in decision theory over recent years concerns the rationality of different risk attitudes exhibited by decision-makers. Numerous philosophers have debated the merits and flaws of various 'precautionary principles', which have become especially pertinent in response to the COVID-19 pandemic (Greenhalgh et al. 2020), and there is much ongoing discussion about the rationality of risk-averse and risk-seeking behaviours more broadly (Buchak 2013; Pettigrew 2015; Briggs 2016; Stefánsson and Bradley 2020; Thoma 2021b). In medical contexts, this matter is complicated by the fact that physicians must often make choices for the benefit of their patients, but the norms of rational choice are conventionally grounded in a decision maker's own attitudes, dispositions, and behaviour. The presence of both clinician and patient in medical contexts raises the question of whose risk attitude matters for the choice at hand and what to do when these diverge. Must doctors make risky choices when treating risk-seeking patients? Ought they to be risk-averse in general when choosing on behalf of others? Or should they remain risk-neutral, in order to secure the best aggregate outcome in the long-run?

There is a nascent literature on permissible risk attitudes in other-regarding choices (Bovens 2019; Blessenohl 2020; Scharding 2021; Thoma 2021a; Zhao 2021). Johanna Thoma (2021a) divides the views on this matter into three categories: a permissive view, according to which any rationally permissible risk attitude can be adopted when choosing for another; a specific requirement view, according to which there is a particular risk attitude that one ought to adopt when choosing for others; and a deferential view, according to which one ought to adopt the risk attitude of the individual on behalf of whom the choice is being made. In this paper, I argue that the deferential view comes closest to getting it right in the context of clinical medicine. I demonstrate how the arguments that underlie widely held antipaternalistic views about medical decision-making can be extended to support antipaternalism about attitudes to risk just as well. Medical professionals have reason to respect not only the basic values and preferences of their patients, but also their attitudes to risk (insofar as these are distinct). However, this view is in need of further refinement. Physicians ought not simply to defer to whatever risk attitudes their patient actually has, but to what risk attitude they would endorse. This amendment to the deferential view is crucial in order to justify my claims to those who conceive of risk attitudes as distinct from an agents desires (e.g. Buchak 2013) and to avoid some potential counterexamples. Lastly, I draw out some implications of the view I propose for health policy, medical practice, and law. While the view defended is demanding, it is inescapable for those concerned to promote patient-centred medical decision-making.

#### References:

Blessenohl, Simon. 2020. "Risk Attitudes and Social Choice." Ethics 130 (4): 485–513.

Bovens, Luc. 2019. "The Ethics of Making Risky Decisions for Others." In Oxford Handbook of Ethics and Economics, edited by Mark D. White, 446–73. Oxford: Oxford University Press.

Briggs, Rachael. 2016. "Costs of Abandoning the Sure-Thing Principle." Canadian Journal of Philosophy 45 (5–6): 827–40.

Buchak, Lara. 2013. Risk and Rationality. Oxford: Oxford University Press.

Greenhalgh, Trisha, Manuel B Schmid, Thomas Czypionka, Dirk Bassler, and Laurence Gruer. 2020. "Face Masks for the Public during the Covid-19 Crisis." BMJ 369 (m1435): 1–4.

Pettigrew, Richard. 2015. "Risk, Rationality and Expected Utility Theory." Canadian Journal of Philosophy 45 (5–6): 798–826.

Stefánsson, Orri H., and Bradley, Richard. 2020. "What Is Risk Aversion?" British Journal for the Philosophy of Science 70 (1): 77–102.

Thoma, Johanna. 2021a. "Taking Risks on Behalf of Another." Manuscript.

----. 2021b. "Time for Caution." Philosophy & Public Affairs.

Zhao, Michael. 2021. "Ignore Risk; Maximize Expected Moral Value." Noûs, 1–18.

## Malik, Uzma

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

Robert Batterman argues that his 'minimal' models often provide better understanding than more accurate detailed ones. I will explain minimal models using Batterman's own example in hydrodynamics (Batterman 2009/2001) to illustrate. I then develop an account of understanding--- a 'local unification' account—that fits Batterman's claims, which I shall argue, serves global and local intuitions about understanding better than available alternatives.

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

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

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

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

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

References:

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

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

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

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

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

## Matarese, Vera

# The epistemic status of atmospheric retrieval models in exoplanetary science

Exoplanetary science provides an exciting frontier for philosophy of science, since its practices are still novel and fluid. This talk focuses on retrieval exoplanet atmospheric models, which are indispensable tools to overcome the lack of knowledge about exoplanets' atmospheres (Madhusudhan 2018), but whose reliability is still under discussion (Barstow et al. 2020). I investigate whether such models are mere heuristic tools that explore different options of atmospheric profiles, or whether they provide epistemic justification for inferring the constitution of exoplanets' atmosphere.

Retrieval methods aim to infer the composition of an exoplanet's atmosphere by exploring a wide range of possible atmosphere models and evaluating which one yields the best fit with the exoplanet's transmission spectrum. The first step of these models is to make a guess of the initial exoplanet's atmospheric state, feed it into forward models derived from first principles, and obtain a synthetic spectrum. The second step consists in comparing the synthetic spectrum with the transmission spectrum and evaluating, in most cases with Bayesian statistics, whether the two spectra match. In case of a negative answer, the initial state gets modified and the resulting synthetic spectrum gets compared again through an iterative process until the best match is found.

When evaluating the alethic justification of retrieval models, different empirical and theoretical considerations come into play. First of all, I consider the relation of retrieval models to theory and data. While they are regarded as data-based models, retrieval models typically include forward models and so should rather be regarded as hybrid models. I argue that thanks to their hybrid nature, retrieval models are more reliable than data-based or theory-based models because they are jointly constrained by theoretical elements as well as the data, and inherit their epistemic justification through both top-down and bottom-up approaches. The theories grounding forward models are a guarantee for the veridicality of the models, while the data, which have some sort of representational content, provide the models with an empirical basis.

Second, I evaluate the following tension. On the one hand, since the ultimate goal of these models is to reveal the composition of the exoplanets' atmosphere, it seems compelling to assign them a representational function in terms of mapping to states of affairs. On the other hand, since the retrieval methods start with an initial guess of the state of the atmosphere, it seems that retrieval models play a merely exploratory and heuristic role. I shift the attention to the function of these models of carving out the space of possible combinations of atmospheric conditions that could generate the transmission spectrum under consideration. I argue that since this function delivers a modal knowledge about the target system, retrieval models are truth-conducive.

Thirdly, I propose to regard retrieval models as perspectival models (Massimi 2018) in order to solve the problem of incompatible models from which retrieval models suffer, as different retrieval models may provide different atmospheric compositions for the same data (Barstow et al. 2020). Different retrieval models should be regarded as complementary rather than contradictory, as each model provides an incomplete representation of the

target system given its level of idealization. However, I also stress some differences from Massimi's perspectival realism, as in my view, retrieval models carry a normative epistemic weight.

References:

Barstow, J. K., Changeat, Q., Garland, R., Line, M. R., Rocchetto, M., & Waldmann, I. P. (2020). A comparison of exoplanet spectroscopic retrieval tools. Monthly Notices of the Royal Astronomical Society, 493(4), 4884-4909.

Barstow, J. K., & Heng, K. (2020). Outstanding Challenges of Exoplanet Atmospheric Retrievals. Space science reviews, 216(5), 1-25.

Madhusudhan, N. (2018). Atmospheric retrieval of exoplanets. doi, 10, 978-3.

Massimi, M. (2018). Perspectival modeling. Philosophy of Science, 85(3), 335-359.

# Maziarz, Mariusz and Serpico, Davide

## N-of-1 Studies as a Valuable Source of Evidence for Precision Medicine

Conflicting results of clinical trials are widely considered a major problem for medicine as they undermine reliable inference and clinical decision-making. Under the strain of empirical literature reporting heterogeneous treatment effects, researchers have pointed at the need of precision/personalised medicine approach to improve diagnosis and treatment by accounting for the variation among patients (Lemoine 2017; Plutynski 2020). In our talk, we approach the problem of treatment effect heterogeneity across individuals and argue that gene-environment interactions (G×E) undermine the results of the standard repertoire of precision medicine, e.g., genome-wide association studies (GWAS) and big data research of electronic health records. We then support the use of pragmatic N-of-1 studies to inform therapeutic decisions. The structure of our talk is as follows.

First, we explain that inconsistencies in response to treatment can depend on the genetic heterogeneity of populations. We focus on the case of asthma, a complex disease characterised by chronic airway inflammation (Global Initiative for Asthma 2019): a variety of pharmacogenetics studies have pointed out the involvement of G×E where the absence/presence of specific genetic variants determine different responses to drugs (Kersten & Koppelman 2017 for a review).

Second, we consider RCTs assessing the efficacy of asthma pharmaceutical therapies (inhaled corticosteroids, leukotriene receptor antagonists) to explain how conflicting results can emerge due to G×E: since RCTs inclusion and exclusion criteria are not based on genetic variability or biomarkers, they usually overlook G×E and population stratification (i.e., the existence of subgroups that differ in genetic terms and thus respond differently to drugs). When major G×E are present, running effective RCTs ideally needs subtyping the population in a way that tracks down actual biological differences; however, this requires (often missing) prior knowledge from GWAS and candidate-gene studies on what specific G×E can affect treatment responses. Given these limitations, most RCTs can only identify effective treatments for the average patient.

Third, we argue that, due to the problems with evidence stemming from RCTs, the evidencebased approach to treatment can be beneficial for patients sufficiently resembling the average participant in clinical trials, but not outliers. While sub-group analysis has proved successful in some cases (Blunt 2019), it is susceptible to reporting spurious findings due to multiple hypothesis testing. We analyse the Global Initiative for Asthma (2019) recommendations and show that, while it acknowledges different treatment responses to standard therapies (e.g., inhaled corticosteroids, p. 52), it relies on the so-called one-sizefits-all approach: patients with poorly controlled symptoms are advised to receive a nextstep treatment.

Given that the standard evidence sources cannot account for inter-patient heterogeneity, we argue that clinical practice should rely (to a greater extent) on pragmatic N-of-1 trials to choose the most appropriate therapy for each patient. Due to their very nature, N-of-1 studies avoid stratification issues entirely and allow to investigate more effectively individual-level response to treatments, bypassing confounding factors like G×E and other

sources of individual differences. We thus believe that this approach can represent a further step forward on the path to more precise pharmaceutical treatments.

#### References

Blunt, C. J. (2019). The Dismal Disease: Temozolomide and the Interaction of Evidence. Available at SSRN 3444926.

Global Initiative for Asthma (2019). Global Strategy for Asthma Management and Prevention. Retrieved from: https://ginasthma.org/wp-content/uploads/2019/06/GINA-2019-main-report-June-2019-wms.pdf. Access: 26th February 2022.

Kersten, E. T., & Koppelman, G. H. (2017). Pharmacogenetics of asthma: toward precision medicine. Current opinion in pulmonary medicine, 23(1), 12-20.

Lemoine, M. (2017). Neither from words, nor from visions: understanding p-medicine from innovative treatments. Lato Sensu: Revue de la Société de philosophie des sciences, 4(2).

Plutynski, A. (2020). Why precision oncology is not very precise (and why this should not surprise us). Philosophical Issues in Precision Medicine.

# McCoy, Ryan

#### The Role of Local Knowledge in Understanding Climate Change

Tension between expert and local knowledge has taken a central role in recent debates concerning the observability of climate and climate change. Given that climate is understood in its scientific sense as an average of variables typically over a 30 year period and developed by rather complicated modeling, some scholars have argued in favor of invisibilism, or the view that climate change is undetectable from the standpoint of local observation (Swim, et al. 2009; Hulme 2009). Moreover, work in experimental psychology has indicated that local observations of climate change have tended to be biased, and there is concern that giving credence to lay observations of climate and climate change will result in evaluations that are as mercurial as the weather (Rudiak-Gould 2013). In contrast, visibilist scholars have argued that the effects of climate change can in fact be locally observed. The rationale for this position has been that stakeholders in indigenous and local communities with prolonged engagement with their environment possess the knowledge and acumen to detect climatic changes at a local level.

In this paper I argue for a qualified version of visibilism that addresses invisibilist concerns, and conclude that local knowledge can importantly contribute to our understanding of climate and climate change. As evidence for this claim, I draw on recent work undertaken by climate researchers in Spain at the Local Indicators of Climate Change Impacts (LICCI) project (Reyes-García, et al. 2020), as well as researchers working with indigenous communities in Australia and New Zealand (Green, et al. 2010, King, et al. 2008). I argue that this research shows that local knowledge can significantly contribute to our understanding of climate and climate change given that (1) local knowledge can provide data that is much more fine grain than is capable for regional climate modeling, (2) this data can be used to track and address impacts on biophysical systems, and (3) local knowledge can fill spatial and temporal gaps in instrumental climatic data.

Lastly, in my conclusion I explore the non-epistemic consequences of including local knowledge within our understanding of climate and climate change. I argue that citizen science initiatives like those undertaken at the LICCI importantly engage non-experts in the production of scientific knowledge. A consequence of this engagement is that it can both facilitate public trust in science, as well as improve our scientific understanding.

#### References:

Green D, Billy J, Tapim A (2010). "Indigenous Australians' knowledge of weather and climate," Climate Change 100:337–354.

Hulme, M., (2009). Why We Disagree about Climate Change: Understanding Controversy, Inaction and Opportunity. Cambridge University Press.

King DNT, Skipper A, Tawhai WB (2008). "Māori environmental knowledge of local weather and climate change in Aotearoa–New Zealand," Climate Change 90(4):385.

Reyes-García V., Fernández-Llamazares Á., García-del-Amo D., Cabeza M. (2020) Operationalizing Local Ecological Knowledge in Climate Change Research: Challenges and Opportunities of Citizen Science. In: Welch-Devine M., Sourdril A., Burke B. (eds) Changing Climate, Changing Worlds. Ethnobiology. Springer, Cham.

Rudiak-Gould P. (2013). "We Have Seen It with Our Own Eyes": Why We Disagree about Climate Change Visibility. Weather Climate and Society: 5:120–132.

Swim J, Clayton T, Doherty R, Gifford G, Howard J, Stern P, Weber E. (2009). Psychology and Global Climate Change: Addressing a Multifaceted Phenomenon and Set of Challenges. Pennsylvania: APA Task Force on the Interface between Psychology and Global Climate Change.

## Meyer, Russell

### An Explanatory Taste For Mechanisms

Accounts of scientific explanation typically lay out explanatory standards: the sorts of things a description needs to provide in order to become a bona fide explanation. A prominent example of this is Mechanism, which provides a set of explanatory standards that require that explanations completely and accurately capture the causal entities and activities underlying phenomena (Bechtel & Abrahamsen 2005, Craver 2007, Glennan 2017). These standards, according to one prominent version of Mechanism, are objective (Craver 2007, 2014). This is because, so the story goes, mechanistic standards of explanation have been developed in order to accurately capture objective explanations – the really existing mechanisms out in the world. This leads to the claim I call the closure of explanatory standards: mechanistic standards of explanation are insulated from judgements and decisions by investigators about what explanatory standards are appropriate, since determining the correct explanatory standards is "...independent of psychological questions about the kinds of explanation that human cognitive agents tend to produce or tend to accept...people in different cultures might have different criteria for accepting or rejecting explanations. These facts... are not relevant to the philosophical problem of stating when a scientific explanation ought to be accepted as such." (Craver 2014 pg. 29)

I raise two problems with this position, and suggest an alternative and more ecumenical way of thinking about explanatory standards. First, Craver's account relies on several ontological claims about mechanisms which, while plausible, fail to guarantee the objectivity of mechanistic explanatory standards to the degree of certainty required. Second, Craver's position itself introduces a value–laden explanatory standard—the 3M requirement (Kaplan & Craver 2011)—which undermines the closure of explanatory standards by revealing how mechanism's standards are selected for their usefulness towards certain investigative ends, rather than because an accurate description of nature requires them.

I show how in practice, the selection of Mechanism's explanatory standards is guided by explanatory taste, shorthand for researchers' interest in particular kinds of questions about nature, which in turn drives the form of the answers – explanations – being sought. Mechanism, for instance, often demonstrates a pragmatically-oriented taste for control over phenomena, and gerrymanders explanatory standards in order to obtain it. I argue that this is a normal and productive feature of science, and is not to the detriment of Mechanism or any other account. I conclude by arguing that objectivity, rather than being obtained through an objectively correct set of explanatory standards, is better thought of as the result of a process of intersubjective criticism, which render visible the contextual values of communities of investigators and allow them to be controlled for (Longino 1990).

#### **References:**

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36(2), 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010

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

Craver, C. F. (2014). The Ontic Account of Scientific Explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), Explanation in the Special Sciences: The Case of Biology and History (pp. 27–52). Springer Verlag.

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

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective\*. Philosophy of Science, 78(4), 601–627. https://doi.org/10.1086/661755

Longino, H. (1990). Science as Social Knowledge: Values and Objectivity in Scientific Inquiry (Vol. 25). Princeton University Press.

# Michelini, Matteo

## Epistemic Vigilance as a Driver of the Emergence of Truthful Communication -- an Agent-based modeling approach

The stability and evolution of honest communication have always represented a puzzle for anthropologists, as our communication system seems to be prone to the limitless spread of misinformation. Sperber et al. (2010) attempted to solve this puzzle on the basis of the available experimental evidence on testimony and communication. They claimed that humans developed a suite of cognitive mechanisms, which they call epistemic vigilance, in order to keep communication advantageous, forcing the speaker to be honest (Michaelian, 2013; Sperber, 2013). As Michaelson (2018) pointed out, epistemic vigilance relies on four different epistemic tools: a past record of the listener's experiences with a given speaker, a social assessment of the speaker's reliability (i.e. public reputation), a consistency check of the new information against the listener's previous beliefs and the ability of the listener to catch non-verbal cues that the speaker may produce. In the account by Sperber et al. (2010), the combination of these four is claimed to be necessary and sufficient for communication to be honest.

In this paper, we study the impact of these tools through the use of an agent-based model. Our contribution is threefold. First of all, our model is the first that approaches the problem of human honest communication from the perspective of indirect reciprocity dynamics (Giardini et al., 2021; Nowak and Sigmund, 1998). This also allows us to situate the epistemic vigilance hypothesis in the broader area of formal evolutionary theory. Secondly, our results are a simulated test for Sperber et al. (2010)'s theoretical intuitions, that epistemic vigilance is sufficient and necessary for honest communication. Finally, a nuanced quantitative perspective on the effectiveness of these four epistemic tools is provided, by studying their assessment in specific social conditions.

We build our model as an enhanced version of the models of indirect reciprocity (Giardini et al., 2021; Nowak and Sigmund, 1998; Ohtsuki and Iwasa, 2004). In such models, agents interact and cooperators may perform altruistic acts towards others. However, defectors may take advantage of cooperative behavior without giving anything in return. Thus, agents use reputation and gossip to ensure defectors do not proliferate. In our version, the agents exchange information, and defectors may fake cooperation, by lying. For this reason, in addition to keeping track of personal and public records of other agents' reputations (two of Sperber et al. (2010)'s tools), agents may be equipped with more specific abilities. They can detect a lie with a certain probability, which depends on how much information an agent already has (beliefs consistency check) and they have a probability to recognize a liar as they listen (spotting behavioral cues).

Our results confirm Sperber et al. (2010)'s hypothesis for most of the plausible sets of parameters, i.e. those that resemble real-world conditions for communication. Indeed, in these cases, the winning strategies are those that rely on truthful communication. Yet, our results also identify under which conditions epistemic vigilance does not generate honest communication. For example, this is the case when agents have either collected too much or too little information in the first phase of a population era. Furthermore, we find that the strength of epistemic vigilance lies in the combination of public reputation and belief

consistency check. The likelihood of producing honest communication with both of these tools is significantly higher than when only one of them is active, even if combined with other tools (spotting behavioral cues or personal records).

#### References:

Giardini, Francesca, Dan Balliet, Eleanor Power, Szabolcs Szamado, and Karoly Takacs (2021). "Four Puzzles of Reputation-Based Cooperation: Content, Process, Honesty and Structure". In: Human nature-An interdisciplinary biosocial perspective. issn: 1045-6767.

Michaelian, Kourken (2013). "The Evolution of Testimony: Receiver Vigilance, Speaker Honesty and the Reliability of Communication". In: Episteme 10.1, pp. 37–59.

Michaelson, Eliot (Nov. 2018). "Lying, Testimony, and Epistemic Vigilance". English. In: The Oxford Handbook of Lying. Oxford University Press, pp. 214–228. isbn: 9780198736578.

Nowak, Martin and Karl Sigmund (July 1998). "Evolution of Indirect Reciprocity by Image Scoring". In: Nature 393, pp. 573–7.

Ohtsuki, Hisashi and Yoh Iwasa (2004). "How should we define goodness?—reputation dynamics in indirect reciprocity". In: Journal of Theoretical Biology 231.1, pp. 107–120. issn: 0022-5193.

Sperber, Dan (Mar. 2013). "Speakers are honest because hearers are vigilant. Reply to Kourken Michaelian". In: Episteme 10. Sperber, Dan, Fabrice Clement, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson (2010). "Epistemic Vigilance". In: Mind & Language 25.4, pp. 359–393.

## Muller, Sean

### On the Nature and Purpose of Peer Review

Modern academic philosophy relies, much like many other academic disciplines within the humanities, social sciences and natural sciences, on peer review by a small number of reviewers in order to determine whether a given analysis should be published or not. Some empirical evidence raises concerns about the efficacy of peer review. For example, there are many accounts of papers being rejected in one or more credible outlets, only to be published in another credible outlet and subsequently being deemed to have made an important contribution to their topic. At minimum such examples, which appear across a wide range of disciplines, suggest inefficiency in the system of peer review. Yet the rejection of work that is subsequently considered epistemically worthy (without substantive changes) also raises the concern that the system regularly produces what may be deemed 'incorrect' decisions. Outside of contexts where triple-blind review is utilised, an increasing number of studies have claimed to find network and prestige effects in peer-reviewed journals suggesting various forms of bias (Haffar, Bazerbachi, and Murad 2019). In some disciplines and outlets, concerns about such dynamics have led to efforts to amend and improve review processes (Smith 2006). More extreme cases also exist where peer review systems have been substantially compromised by outrightly corrupt and unethical conduct (Biagioli and Lippman 2020).

In this broad context, it has recently been argued that the traditional approach to peer review should be entirely reconsidered. One contribution has argued that 'prepublication peer review should be abolished' (Heesen and Bright 2021), while a related contribution attempts to make the case for 'crowdsourced' post-publication peer review (Arvan, Bright, and Heesen 2022). The latter appeals to the Condorcet jury theorem to establish a claim that such a process would produce epistemically superior outcomes.

The present analysis has two main objectives. First, we show that the propositions of recent contributions are flawed in as much as the results rest on significant omissions. Among these are the influence of prestige and networks, but the more fundamental problem arises from how those studies conceive of the purpose of the peer review process. We illustrate this concern by relating the stance taken in those analyses to Kuhn's characterisation of normal and revolutionary science (Kuhn 1962). Second, we develop an alternative framework for thinking about the dynamics and outcomes of peer review. In particular, one that allows for an association between the nature of ideas and the manner in which they are received (adjudicated) by peers. In that framework the structure of the academic community is such that 'peer review by jury' can produce worse outcomes than traditional peer review. Furthermore, we argue that these effects will vary across epistemic contexts. It follows that practical adoption of such proposals carries risk of real epistemic harm. Given this, there is a burden of proof on advocates of those positions that has not yet been met. Our analysis concludes by considering how extant proposals might be adapted to mitigate these concerns and emphasises the importance of maximally-blinded review processes.
References:

Arvan, Marcus, Liam Kofi Bright, and Remco Heesen. 2022. 'Jury Theorems for Peer Review'. The British Journal for the Philosophy of Science, January, 719117. https://doi.org/10.1086/719117.

Biagioli, Mario, and Alexandra Lippman, eds. 2020. Gaming the Metrics: Misconduct and Manipulation in Academic Research. Infrastructures. Cambridge, MA: MIT Press.

Haffar, Samir, Fateh Bazerbachi, and M. Hassan Murad. 2019. 'Peer Review Bias: A Critical Review'. Mayo Clinic Proceedings 94 (4): 670–76. https://doi.org/10.1016/j.mayocp.2018.09.004.

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.

Kuhn, Thomas S. (1962) 1996. The Structure of Scientific Revolutions. 3rd ed. Chicago: University of Chicago Press.

Smith, R. 2006. 'Peer Review: A Flawed Process at the Heart of Science and Journals'. Journal of the Royal Society of Medicine 99 (4): 178–82. https://doi.org/10.1258/jrsm.99.4.178.

# Ney, Milan

## Dyamic Monism for Biological Function

Function has been a central topic in the philosophy of biology for more than half a century. Since the 1960s, various 'theories of function' have been proposed. Discussions between proponents of these views often end up turning on nothing more than the "dull thud of conflicting intuitions" (Bigelow and Pagetter (1987: 196). Is digging, for example, a function of turtle flippers if it is important to their reproduction even if they haven't been selected for it? Usage in the scientific community doesn't seem to settle such questions and intuitions differ. Thus, various alternatives have been proposed to the project of finding the one correct definition of \_function\_ for biology. Most philosophers of biology now endorse some form of pluralism. They accept that \_function\_ is used with different senses within biology and that different theories of function account for different senses. Yet, many of the same issues repeat themselves. Is there a legitimate theory of function that excludes turtle flipper's digging?

I propose an alternative approach: Dynamic Monism. It is a meta-theory of function, i.e. a theory about the nature of theories of function. I draw on the work of Peter Ludlow (\_Living Words\_, 2014) on the Dynamic Lexicon and the underdetermination of meaning. I argue that there is one skeletal 'core' meaning of \_function\_ as used within the life sciences. This skeletal meaning is largely constituted by a few canonical judgments, such as 'hearts have the function of pumping blood'. It allows function to be applied to new cases through implicit or explicit analogy.

The skeletal meaning is heavily underdetermined. For many cases, it does not determine whether 'function' applies. Hence, scientists implicitly and explicitly sharpen it. They do so, primarily, by adding new canonical judgments, such as 'making a thumping noise is (/is not) a function of hearts'. In different contexts, the meaning is sharpened differently. While reducing the underdetermination of \_function\_'s meaning, such sharpenings never completely eliminate it. There will always be cases that cannot be decided based on the agreed-upon canonical cases.

Theories of function, like Cummins-style accounts, selected-effect accounts and organisational accounts of function, are also ways of modulating the meaning of \_function\_. Rather than introducing new canonical judgments, they introduce an explicit rule according to which \_function\_ is to be applied. Thus, they further reduce its underdetermination.

Theories of function serve various roles. They allow us to categorise different ways of modulating the meaning of \_function\_ by adding canonical judgments. We group them according to which theories they are compatible with. They further clarify disagreements in which scientists use \_function\_ with differently modulated meanings. Finally, they allow us to justify certain practices in science. For example, the selected-effect account vindicates the practice of explaining the existence of traits by appealing to their functions.

I contrast Dynamic Monism with other meta-theories of function implicit in the literature: Hard Monism, Fuzzy Monism, Hierarchical Monism, Between-Discipline Pluralism and Between-Discipline Pluralism. I argue that only Dynamic Monism can account for the diversity in which function is used in successful biological practice, and avoid making prescriptions that are likely to hinder the future development of \_function\_-discourse in biology.

References:

Bigelow, J., & Pargetter, R. (1987). Functions. \_The Journal of Philosophy, 84\_(4), 181–196

Cusimano, S., & Sterner, B. (2019). Integrative pluralism for biological function. \_Biology & Philosophy\_, 34(6), 1-21.

Doolittle, W. F. (2013). Is junk DNA bunk? A critique of ENCODE. \_Proceedings of the National Academy of Sciences\_, 110(14), 5294-5300.

Ludlow, P. (2014). \_Living Words: Meaning Underdetermination and the Dynamic Lexicon\_. Oxford: Oxford: Oxford University Press.

Neander, K. (2017). Functional analysis and the species design. \_Synthese, 194\_(4), 1147-1168.

# Nissan-Rozen, Ittay and Liron, Amir

## Triangulation, Incommensurability, and Conditionalization

According to Methodological Triangulation there is an epistemic value in using several methods to establish the same scientific claim. Hessen et al. (2016) presented three formal results that can be viewed as expressing two different senses for the term "epistemic value" in the above characterization. The first sense is that of independent reliability: when several reliable methods support the same claim independently, the overall (objective) probability of the claim to be true increases with the number of methods. The second sense is that of reliability likelihood: when at least two methods, with unknown levels of reliability, agree, the probability that some of these methods are reliable increases with the number of methods.

After discussing some limitations of these two senses, we suggest a third one that applies to cases in which different methods rely on partly incommensurable theories, that nevertheless – give predictions regarding the same phenomena. We model such cases using different probability distributions that are defined over partly overlapping algebras. We assume that although all theories assign a probability value to some hypothesis, H, none of the theories assign a probability value to the claim that one of the other theories predicts H with a given probability. We argue that this nicely captures the situation in a wide range of interdisciplinary scientific discussions. We also argue that at least in many such cases, it is natural to demand that after learning that a method, associated with a given theory, supports H to a given degree, a rational agent should not change her conditional credence in the reliability of other methods, that are associated with other theories, given H. We show that this last assumption amounts to a commitment to using Jeffrey's conditionalization as an updating method and that when all methods are reliable, rational credence in H increases with the number of methods used. We also provide a general formula for computing this credence and show that the order of using different methods does not matter.

We then demonstrate how our account can be applied to the case of using different methods to gain knowledge about casual mechanisms. We do that by exploring the example of the research regarding gendered pathways to women's incarceration. The relevant scientific literature identifies three different pathways to women's incarceration. However, some of the measures (e.g., substance use) used in the research regarding each one of these pathways are endogenous to the pathways. Thus, it is not always possible to assess the exogenous effect of one pathway given another one. We argue that this should be understood as a type of partial incommensurability between the theories that describe each one of the pathways and so our account can capture this case. Specifically, we argue that even though no statistical evidence regarding the effect of one pathway given another is available, the fact that a given woman follow more than one pathway should increase one's confidence in her future incarceration. We point to some implications to rehabilitation policy.

References:

Heesen, R., Bright, L.K. & Zucker, A. Vindicating methodological triangulation. Synthese 196, 3067–3081 (2019)

Kuorikoski, J., & Marchionni, C. (2016). Evidential Diversity and the Triangulation of Phenomena. Philosophy of Science, 83(2), 227-247.

Salisbury, Emily J., and Patricia Van Voorhis. "Gendered pathways: A quantitative investigation of women probationers' paths to incarceration." Criminal justice and behavior 36.6 (2009): 541-566.

# Pain, Ross

## Stone Tools, Predictive Processing, and the Evolution of Language

A recent development in archaeology is the sub-field of experimental archaeology. Experimental archaeologists perform experiments in the present and use them to project inferences to the past. In the discipline of cognitive archaeology, researchers use various neuroimaging techniques to identify which areas of the brain are co-opted by toolmaking tasks. This technique has revealed considerable neural overlap between the areas of the brain co-opted by Early Stone Age toolmaking and those recruited during the production of language (e.g. Stout et al. 2021). Such results lend weight to tool-language co-evolutionary hypotheses.

This paper has two parts. In the first part I use a prediction error minimisation model (Hohwy 2013) to add computational detail to these findings. I show that, when applied to the actions required to produce Late Acheulean tools, an error minimisation framework reveals that complex structured representations--that is, representations with nested part-whole relations--were present in the Early Stone Age. As representations of this kind are thought to be an important step in the emergence of syntax, it is plausible that toolmaking facilitated the development of cognitive structures that were co-opted and refined in the evolution of language. An error minimisation model thus demonstrates an information processing overlap between toolmaking and language, which complements the neural overlap revealed by experimental archaeology.

In the second part of the paper, I situate the above proposal within the literature on language evolution. An important issue is the evolution of structured representations. On some accounts, structured representations are treated as a necessary condition for language production. Consequently, explaining the evolution of language requires positing an evolutionary transition from cognitive systems without structured representations to cognitive systems with structured representations. The transition might be sudden (e.g. Berwick and Chomsky 2016) or gradual (e.g. Planer and Sterelny 2021). This is typically achieved by positing an associated transition from sequential information processing to hierarchical information processing. Others emphasise evolutionary continuity, and want to avoid a commitment to transitions, either between non-structured and structured representations or between sequential and hierarchical processing. They thus deny that structured representations are required for language production (e.g. Frank, Bod, and Christiansen 2012). The error minimisation account occupies a unique place within this dialectic, because the transition between non-structured representations and structured representations can occur without positing a transition between sequential and hierarchical processing. This is because the former transition is made possible by increasing the number and sophistication of layered models in a cognitive system, which nonetheless remains governed by the principle of error minimisation. I argue that this feature makes the account an attractive choice for those attempting to produce gradualist explanations of language evolution.

In sum, I motivate two claims. First: the error minimisation approach demonstrates that producing the tools of the Early Stone Age required sophisticated structured representations. Second: the error minimisation approach can account for the evolution of

structured representations without needing to posit a corresponding transition in types of processing.

References:

Berwick, Robert C., and Noam Chomsky. 2016. Why Only Us: Language and Evolution. MIT Press.

Frank, Stefan L., Rens Bod, and Morten H. Christiansen. 2012. 'How Hierarchical Is Language Use?' Proceedings of the Royal Society B: Biological Sciences 279 (1747): 4522–31. https://doi.org/10.1098/rspb.2012.1741.

Hohwy, Jakob. 2013. The Predictive Mind. Oxford, New York: Oxford University Press.

Planer, Ronald J., and Kim Sterelny. 2021. From Signal to Symbol: The Evolution of Language. Cambridge, MA, US: MIT Press.

Stout, Dietrich, Thierry Chaminade, Jan Apel, Ali Shafti, and A. Aldo Faisal. 2021. 'The Measurement, Evolution, and Neural Representation of Action Grammars of Human Behavior'. Scientific Reports 11 (1): 13720. https://doi.org/10.1038/s41598-021-92992-5.

# Piotrowska, Monika

## A Depressed Mouse Is a Depressed Mouse Is a Depressed Mouse: A Critique of the Commitment to Standardization in Behavioral Neuroscience

We study suffering because alleviating it is good, and inflicting it is wrong. Yet we can't study suffering in an animal that can't suffer, and we can't study suffering without inflicting it. -Garner 2020, 82

Depression is such a common disorder that it is the leading cause of disability worldwide. Treatments are often ineffective, and the development of new, more effective antidepressants is slowed by the fact that roughly 90% of compounds that seem to work on animals fail to work on humans (Garner 2014). In other words, a big problem for developing more effective treatments for depression is the difficulty of extrapolating from seemingly effective treatments for animals to effective treatments for human beings. One obstacle to effective extrapolation is the fact that animal research on depression tends to be done almost exclusively on male animals (Beery and Zucker 2011; Shansky 2019) but depression occurs much more frequently among women than men—by an almost 2:1 ratio. In the face of this fact, the easy solution would seem to be improving the male-female sex ratio in animal research. But such a remedial step does very little when the problem itself arises from a certain type of theoretical commitment—vis., the commitment to standardization. It is because scientists need standardized animals to control variation that they end up focusing on, for example, only one sex. It is this theoretical commitment to standardization that has resulted in female subjects being excluded from animal research (their estrus cycle is a potential source of variability) and it is the drive toward this commitment that I aim to explore. More concretely, I will argue that seemingly reasonable choices about what to standardize in behavioral neuroscience research on depression often hinder extrapolation from rodents to humans.

The presentation will begin with a brief overview of the difficulties involved in modeling depression in rodents. These difficulties help explain why behavioral neuroscience research on depression is highly standardized. I'll then explain how standardizing behavioral tests for depression, the subjects of these experimental tests, and the testing environment is hurting extrapolation efforts. I'll end by considering a shift in perspective that might help us identify unreasonable standardization choices.

References:

Beery, A. K., & Zucker, I. (2011). Sex bias in neuroscience and biomedical research. Neuroscience and biobehavioral reviews, 35(3), 565–572. https://doi.org/10.1016/j.neubiorev.2010.07.002

Garner J. P. (2014). The significance of meaning: why do over 90% of behavioral neuroscience results fail to translate to humans, and what can we do to fix it?. ILAR journal, 55(3), 438–456. https://doi.org/10.1093/ilar/ilu047

Garner. J.P. (2020). The Mouse in the Room. In: Principles of Animal Research Ethics. Edited by: Tom L. Beauchamps and David Degrazia, Oxford University Press.

Shansky, R.M. (2019). Are hormones a "female problem" for animal research? Science 364 (6443): 825-6.

Würbel, H. (2000). Behaviour and the standardization fallacy. Nature Genetics 26, 263. https://doi.org/10.1038/81541

# Sassoon, Judyth

### What is a plesiosaur?

Philosophy of paleontology is a new discipline exploring how palaeontologists gain knowledge of the past from fossils (Currie, 2019). To date it has mainly focused on epistemic questions, such as the incompleteness of the fossil record and the scientific strategies that mitigate those limitations (Cleland, 2002; Turner 2011). What has not been discussed in sufficient detail is the actual process of fossil interpretation, especially the influence of ideas brought to fossil specimens by the palaeontologists themselves. Since the beginning of palaeontology as a science, palaeontologists have worked within generally accepted conceptual frameworks that provided the backdrop for their interpretation of specimens (Rudwick, 2008). T.D. Johnston (2021) called these frameworks "theories of relatedness". A theory of relatedness makes certain assumptions and influences the way comparative data is interpreted. At the same time, empirical comparisons themselves can also shape theories of relatedness. The meanings assigned to fossils arise out of this reciprocity. Here, I will show how such dynamics have modified the conceptualisation and meaning of one group of extinct animals: the plesiosaurs.

Plesiosaurs were marine reptiles that evolved from a terrestrial ancestor in the early Mesozoic Era. Fossils show that plesiosaurs had a unique four-finned body "type" which has not been repeated by any animal group following their extinction. Since their 'rebirth' in 1821 as scientific concepts and palaeontological icons, plesiosaur fossils have been studied and interpreted against a background of slightly different "theories of relatedness". The interpretations of plesiosaur specimens emerged from the referencing of comparative observations to such theories, and it is also possible to find examples of how data from plesiosaur fossils shaped the overarching theories of relatedness themselves. Here I will present several conceptualisations of the meaning of "plesiosaur", as it emerged from the reciprocity between empirically derived comparative observations and theoretical backdrops. I discuss three theories of relatedness within which the idea of the plesiosaur has been conceptualised in slightly different ways: (1) natural theology, (2) vertebrate archetype (3) phylogenetic.

In the 19th Century, the first detailed plesiosaur description was made by William Conybeare who coined the Greek name pleios sauros ("near reptile"). This name itself reflected the way plesiosaurs were conceptualised within a natural theological theory of relatedness. At the same time, Richard Owen was working on his own theory of relatedness, which emerged as the vertebrate archetype. Empirical date from a plesiosaur specimen helped Owen to form his ideas in 1840. Owen's theory was a peculiar mixture of natural theology, continental Naturphilosophie and the comparative anatomist's need for a general, experimentally derived reference "schema". The Darwinian revolution of 1859 opened a new vista in palaeontology, enabling fossil workers to situate specimens into ancestral lineages. It brought with it new possibilities to explain the two broad morphologies adopted by the plesiosaur body "type": plesiosauromorphs (small heads and long necks) and pliosauromorphs (large heads and short necks). These examples will be used here to highlight the reciprocal dynamic between comparative observations and the overarching theories of relatedness in the process by which palaeontologists make sense of fossils.

#### References:

Cleland, C.E. (2002) Methodological and epistemic differences between historical science and experimental science. Philosophy of Science 69, 447-451. doi.org/10.1007/s10539-019-9682-2

Currie, A. (2019) Scientific Knowledge and the Deep Past: History Matters. Elements in the Philosophy of Science. Northcott R. & Stegenga, J. (eds) Cambridge: Cambridge University Press. doi.10.1017/9781108582490

Johnston, T.D. (2021) The pre-Darwinian history of the comparative method, 1555-1855. History and Philosophy of the Life Sciences. Nov;43,118. doi: 10.1007/s40656-021-00474-8.

Rudwick, M.J.S (2008) The meaning of fossils: Episodes in the history of palaeontology. University of Chicago Press.

Turner, D. (2011) Paleontology: A Philosophical Introduction. Cambridge University Press.

# Segarra, Adrià

## A Hybrid Theory of Induction

A theory of induction is an account of inductive support (Hawthorne, 2018), and approaches to induction can be roughly divided into formal and material. According to formal theories of induction, inductive inferences are warranted by rules, like those of Bayesianism, which allow us to articulate measures of inductive support. Rules of induction, however, are unable to warrant all and only the good inductive inferences (Norton (2005); Lipton (2004, ch. 1)). According to material theories of induction (e.g. Brandom, 2000; Norton, 2021), good inductive inferences are not warranted by rules but by matters of fact. Norton (2021), in particular, convincingly argues that inductive inferences are warranted by local facts about the matter of the induction. Material theories, however, do not offer accounts of inductive support (Peden, 2019, pp. 677-678). As a consequence, material theories are not full-fledged theories of induction but (narrower) theories of inductive warrant.

In this talk I present a Hybrid Theory of Induction (HTI). This theory is hybrid because it tells us that both rules and local facts play a role in understanding inductive support. In developing the HTI I articulate an account of inductive support that takes on board Norton's lessons on inductive warrant. According to the HTI, inductive inferences follow rules, and rules are warranted by local matters of fact. This view acknowledges that we need rules of induction in order to understand and articulate measures of inductive support, and we need local material facts in order to define the domains of application of those rules and the degree to which they are warranted in a context. Thus, the warranting role is reserved for material facts, but both rules and matters of fact play a role in determining the degree of inductive support of an inference. In this talk I will expand on the role of rules and matters of fact within the HTI and the connections between them.

The HTI has several advantages over other approaches to induction. Crucially, it addresses the aforementioned challenges to both formal and material theories of induction. Furthermore, I argue that the HTI captures a view that is widespread in contemporary and classic literature, although not explicitly stated. Making this view explicit can help resolve some current debates. Most formal theorists, for example, acknowledge that the rules they defend are not universal, and are only appropriate under certain circumstances. At the other end of the spectrum, Norton also acknowledges that, although the warranting work is done by facts, rules are valid within restricted domains. Thus, the apparent divide between formal and material theorists can be bridged if we realise that their views are compatible and complementary. The apparent disagreement between them is only a difference in focus, with both sides exploring the role of one component (rules or facts) in understanding induction, while implicitly acknowledging the presence of the other. The HTI articulates this shared underlying view, addresses the main challenges to both formal and material theories and explains away the disagreement between defenders of both approaches.

#### **References:**

Brandom, R. (2000). Articulating Reasons: An Introduction to Inferentialism. Harvard University Press.

Hawthorne, J. (2018). Inductive Logic. In E. N. Zalta (Ed.), The stanford encyclopedia of philosophy (Spring 2018). Metaphysics Research Lab, Stanford University.

Lipton, P. (2004). Inference to the best explanation (2nd ed.). London, UK: Routledge.

Norton, J. D. (2005). A Little Survey of Induction. In P. Achinstein (Ed.), Scientific evidence: Philosophical theories and applications (pp. 9–34). Baltimore: Johns Hopkins University Press.

Norton, J. D. (2021). The Material Theory of Induction. BSPS Open. Calgary, Alberta: University of Calgary Press.

Peden, W. (2019). Direct Inference in the Material Theory of Induction. Philosophy of Science, 86(4), 672–695.

# Sequeiros, Sofia Blanco

# Discordant Evidence, Evidential Reasoning, and Scientific Practice

In this paper, I analyze the concept of discordant scientific evidence and its epistemic and methodological implications for scientific reasoning and inference. The main thesis of this paper is that evidential discordance is a problem for scientific inference and practice especially in decision-making contexts, and that discordance develops into scientific evidence when the inferential and evidential relationships between data and phenomena are misconstrued during the evidence-generating process.

The argument proceeds as follows. First, I define the concept of discordant evidence. At its most general, evidential discordance refers to the situation where evidence both confirms and disconfirms the same scientific hypothesis, theory or claim (Stegenga 2009, 2012; Hey 2015). This can mean either a body of evidence that provides support both for and against a hypothesis, or individual pieces of evidence that contradict each other. I focus on persistent discordance, i.e. discordance that cannot be explained with a typo in a spreadsheet, or the like. My analysis meant to complement the more formal accounts of evidential discordance and amalgamation (e.g. Claveau 2013, Landes 2020).

I argue that discordance introduces uncertainty into decision-making, which then decreases its efficiency and its likelihood of success. Discordance can be fruitful for scientists in the sense that disagreement provides fertile ground for new hypotheses and discoveries, but I argue that with respect to decision-making, discordance in evidence presents a significant issue. It generates uncertainty within further scientific inference, practice, and evidential reasoning.

Next, I use the distinction between data and phenomena (Bogen and Woodward, 1988) to explain how discordance develops into scientific evidence during the evidence-generating process. For Bogen and Woodward, data are local, context-dependent observations, recorded as data points in specific research settings and under certain conditions. Claims about phenomena, on the other hand, are nonlocal. I propose that discordance is the outcome of a process of generating evidence where the evidential and inferential relationships between data and phenomena are misconstrued. Data are taken to be nonlocal when they are local, or vice versa; or the scope of claims about phenomena is understood as more nonlocal or local than it is. I also use the concept of enriched evidence (Boyd, 2018) to show how uncertainty, and thereby discordance, enter scientific evidence during the evidence-generating process.

Last, I illuminate my conceptual analysis with a case example from economics of crime: studies on the causal relationship between the length of a prison sentence and recidivism. I show how epistemic uncertainty about the relationship between a prison sentence and the likelihood of committing a new crime has meant that the relationship between data and the phenomena that the data is meant to track has been misconstrued. This has created a discordant body of evidence on the effects of prison sentences on recidivism. I conclude with a discussion on how discordance and its effects for scientific inference and evidential reasoning can be mitigated.

References:

Bogen, J., & Woodward, J. (1988). Saving the phenomena. The philosophical review, 97(3), 303-352.

Boyd, N. M. (2018). Evidence enriched. Philosophy of Science, 85(3), 403-421.

Claveau, F. (2013). The independence condition in the variety-of-evidence thesis. Philosophy of Science, 80(1), 94-118.

Hey, S. P. (2015). "Robust and discordant evidence: Methodological lessons from clinical research". Philosophy of Science, 82(1), 55-75.

Landes, J. (2020). Variety of evidence. Erkenntnis, 85(1), 183-223.

Stegenga J. (2012) "Rerum Concordia Discors: Robustness and Discordant Multimodal Evidence". In: Soler L., Trizio E., Nickles T., Wimsatt W. (eds) Characterizing the Robustness of Science. Boston Studies in the Philosophy of Science, vol 292. Springer, Dordrecht

Stegenga, J. (2009). "Robustness, discordance, and relevance". Philosophy of Science, 76(5), 650-661.

# Skaf, Rawad El and Palacios, Patricia

## What Can We Learn (and Not Learn) from Thought Experiments in Black Hole Thermodynamics?

Black hole thermodynamics (BHT) is a discipline that combines theoretical statements coming from three main theories: quantum mechanics, general relativity and thermodynamics. Although BHT has attracted a great deal of attention in the last decades, it still lacks direct empirical support. In fact, thought experiments (TE) instead of real experiments have proved to be one of the most important tools for getting insight about the thermal properties of black holes. But to what extent can we trust the results coming from TEs in BHT?, and what are the limitations of the knowledge that can be obtained on the basis of TEs?

Some physicists (e.g. Susskind 2008, Polchinski 2017) have stressed the importance of TEs for bringing to light paradoxes between fundamental theories in the context of black holes, and some philosophers of science have even suggested that black hole TEs can give some theoretical support to the idea that black holes have thermodynamic properties (Curiel 2015, Wüthrich 2019). But despite the essential role that TEs seem to play in BHT, there has been surprisingly little philosophical work on this topic. In fact, most of the philosophical work around BHT has focused either on the main calculations that give theoretical support to the idea that black holes are thermodynamic objects (e.g. Dougherty and Callender 2019, Wallace 2018, 2019) or on the use of analogue experiments, which are real experiments performed in systems different but analogous to black holes (e.g. Crowther et al., 2021; Dardashti et al., 2017, 2019).

In this contribution, we will start filling this gap by systematically analyzing the role of TEs for increasing our confidence in BHT and also by stressing the limitations of TEs for conclusively confirming untested theoretical predictions. More precisely, in the first part of the talk, we will summarize the use of TEs in BHT, focusing on Wheeler's TE and Geroch's engine TE. We will then argue that Wheeler's TE reveals an inconsistency between general relativity's no-hair theorem and the second law of thermodynamics, and we will explain how Bekenstein (1972) resolves this inconsistency by formulating the Generalised Second Law. In contrast, we will argue that Geroch's TEs reveals an inconsistency between the general relativity's redshift and the second law of thermodynamics, both in its Kelvin and in its generalised form. We will then discuss different resolutions offered in the literature, such as Bekenstein's entropy bound (1981) and Unruh and Wald (1982)' buoyancy effect. In the third part, we will analyze the type of evidence coming from Wheeler's and Geroch's TEs by engaging with the philosophical literature on TEs. We will argue, following El Skaf (2021), that the principal epistemic function of these TEs, like many other scientific TEs, is to reveal "external" inconsistencies. Finally, we will contrast the type of evidence coming from TEs with the type of evidence coming from other means such as analogue experiments and mathematical derivations. This will serve to illustrate the importance, but at the same time, the limitations of the use of TEs for learning about BHT.

# Spagnesi, Lorenzo

### Regulative Idealization: A Kantian Approach to Idealized Models

Scientific models, including those used in physics, chemistry, biology, economics, and geology, typically contain idealizations, or assumptions that are known not to be true (Cartwright 1983; Potochnik 2017). Scientists for example, assume frictionless planes, infinitely sized populations, and perfectly rational agents. Ideal assumptions such as these intentionally misrepresent the empirical system that is being studied, and yet they allow for the formulation of theories and laws of nature in a variety of scientific domains. While the benefits of idealizations are undeniable, philosophers of science have reached no consensus on how we can use ideal assumptions to know how things work in the real world (Godfrey-Smith 2009). Some have argued that the inaccuracy of idealizations implies the falsity of scientific laws (e.g. Cartwright 1983). For others, these are heuristic tools that may disappear with the advancement of science (e.g. Wimsatt 1987). A third position is that they are ineliminable, yet felicitous falsehoods that advance our understanding in a way similar to other forms of representations in art and literature (e.g. Elgin 2017).

During the Enlightenment, the philosopher Immanuel Kant developed a normative account of scientific investigation that can inspire a new approach to the contemporary debate. Kant argued that scientific investigation is possible only if guided by ideal assumptions—what he calls regulative ideas. These ideas are not true of any object of nature, and yet they are not mere heuristic tools or fictional representations of reality. They are necessary rules governing the construction and assessment of scientific explanations. To use some of Kant's examples, the ideal assumption of fundamental power is the rule that allows to seek empirical explanations of various powers in psychology; the ideal assumption of organized beings is the rule that allows to seek physical explanations in biology. In short, regulative ideas set the explanatory goals of scientific investigation. As such, they regulate the investigation of what reality is like and significantly contribute to our knowledge of nature.

In this paper, I suggest that Kant's account of regulative ideas can help us reconcile the ubiquity of idealizations in contemporary science with a realist commitment to scientific knowledge. First, I develop a Kantian stance on idealizations that emphasizes the necessary, yet merely normative function of scientific idealizations. In other words, I argue that idealizations are not defined by their descriptive relation to objects, but rather by the goals and standards they set up for empirical investigation. As such, idealizations are to be thought as explanatory ideals that are indispensable to the scientific investigation of nature. In the second part of the paper, I evaluate the benefits of a Kantian stance on scientific idealization on the basis of its capacity to: (i.) find an alternative to heuristic and fictionalist readings of scientific models; (ii.) explain the revisability and progress of scientific inquiry; (iii.) elucidate the shareability of scientific models by an epistemic community.

#### References:

Cartwright, Nancy 1983. How the Laws of Physics Lie, Oxford: Oxford University Press.

Elgin, Catherine Z. 2017. True Enough, Cambridge, MA: MIT Press.

Godfrey-Smith, Peter 2009. "Models and Fictions in Science," Philosophical Studies, 143, 1, 101–116.

Potochnik Angela 2017. Idealization and the Aims of Science. Chicago: University of Chicago Press.

Wimsatt, William C. 1987. "False Models as Means for Truer Theories," in Nitecki M.H. and Hoffman A. (eds.), Neutral Models in Biology, 23–55, New York: Oxford University Press.

# Steinbrink, Lukas Christian

## Social Science Denialism

Although historically somewhat overshadowed by the debate about pseudosciences, the phenomenon of science denialism has nevertheless attracted the attention of philosophers of science. Paradigmatic examples include denialism in the areas of biology (theory of evolution), medicine (HIV/AIDS, vaccines and recently COVID-19), climate science (global warming denialism) and epidemiology (health effects of tobacco consumption). With the notable exception of historiography, where various forms of genocide denialism constitute their own area of research, the examples studied are usually taken from the natural sciences. Analogous to the focus in the debate about the demarcation problem, the social sciences seem to be an afterthought in the philosophical reflection about the phenomenon of science denialism. There are various possible justifications for this: (a.) maybe there is no such thing as social science denialism; (b.) if there is such a thing, then it may not a problem that is deserving of comparable attention to the case of denialism in the natural sciences; and finally, (c.) it might be implicitly assumed that a satisfactory account of science denialism simpliciter, based on the examples taken from the natural sciences, can just be extended to cover the social sciences. In this paper I argue that these three claims are wrong. In order to do this, I first dispel a conceptual worry about 'social science denialism', before looking at specific examples. The conceptual worry can be summarized like this: Denialism requires the existence of a robust consensus within the scientific community that the movement in question can deny. The social sciences have a structure that is characterized by a plurality competing paradigms, methodologies and theoretical perspectives. Therefore, no such robust consensus exists to be denied. The upshot of the rebuttal is that the requirement that the community in question must encompass the whole discipline is too strong and that there are nevertheless important areas of convergence. Second, a look at specific cases reveals that denialism with regard to the social sciences has consequences, which are potentially socially harmful, and that it exhibits certain epistemic features, which – while present in some cases of denialism in the natural sciences – give social science denialism a unique and interesting profile. I emphasize three aspects in particular: (1) a grossly oversimplified understanding of the relevant field; (2) the generalization of superficial criticisms from parts of a discipline to the whole; and (3) a promotion of alternative, blatantly unscientific explanations.

#### References:

Hansson, Sven O. 2017: "Science denial as a form of pseudoscience", in: Studies in History and Philosophy of Science, 63: 39–47.

Levy, Neil 2019: "Due deference to denialism: explaining ordinary people's rejection of established scientific findings", in: Synthese (2019) 196:313–327.

Torcello, Lawrence 2012: "The Trouble with Pseudoskepticism", in: Skeptical Inquirer 36 (3).

Slater, Matthew H. & Huxster, Joanna K. & Bresticker, Julia E. 2019: "Understanding and Trusting Science", in: Journal for General Philosophy of Science (2019) 50:247–261.

# Suárez, Javier and Stencel, Adrian

# The Major Transitions in Coercion: A radical approach to the origins of the biological hierarchy

Life on Earth is hierarchically organized. Genes form genomes, cells form multicellular organisms or (endo)symbiotic assemblages, multicellular organisms associate to form eusocial colonies, etc. A key question in contemporary philosophy of biology concerns the nature of the biological constraints that need to evolve such that previously independent biological units are lumped together to form a higher-level unit in the biological hierarchy. The Major Transitions in Evolution (MTE) project, originally systematized by Maynard-Smith & Szathmary (1995), and substantially altered since (Michod 1999; West et al. 2006; Szathmary 2015; Agren et al. 2019; Durand et al. 2019; Herron 2021), focuses on answering this question. This project is built upon the 'cheater point of view', i.e., the premise that selection at the higher-level will systematically be disrupted by selection at the lower-level due to the selfish behaviours of the lower-level units (Lloyd & Wade 2019). Hence, the origin of the biological hierarchy is subsumed under the question about the set of biological mechanisms that transformed independently living and selfish biological elements (or individuals, if the question concerns evolutionary transitions in individuality, e.g., Michod 1999) into a cooperative unit, whose elements are altruistic to each other because they share a common evolutionary interest (Okasha 2018).

The purpose of this talk is to question the necessity of cooperation in the inner logic of the MTE. We contend that, given that the MTE project concerns the evolution of higher-level forms of organization, then while cooperation may be a way of achieving organization at the higher-level (i.e., it may be sufficient), it not the only way (i.e., it is not necessary). Particularly, drawing on Stencel & Suárez's (2020) work, we argue that coercion may be as effective as cooperation in triggering a MTE. Hence, the role of coercion in the MTE should not be ignored, but rather invigorated, contrary to the reductionistic and misguiding tendency to conceive questions about the MTE exclusively in terms of questions about cooperation. Our goal is thus to transform the question of the MTE from a question about the origins of cooperation into a question about how coercion has triggered the MTEs, and what role it still plays in maintaining the stability of higher-levels of biological organization. We show that thinking the MTE in terms of coercion rather than in terms of cooperation, completely changes the logic of the process, including the type of questions being asked and the type of biological evidence that needs to be found to discover a MTE. Given the relevance that coercion plays in our conception of the MTE, we refer to our project as the Major Transitions in Coercion.

#### References

Durand, P. M. et al. 2019. Cell Death in Evolutionary Transitions in Individuality. Yale J Biol Med. 92(4): 651-662.

Herron M. D. 2021. What are the Major Transitions? Biology & Philosophy 36:2.

Maynard-Smith, John, and Eors Szathmáry. 1995. The Major Transitions in Evolution. Oxford University Press.

Lloyd, Elisabeth A, and Michael J. Wade. 2019. 'Criteria for Holobionts from Community Genetics'. Biological Theory 14 (3): 151–70.

Okasha, Samir. 2018. Agents and Goals in Evolution. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198815082.001.0001.

Stencel, A., Suárez, J. 2021. Do Somatic Cells Really Sacrifice Themselves? Why an Appeal to Coercion May be a Helpful Strategy in Explaining the Evolution of Multicellularity. Biol Theory 16, 102–113

# Sung, Richard D. and Holman, Bennett

### Against evidentiary pluralism in medicine

In a recent issue of the BMJ Evidence-based Medicine (EBM), a letter signed by an interdisciplinary group of 42 scholars urged medical researchers to abandon the priority currently placed on randomized controlled trials (RCTs) in the evaluation of medical evidence. Among the signatories were Nancy Cartwright, Jacob Stegenga, Alex Broadbent, and other notable philosophers of science. This preponderance of philosophers on the list is less surprising given that effort to extend the notion of causation has been spearheaded by a network of philosophers of medicine collectively working under the label of "EBM+". In light of this origin, what is particularly impressive is the group's success at reaching outside of philosophy to medical researchers, including those at the center of the movement they seek to reform (e.g., Jeffery Aaronson and Trish Greenhalgh at the Center for Evidencebased Medicine at the University of Oxford). This is doubly impressive given the sweeping changes the letter encourages in relation to the evaluation of medical evidence in general. However, while avant-garde in medical research, the positions expressed in the letter could be fairly characterized as the received view amongst philosophers of science. We will argue that this stems in part from the fact that view emerged not from within a discussion of medical evidence, but from a previously conceived view of causation which was then grafted onto debates in medical research as a potential application of the existing philosophical view.

To substantiate our position, we provide a historical account of the development of EBM+ and identify its tenets. Our analysis shows that 1) EBM+'s epistemology of causation is motivated by a specific problem in the philosophy of causation (i.e., accounting for causal discovery), 2) that EMB+ is predicated upon specific metaphysical and probabilistic positions regarding the epistemology of causation, and that 3) EBM+'s epistemology of causation is derived from a set of meta-epistemological desiderata which fail to account for the social context of medical research. After presenting our genealogy of EBM+, we survey objections to EBM+ (Holman, 2019; Andreoletti and Teira, 2019) as well as how these critiques have been responded to. Here, we argue that the responses to the critiques of EBM+ have either begged the question by appealing to the same meta-epistemological standards under dispute (Williamson, 2020) or appeal to case studies that do not probe circumstances where critics argue EBM+ would fail (Gillies, 2019; Auker-Howlett and Wilde, 2020). As such, we conclude that while EBM+ may have certain philosophical merits, these are not sufficient grounds to abandon the priority placed on RCTs, in part because such an acontextual understanding of causation has led proponents of EBM+ to be insensitive to the material realities of medical research which could undermine the applicability of their proposed reforms.

#### References:

Andreoletti, M. and Teira, D., 2019. Rules versus standards: what are the costs of epistemic norms in drug regulation?. Science, Technology, & Human Values, 44(6), pp.1093-1115.

Auker-Howlett, D. and Wilde, M., 2020. Reinforced reasoning in medicine. Journal of Evaluation in Clinical Practice, 26(2), pp.458-464.

Gillies, D., 2019. Holman's criticisms of EBM+.

Holman, B., 2019. Philosophers on drugs. Synthese, 196(11), pp.4363-4390.

Williamson, J., 2021. The feasibility and malleability of EBM+. Theoria: An International Journal for Theory, History and Foundations of Science, 36(2), pp.191-209.

# Sus, Adán

# Symmetries, the Helmholtz-Lie-Weyl theorem and Earman's principles

Earman(1989) states two symmetry principles, understood as conditions of adequacy between dynamics and spacetime structures in a theory of motion, that amount to prescribing the extensional equivalence of dynamical and spacetime symmetries. Much in recent debates about symmetries in physics is related to the discussion about the validity, limitation or refutation, and eventual justification of Earman's principles. Examples of this are some recent defenses of the dynamical approach to relativity theory that take the principles as analytic or, contrarily, claims to the effect that there are historical counterexamples to the equivalence between spacetime and dynamical symmetries.

In this talk, I intend to tackle directly the question of the justification of Earman's principles with the aim of providing an improved version of a principle that relates spacetime symmetries and dynamical symmetries. Part of the problem with the original formulation is that it is not entirely clear what the two extremes of the equivalence, namely spacetime symmetries and dynamical symmetries, are. So one necessary component of the approach here presented will involve a clarification of what these terms mean in order to be part of the formulation of a principle that plays an essential role in the identification of spacetime structures in a theory.

In order to do so, I make contact with two different discussions in the context of spacetime theories. The first one is the so-called Problem of Space as addressed by authors like Riemann, Helmholtz, Poincaré, Lie and Weyl. (Dewar(2020), Sholz(2019) are some recent presentations of this problem). One can argue that the main motivation behind these different proposals has to do with stating the conditions from which spatial geometry (in the original pre-relativistic context) and spatiotemporal local chrono-geometry (in Weyl's version of the problem) can be derived. A dimension of this discussion, that proves extremely fruitful for the formulation of a symmetry principle, is that it hints to a explicit connection between spacetime geometry and certain conditions (like free mobility in the original formulation) that can be interpreted dynamically. I propose and discuss a general formulation of these results (a Helmholtz-Lie-Weyl theorem) and extract from it a promising perspective for the characterization of spacetime symmetries.

On the other hand, we have the discussions about the physical and empirical significance of different symmetry transformations. There is a natural approach to this problem, developed mainly by David Wallace, that makes such a significance dependent on questions about how the symmetry transformations of a given subsystem extend to transformations that include the environment (other subsystems) that interacts with it, together with the explicit consideration of how the symmetries affect the physical process of measuring. I fully endorse this approach and put it to work to provide an interpretation of dynamical symmetries, dependent on the characterization of the workings of the measuring procedures implicit in the theory, that allows to establish the link between dynamical and spacetime symmetries.

The result of combining appropriately these two discussions is the formulation of an improved symmetry principle, together with a framework in which to discuss its

foundations, that will allow us to understand better its limitations and confront its counterexamples.

#### References:

Dewar, N. and Eisenthal, J. (2020). "A Raum with a View." In Claus Beisbart, Tilman Sauer & Christian Wüthrich (eds.), Thinking About Space and Time: 100 Years of Applying and Interpreting General Relativity. Springer. pp. 111-132.

Earman, J. (1989). World Enough and Space-Time. Absolute versus Relational Theories of Space and Time. MIT Press.

Sholz, E. (2019). "The changing faces of the problem of space in the work of Hermann Weyl." In Carlos Lobo and Julien Bernard (eds.) Weyl and the Problem of Space. Springer Verlag.

Wallace, D. (forthcoming). "Observability, redundancy and modality for dynamical symmetry transformations". In J. Read and N. Teh (eds.). The Philosophy and Physics of Noether's Theorem. Cambridge: Cambridge University Press.

# Taşdan, Ufuk, Roberts, Bryan W., and Thebault, Karim P. Y.

## Spacetime Conventionalism Revisited

Poincare (1914) and Reichenbach (1957) famously argued that the geometry of space, and by extension spacetime, is a matter of convention: spacetime geometry is empirically underdetermined in such a way that, together with an appropriate empiricist stance, the choice of geometry is conventional. The subsequent discussion was primarily focused on the underdetermination of relativistic spacetime models by empirical observation. Thus, Reichenbach's (1957) conventionality thesis, which claimed that `universal forces' give rise to empirically adequate but distinct spacetime geometries, is typically viewed as a statement about the underdetermination of relativistic spacetimes. More recently, Weatherall and Manchak (2014) have given an elegant and precise statement of this approach, defining `universal forces' to be characterised by a field that accelerates a test particle accelerates away from geodesic motion. Their conclusion is that, on this reading of conventionality, the thesis is provably false. In this context, it is worth mentioning the possibility, suggested by Carnap (1966) in the introduction to Reichenbach's last book, that such forces could be understood as `universal effects' rather than actual forces to imitate the complex nature of the effect of the curvature. A similar idea is considered by Dieks (1987).

In this paper, we generalise the concept of a `universal force' to a less-restrictive form, following the suggestion of Carnap (1966), and to some extent Dieks (1987), to consider a more general class of `universal effects' as evidence for conventionality. Here, we no longer assume that it is the spacetime models that are underdetermined, but a more general class of geometric properties; and, we no longer assume that forces must take their standard form as tensor fields, but allow for more general effects. Our particular focus will then be on a putative conventionality associated with matter-coupled and geometric degrees of freedom.A simple example arises in the cosmological constant problem: While Einstein treated as a geometric object on the left-hand side of Einstein's equation, contemporary physicists treat it as a matter-energy field on the right-hand side. Our key move in the reframing of conventionality question is to consider the possibility of using the Weyl and Ricci tensors as a means to distinguish 'geometric' as opposed to 'matter-energy' aspects of gravitation, respectively. In this context, the Einstein equations relate the `volume' aspects of geometry to stress-energy, and are silent with regard to the `shape' aspects. We interpret Weyl curvature, and the associated gravitational excitations, as the `free-geometric' degrees of freedom, and the Ricci curvature, and the associated gravitational excitations, as the `matter-coupled' degrees of freedom. We propose that once understood as manifestations of the free gravitational field and analysed in terms of Weyl curvature, tidal forces are evidently non-conventional.

#### References:

Carnap, R. (1966). An introduction to the philosophy of science (Dover ed ed.). Dover. Dieks, D. (1987). Gravitation as a universal force. Synthese, 73 (2), 381{397. Poincare, H. (1914). 1919. science and method. Trans. Francis Maitland. London, Edinburgh,

Dublin, New York: Thomas Nelson and Sons.

Reichenbach, H. (1957). The philosophy of space and time. New York.

Weatherall, J. O., & Manchak, J. B. (2014). The geometry of conventionality. Philosophy of Science,

81 (2), 233-247.

# Thompson, Morgan

# Characterizing and Measuring Racial Discrimination in Public Health Research

Researchers in public health use racial discrimination in their causal explanation of racial health disparities. I argue that different characterizations of racial discrimination (and different measurement tools) are needed to satisfy this causal explanatory goal.

Many researchers measure experiences of racial discrimination via self-report scales, such as Williams's Everyday Discrimination Scale (1997) and Krieger's Experiences of Discrimination scale (Krieger et al. 2005). These scales require participants to first identify instances of discrimination and then provide responses to the scale.

There are two challenges to this way of measuring racial discrimination. First, it does not measure intersectional discrimination where a discriminatory experience is due to membership in multiple social groups (Harnois, Bastos, and Shariff-Marco 2020). Second, some experiences of discrimination are such that the cause of the event is ambiguous to the victim (call this ambiguous discrimination). This second type appeals to attributionally ambiguous microaggressions, which are slights that may or may not be attributable to the victim's social group membership. Paradigmatic cases are ones where alternative non-social-identity-based explanations of the event are plausible.

Intersectional discrimination could be measured by modifying Williams' scale to allow participants to select multiple reasons why a discriminatory experience has occurred. Krieger's scale cannot be adapted because her scale builds in the cause of the discriminatory experience as race, color, or ethnicity.

The second type constitutes a more fundamental challenge. Experimentation in psychology requires designing experimental conditions that generate the right kind of reactivity in the subject while suppressing other reactivity (Feest and Moore 2022). My contention is that these scales generate reactions about experiences of attributable racial discrimination and suppress reporting of ambiguous discrimination.

First, some discriminatory actions measured by Williams' and Krieger's scales are less plausibly given alternative explanations, such as being threatened or harassed, being called names or insulted, and being followed around stores. As a result, participants will be less likely to identify cases where they are uncertain about whether their race was a cause of the other person's behavior and so, ambiguous discrimination will not be reported.

Second, some questions require attribution of normative concepts and so the stakes of attribution are higher. Both scales refer to disrespectful treatment, discriminatory treatment, and being made to feel inferior. In cases where the behavior is ambiguous between being caused by one's race and being caused by some other non-social-identity feature, I predict individuals are less likely to agree that the actor was acting disrespectful, discriminatory, or as if they are inferior and thus less likely to report these cases.

It is important to measure both types of discrimination because I predict they have different psychological effects, which in turn could lead to different negative health outcomes. Attributable discrimination is likely to cause emotions like anger or coping mechanisms like

John Henryism; whereas ambiguous discrimination is more likely to cause psychological moderators like perseverative cognition and depressive rumination. If my arguments are correct, public health researchers need multiple characterizations of racial discrimination and new measurement scales to satisfy their causal explanatory goals.

#### References:

Brigandt, I. (2010). The epistemic goal of a concept: accounting for the rationality of semantic change and variation. Synthese, 177(1), 19-40.

Feest, U., & Moore, I. (2022). Data Quality, Experimental Artifacts, and the Reactivity of the Psychological Subject Matter. European Journal of Philosophy of Science.

Harnois, C. E., Bastos, J. L., & Shariff-Marco, S. (2020). Intersectionality, contextual specificity, and everyday discrimination: Assessing the difficulty associated with identifying a main reason for discrimination among racial/ethnic minority respondents. Sociological Methods & Research, 0049124120914929.

Krieger, N., Smith, K., Naishadham, D., Hartman, C., & Barbeau, E. M. (2005). Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health. Social science & medicine, 61(7), 1576-1596.

Williams, D.R., Yu, Y., Jackson, J.S., and Anderson, N.B. "Racial Differences in Physical and Mental Health: Socioeconomic Status, Stress, and Discrimination." Journal of Health Psychology. 1997; 2(3):335-351

# Tulodziecki, Dana, Porterfield, David Marshall, and Kruse, Colin P.S.

# Experimental control, validation, and replicability in omics research

'Omics research' describes a collection of allied biotechnology approaches similar to or enabled by DNA sequencing technology developed for the human genome: genomics (DNA) follows into transcriptomics (RNA), proteomics (protein), and metabolomics (small metabolite molecules), mirroring the natural flow of biological information (genotype) into biological expression (phenotype). Omics research is fast becoming a standard approach for contemporary life sciences research, especially in medicine, where it forms the basis of biomedical technologies seeking to give rise to personalized and precision medicine ([1], [3]).

Our goal in this paper is to argue that omics research is plagued with significant epistemological problems: when performing omics experiments, sample preparation, experimental error, and sample biological variability are all areas in which validation, experimental control, and replicability are lacking.

First, due to the complexity and expense of omics experiments, most omics experiments are not run in replicate. Under 'typical' (non-omics) laboratory conditions, during the course of an experiment, a specimen is sampled more than once and each individual sample processed separately in order to determine both (a) the reliability of the sampling process, and (b) the representativeness of the sample. In contrast, during the course of many omics experiments, usually each sample is analyzed a single time. In some even worse scenarios, multiple samples that should have been processed separately (such as samples from different plant specimens that are extremely valuable individually) are consolidated and processed together as one mixed sample. We will argue that therefore both (a) and (b) are called into question; as a result, it is unclear to what extent the data from such experiments is compromised. This has potentially far-reaching consequences, since such data might be used as the basis for policy decisions (specific health risk factors) and research investments (lead drug compounds, tax dollar research grants, etc.).

Second, omics instrumentation standards themselves are problematic. Some existing standards are supposed to ensure precision and accuracy by calibrating individual samples for analysis to their respective ribosomal RNA (an abundant and species independent type of gene). We will argue that these standards involve a significant underlying assumption, namely that ribosomal RNA is a consistent -- and thus universally appropriate -- calibration reference for all organisms. As we will show, however, there is good reason to think that there are exceptions to this.

We will then illustrate the magnitude of these epistemological issues through an example of a serious biological mishap: the mutational drift in Henrietta Lacks's famous cells ([2], [4]). We will argue that the adequate application of proper epistemological standards involved in omics analyses could have prevented inconsistencies and potentially erroneous findings in well over 500 peer-reviewed manuscripts.

We end by categorizing the different problems outlined above and argue that, while some of them are practical and thus at least have a theoretical solution, there also remain more serious epistemological issues for which it is not easy to see how they could be resolved, even in principle. To wrap up, we discuss the implications of this for the results obtained through such research.

#### References:

Barh, D., Zambare, V. and Azevedo, V. eds., 2013. Omics: applications in biomedical, agricultural, and environmental sciences. CRC Press.

Horbach, S.P. and Halffman, W., 2017. The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. PloS One, 12(10), p.e0186281.

Karczewski, K.J. and Snyder, M.P., 2018. Integrative omics for health and disease. Nature Reviews Genetics, 19(5), pp.299-310.

Scherer, W.F., Syverton, J.T. and Gey, G.O., 1953. Studies on the propagation in vitro of poliomyelitis viruses: IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix. The journal of experimental medicine, 97(5), pp.695-710.

## Wells, Aaron

## Du Châtelet on Newton's Rules for Reasoning in Natural Philosophy

Emilie Du Châtelet is well-known for her complete French translation, with commentary, of Newton's Principia. But there is a long-standing interpretive question about her relationship to Newton's methods in her earlier work on philosophy and physics, the Institutions de physique (first edition Paris, 1740). As William Barber noticed in the 1960s, drafts of this work (which only partially survive) indicate a crucial role for at least some of the Principia's famous four rules of reasoning in natural philosophy. By contrast, the published version of the work does not begin with these rules. Du Châtelet instead seems to add the two rationalist principles of contradiction and sufficient reason. Barber concludes that Du Châtelet began as a doctrinaire Newtonian in the early drafts, but then underwent a "Leibnizian illumination," which is displayed in the "metaphysical foundation" he takes to be added in the published Institutions.

Barber makes several problematic assumptions that have been questioned by more recent scholarship. One is that Newton's rules for reasoning in natural philosophy have nothing to do with metaphysical foundations. Another is the narrative that Du Châtelet simply converted, more or less passively, from a Newtonian system to a Leibnizian one when she became aware of the latter's work. Ursula Winter has persuasively argued, for example, that Du Châtelet already knew quite a bit about Leibniz's system when she began drafting her Institutions. Yet we still find many commentators describing Du Châtelet as simply turning from 'Newtonian' to 'Leibnizian' method. It remains an open question why this occurred. There is, moreover, a surprising lack of discussion of Barber's original manuscript source of evidence.

In this talk I first call attention to six neglected features of Du Châtelet's manuscript presentation of Newton's rules. Here, already, she departs from some Newton in important way. I then offer some interpretive hypotheses about how Du Châtelet's views evolved on five of these six points.

Unlike Newton, Du Châtelet's manuscript (i) presents Newton's Rule 3 as a strict law, and also (ii) as a matter of common-sense agreement among "all philosophers," while dropping (iii) Newton's notorious reference to intension and remission. Moreover, (iv) she holds that by Rule 3, we can regard some qualities of bodies not only as "universal," but as "inherent": on this point, she also goes beyond what Newton says explicitly. However, her framing of the significance of Rule 3 stays close to Newton in some other respects. For (v) she emphasizes, like Newton, that we have no non-inductive access to the properties of bodies or matter. Also as in Newton, (vi) the key application is the argument for universal gravitation is stronger than the case for the universal impenetrability of matter.

Turning now to interpretive hypotheses: the published Institutions seems to at least implicitly reject (i) and (ii). Yet Du Châtelet explicitly endorses (v), as well as relying, I'd argue, on something like Newton's Rules 3 and 4 in her very positive presentation of Newton's empirical arguments for universal gravitation; this parallels point (vi). Neither the principle of contradiction nor the principle of sufficient reason can do the required work here. Relatedly, she seems to accept the universality plank of (iv). It can also be argued that she continues to accept the inherence claim, in the following sense: the universal qualities in question must be seen as grounded in nonrelational physical qualities. This can be seen as bolstering her interest in an aether explanation of gravitation. While most would consider the inherence claim to be a major departure from Newton (though not from Newtonians such as COtes), it is significant that Du Châtelet already took this step in the early manuscript stages of the Institutions.

The details of these interpretive hypotheses are controversial and cannot be fully defended here. The general point I want to insist on is that Du Châtelet did not undergo a simple conversion from one existing system to another. She instead engaged in an active and original synthesis of methodological ideas from both Newton and the Continental rationalists.

#### References:

Barber, W. H. (2006/1967). Mme Du Châtelet and Leibnizianism: The Genesis of the Institutions de physique. In J. P. Zinsser & J. C. Hayes (Eds.), Emilie Du Châtelet: Rewriting Enlightenment Philosophy and Science (pp. 5–23). Oxford: Voltaire Foundation.

Brading, K. (2019). Émilie Du Châtelet and the Foundations of Physical Science. New York: Routledge.

Harper, W. & G. E. Smith. (1995). Newton's New Way of Inquiry. In J. Leplin (Ed.), The Creation of Ideas in Physics (pp. 113–166). Dordrecht: Kluwer Academic Publishers.

Janiak, A. (2021). Emilie Du Châtelet's Break from the French Newtonians. Revue d'histoire des sciences 74-2, 265-296.

Winter, U. (2012). From Translation to Philosophical Discourse—Emilie Du Châtelet's Commentaries on Newton and Leibniz. In R. Hagengruber (Ed.), Emilie Du Châtelet between Leibniz and Newton (p. 173–206). Dordrecht: Springer.

# Whiteley, Cecily

## Two Approaches to the Science of Sleep Experience

The standard approach to the science of sleep experience proceeds by way of phenomenological definitions. Consciousness researchers examine the phenomenological features of sleep experience in order to draw a definitional line between those conscious experiences in sleep that count as 'dreams' as opposed to 'dreamless' experiences. These definitions subsequently form the explanatory targets of empirical research into the biological functions and neural basis of dreaming. This 'phenomenological approach' in consciousness science is most clearly exemplified in the work of the growing number of philosophers and neuroscientists who endorse simulation models of dreaming. According to these models, a large subset (but not all) of our conscious sleep experiences are unified, and thus warrant classification into a single neurobiological kind 'dreaming', in virtue of sharing a distinctive simulation-like phenomenology (Revonsuo, Tuominen, & Valli 2015, Revonsuo 2006, Thompson 2015, Windt et al. 2016, Windt 2015, 2018, 2020). That is, on this view, the majority of our sleep experiences are unified through sharing a 'core' of phenomenological properties: they each involve immersive, vivid 'here and now' experiences of a self in a world.

Despite its widespread adoption by consciousness scientists, little philosophical attention has been paid to the question of whether the phenomenological approach is a good approach to adopt in consciousness science -- whether its underlying methodological and metaphysical assumptions are epistemically sound. This is likely due to the fact that it is unclear that there are serious methodological problems associated with this approach, and that there is an alternative methodological framework available for consciousness science to take up. This paper takes up this challenge and offers a critique of the phenomenological approach alongside the provision of a new methodological framework for the science of sleep experience informed by philosophy of science.

Against the phenomenological approach, two objections are put forward. First, contrary to claims made by proponents of simulation views, I argue that phenomenological definitions of dreaming like the simulation view above are largely untestable, and insensitive to revision in light of empirical evidence. That is, both the empirical considerations offered in favor of these definitions, as well as the reliability of the predictions which are said to follow from them, depend upon a prior acceptance of the definitions these are meant to support. This raises serious questions about the scientific status of the phenomenological approach to dreaming.

Second, I argue that phenomenological definitions are problematic in so far as they build in a priori assumptions about the number of neurobiological kinds present in sleep experience; assumptions which should be left for a posteriori empirical investigation to establish. I argue that the a priori assumption that dreaming is a single kind of state plays an active organizing role in dream science which often goes unnoticed, shaping not only the way empirical results are interpreted, but also the way in which experiments in consciousness science are constructed and designed. I demonstrate this with reference to two case studies: recent research on the neural correlates of dreaming and the nature of lucid dreaming. These methodological obstacles motivate the provision and uptake of an alternative approach to sleep experience. According to the alternative 'natural kind' framework I propose, consciousness science should proceed by identifying the natural clusters of phenomenological, neurophysiological and functional properties in sleep which track distinct global states of consciousness in the sense recently discussed in consciousness science (Bayne et al. 2016). This encourages a theoretical openness to a new possibility viz. that 'dreams' and 'sleep experience' may not form natural kinds.

#### References:

Revonsuo, Antti & Tuominen, Jarno. (2015). The Simulation Theories of Dreaming: How to Make Theoretical Progress in Dream Science. Open MIND. 1–8. 10.15502/9783958570894.

Windt, J. M., Nielsen, T., & Thompson, E. (2016). Does Consciousness Disappear in Dreamless Sleep?. Trends in cognitive sciences, 20(12), 871–882. https://doi.org/10.1016/j.tics.2016.09.006

Bayne, Tim, Jakob Hohwy, and Adrian M. Owen. "Are there levels of consciousness?." Trends in cognitive sciences 20.6 (2016): 405-413.

Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J. J., Riedner, B., ... & Tononi, G. (2017). The neural correlates of dreaming. Nature neuroscience, 20(6), 872-878.

Windt, Jennifer M. "The immersive spatiotemporal hallucination model of dreaming." Phenomenology and the Cognitive Sciences 9.2 (2010): 295-316.
## Żuradzki, Tomasz

## Embryo-like structures, value-loaded metaphysics of science, and regulation of biomedical research

Philosophical discussions on non-epistemic values in science usually focus on the role of values in inspiring scientific questions, affecting scientific methodologies, and setting the level of evidence needed for drawing conclusions (Elliott, 2017). On the examples of human embryos and embryo-like structures (Ankeny, Munsie, & Leach, 2022), we analyze the role of non-epistemic values in classificatory practices in biomedicine. In particular, we focus on cases when regulatory mechanisms limit research, either in the forms of direct bans (e.g., some cases of human embryo research) or indirect incentives (e.g., bans on public funding or patenting some procedures regarding human embryos and embryo-like structures). Thus, we challenge all these philosophical theories of classification and kinds that do not accommodate the role for non-epistemic values (Khalidi 2013).

We will discuss the 14-day rule stating that in vitro research on human embryos and some embryo-like structures is permissible, but only until two weeks after fertilization or creation (Matthews & Moralí, 2020). One may interpret this rule as assuming that purely value-free biological facts about human embryos (e.g., individuation, i.e., the fact that embryos can no longer twin, or the first appearance of the primitive streak around this time, which is a precondition for the capacity to feel pain) ground the moral or legal status of organisms. Such a "metaphysics-first" approach tries to settle the metaphysical question of what a human embryo is – what is its essence or definition – and from there derive normative conclusions about, e.g., regulations on research (e.g. Lee 2004).

In this paper, we argue that this view is mistaken because, in particular, in the case of research-oriented biological classifications, there is no value-free (or interest-free) metaphysics of science (cf. Dasgupta, 2017). Our approach first takes into consideration the interplay between epistemic and non-epistemic values in real cases of biological classifications (Reydon & Ereshefsky, 2022), and then draws conclusions about metaphysics, i.e., how human embryos or "synthetic human entities with embryo-like features" (Aach et al., 2017) may be classified to suit a given value framework.

For example, human parthenogenetic stem cells are excluded from the patenting prohibition of procedures based on hESC by the European Biopatent Directive, because such stem cells have been defined differently than human embryos or other types of stem cells: the parthenogenetic ones do not have 'the capacity' to develop into a (born) human being, i.e. totipotency (see European Court of Justice 2014). However, the capacities of some of the embryo-like structures may be measured after the realization of these capacities in controlled environments (Fagan, 2013). Since their capacity is context-dependent, there is no such thing as a value-neutral environment in which we can judge the embryo's or stem cells' 'genuine' intrinsic potential (Piotrowska, 2020). Thus, any forward-looking definition of embryos or embryo-like structures is grounded in specific normative evaluation on what counts as 'normal environment'.

We conclude with a few remarks on the role of philosophers of science and bioethicists in the realm of science policy and we argue that major normative and regulatory issues in

biomedical research would benefit from the tighter integration of these two disciplines (Lohse, Wasmer, & Reydon, 2020).

References:

Ankeny, R. A., Munsie, M. J., & Leach, J. (2022). Developing a reflexive, anticipatory, and deliberative approach to unanticipated discoveries: Ethical lessons from iBlastoids. American Journal of Bioethics, 22(1), 36-45.

Dasgupta, S. (2017). Essentialism and the Nonidentity Problem. Philosophy and Phenomenological Research, 46(3), 540-570.

Fagan, M. B. (2013). The stem cell uncertainty principle. Philosophy of Science, 80(5), 945-957.

Lohse, S., Wasmer, M. S., & Reydon, T. A. (2020). Integrating philosophy of science into research on ethical, legal and social issues in the life sciences. Perspectives on Science, 28(6), 700-736.

Reydon, T. A., & Ereshefsky, M. (2022). How to Incorporate Non-Epistemic Values into a Theory of Classification. European Journal for Philosophy of Science, 12(1), 1-28.

## Symposia

## **Comparative Cognitive Archaeology**

### Adrian Currie, Marta Halina, Anton Killin, and Andra Meneganzin

This symposium explores the possibility of comparative cognitive archaeology, that is, exploring the space of hominin cognitive capacities by drawing on the archaeological record to infer differences and similarities between both early behaviourally modern Homo sapiens and between them and so-called 'archaic' hominins. Doing so would allow us to map the space of human intelligence in terms of our recent phylogenetic history and potentially reveal diverse ways of exhibited intelligence in hominins.

In the last twenty years, our perception of the evolution of our species has become both more complex and less linear. First, overturning assumptions of linear evolutionary progress it is now accepted that at least five other hominin lineages shared the globe with Homo sapiens fifty thousand years ago. Rather than a tree, we have a bush. Further, admixture events appear to have been fairly common at least between H. sapiens, Neanderthal and Denisovans (and perhaps other "ghost" lineages). Rather than independently evolving, parallel lineages we have a complex picture of admixture. Second, the cognitive and technological sophistication of our relatives, particularly Neanderthal, is increasingly accepted. These other hominins were not comparative dullards, but seem to have been innovative: cultural exchange between them and H. sapiens was putatively far from oneway. These two sets of discoveries have reshaped how we think about the evolution of our species and the extinction of our relatives. We now see ourselves as the last surviving twig of a recently flourishing bush.

Most recent philosophical attention to paleoanthropology has either been synthetic work generating narratives of human evolution—or has grappled with epistemic questions concerning how inferences may be made from the archaeological and paleontological records. This symposium opens a new set of questions by placing the evolution of our species in a comparative cognitive context. That is, it recognises that by comparing and contrasting the cognitive acumen of ourselves and our relatives we gain a richer conception of the evolution of intelligence in human-like creatures: of the biological, cultural, environmental and demographic pre-requisites, of the cognitive contingencies and necessities. Further, the project aids in developing a space of the diversity of intelligence. In this, it is closely aligned with projects in comparative psychology, but with telling and fascinating differences.

Comparative psychology is a set of experimental practices aimed at establishing and comparing the cognitive capacities of different lineages in a phylogenetic context. Comparative psychologists seek to establish whether problem solving, mind-reading, mirror-recognition, and so on, are present in a bewildering array of organisms, from elephants to non-human primates to corvids, cephalopods and insects. Like paleoanthropology, both have a long history of anti-anthropocentric aims: comparative psychology has emphasized the cognitive sophistication of non-human animals similarly to how paleoanthropologists have emphasized the cognitive sophistication of 'archaic' hominins. Further, both have

similar debates concerning the forms of cognition required to perform particular behaviours: to what extent does, say, Neanderthal and corvids building sophisticated tools, justify inferring sophisticated folk physics or mental time travel to them? Although philosophers have discussed these in isolation, this symposium is the first to bring the two into dialogue.

Differences between comparative psychology and comparative cognitive archaeology are also telling. First, where the former is experimental the latter is historical, relying on inferences from the archaeological record. In a comparative context, instead of asking how a particular signal from the record might tell us about the cognitive capacities of a particular lineage, we ask how differences in assemblages across groups might tell us about differences in their underlying cognition. Further, practices of hypothesis acceptance ('Morgan's canon' in comparative psychology, for instance) potentially play out quite differently in these differing epistemic contexts.

The second major difference is phylogenetic scope: instead of all animals (and potentially including bacteria, plants and fungus in some contexts!) comparative cognitive archaeology is interested in an extremely tight-knit set of lineages. This promises to uncover parallel (as opposed to convergent) patterns of evolution. For instance, although Neanderthals were evolutionarily separated from H. sapiens by over half a million years, paleoanthropologists now think that many of the underlying cognitive pre-requisites of 'distinctive' sapiens intelligence were present in the common ancestors of such groups. Understanding these patterns of cognitive evolution are crucial for getting to grips with how human-like intelligence evolves.

Overall, then, this symposium explores questions at the cutting edge of understanding human intelligence by exploring the possibilities of comparative cognitive archaeology. We draw on comparisons with experimental studies of non-human animal intelligence, as well as recent work on inferences within both archaeology and comparative biology to identify the potential and challenges of understanding the evolution of intelligence in hominins in local phylogenetic contexts, and further how this may build a space of hominin intelligence.

## Feminist Philosophy of Biology Beyond Gender

#### Azita Chellappoo, Saana Jukola, David Lambert, Tamar Schneider, and Sophie Viegl

Feminist epistemology and philosophy of science have become major branches within philosophy of science more generally, with insights from these fields bearing on longstanding debates surrounding notions of objectivity, knowledge, and scientific methodology, as well as the role of values in science. Although feminist philosophers of science have held a wide variety of positions, often in conflict with one another, there has been a general consensus regarding the inseparability of the epistemic from the social, political and cultural, and the construction of knowers not as isolated islands, but rather as fundamentally and inextricably socially situated.

The biological sciences in particular have often been a central target of feminist philosophical critique. Feminist philosophers of biology have, for example, challenged sexist and androcentric biases in fields such as sociobiology, evolutionary psychology, and primatology (Bleier, 1984; Haraway, 1989; Lloyd, 2003), queried our practices of sex categorisation (Fausto-Sterling, 2000), and challenged reductionist tendencies in models of the effects of sex hormones in development (Longino, 1990). Perhaps unsurprisingly, feminist philosophy of biology has often primarily been concerned with gender as a central analytical category. Many of the areas in which significant contributions have been made have been those where gendered assumptions, biases, or values have a clear role in shaping scientific knowledge production.

We do, however, think that this preoccupation restricts the full potential of feminist epistemology in the philosophy of science. We regard feminist epistemologies as first and foremost tools to uncover (in various ways, and by varying means) how societal power relations are inscribed in scientific inquiry. Given the wealth of feminist scholarship demonstrating the long reach of societal power relations, and the social situatedness of knowledge, it seems that there should be no conceptual or pragmatic restriction on which phenomena are subject to 'feminist' critique. The tools of feminist epistemology in general, and feminist philosophy of biology in particular, that have been critical in uncovering the damaging role of gender biases in many areas of the biosciences, can and should be fruitfully applied in areas beyond the 'usual suspects'. This perspective leads us to the identification of two lacunae in the literature: firstly, the comparative lack of attention to other forms of social hierarchy and their intersections in the context of critiques of science and scientific knowledge production. Although there is a wealth of literature on intersectionality and its implications for epistemology, the application of the insights of feminist philosophy of biology to the operation of social categories other than gender remains underdeveloped. Secondly, the tools that feminist philosophy of biology can offer, including the subtle role of non-epistemic values in shaping adoption of certain epistemic values that guide theory choice, have not been widely deployed to interrogate areas of the biosciences that do not appear to directly engage with social hierarchies or social relations. Nevertheless, these tools could be useful in understanding how knowledge production works in these areas.

In this symposium we aim to explore and go some way towards beginning to resolve these lacuna or areas of underdevelopment by (1) addressing broad conceptual questions that arise when understanding how feminist philosophy of biology can be fruitfully applied 'beyond gender', (2) exploring concrete examples within scientific practice where drawing on feministphilosophy of biology could be generative and is, as yet, underdiscussed, and (3) considering how feminist epistemology might bear on questions of understanding knowledge production in the philosophy of biology community itself.

#### Conceptual Framework for Feminist Philosophy of Biology Beyond Gender

Clarifying the conceptual grounding for expanding feminist philosophy of biology beyond the areas in which it has typically been implied is critical for understanding when and where this approach will be successful or useful. The contribution by Sophie Veigl ("Conceptual resources for a feminist philosophy beyond gender") will examine the conceptual and methodological toolboxes of recent approaches within feminist epistemology and bring them together to explore a first and tentative framework for feminist philosophy of science beyond gender.

#### Feminist Philosophy of Biology Beyond Gender: Cases from Scientific Practice

This symposium will include contributions that attend to particular areas of scientific practice and draw out how feminist philosophy of biology provides unique and worthwhile theoretical tools for analysing these domains. Tamar Schneider ("Bacteriology between chemical interactions and pathogenic individuals") draws on the work of feminist empiricists to explore two perspectives in microbiology as individualist and interactionist, connecting their background beliefs with notions of causality. This case study goes some way towards demonstrating the fruitfulness of feminist philosophy of biology in areas where there is no immediately obvious role for the social or social power relations. In addition, Azita Chellappoo ("Obesity Science" & Standpoint Theory") argues for the importance of standpoint theory in analysing and critiquing the methods of, and the knowledge produced by, "obesity" science. This case turns feminist epistemological tools towards the category of 'fatness', suggesting ways in which starting research from 'fat lives' can remedy ideological distortions.

#### Feminist Epistemology and Philosophers of Biology

Finally, whilst feminist critique of the sciences has been ongoing for decades, feminist critique of philosophy of science remains much rarer. Saana Jukola and David Lambert ("The philosophy of biology community as an epistemic community – Lessons from feminist social epistemology") draw on Helen Longino's account of science as a social enterprise to argue that analysing the community of philosophers of biology as an epistemic community, in which transformative intersubjective criticism is ideally achieved, can help pinpoint mechanisms which make philosophy of biology vulnerable to bias. They focus on practices that hinder the development of critical points of view.

# Exploring the senses of exploration in experimental neuroscience

### David Colaco, Jacqueline Sullivan, Pamela Rienagel, and Philipp Haueis

Experimentation in biology is often characterized by the need to explore unknown aspects of a target system or model organism. Exploration is often necessary for scientists to: isolate identifiable components within complex systems and bring them under experimental control, check if applying experimental techniques produces reliable results rather than artefacts, and generate hypotheses that can be subject to further experimental test.

Despite its value to experimental biology, exploration raises challenges for both practicing scientists and philosophers of science. On the practical side, funding schemes favor hypothesis-testing over exploratory research, leaving the impression that the former is epistemically superior to the latter (Haufe 2013). Biologists are often taught or required to couch their experiments in the language of "testing hypotheses". The effect is that many published papers state hypotheses merely for presentational purposes, while concealing the potentially exploratory nature of the experimental research (Rowbottom and Alexander 2012). The epistemic effects of this practice remain largely unexplored.

On the philosophical side, the label "exploratory experimentation" was initially pitched against the standard picture of experiments testing hypotheses derived from theory (Steinle 1997). Case studies from the life sciences have been instrumental in shaping philosophical thinking about the role exploration plays in experimentation (e.g., Burian 1997, Franklin 2005; Schickore 2016, 2018). More recently, philosophers of neuroscience have begun to consider the role exploration has historically played and continues to play in experimental neuroscience (Colaço 2018; Gamboa 2020; Haueis 2014, 2016, PSA 2021 presentation; Sullivan 2020, PSA 2021; Venturelli 2021). This work suggests that while exploratory experiments in neuroscience do not test hypotheses, they are nonetheless informed or constrained by theory. It is thus an open question whether the distinction between "theory-driven"/"confirmatory" and "exploratory" experiments adequately captures the kinds of experimental practices in neuroscience or whether it actually obscures the nature of experimentation in neuroscience.

This symposium brings together philosophers and practicing scientists to discuss the actual role that exploration plays in experimentation in neuroscience and the life sciences more broadly. The common theme among the individual papers is that case studies from neuroscience suggest that exploration is a prominent feature of all experiments. This broad sense of "exploration" has to do with the fact that experiments aim to discover something that researchers do not yet know (Rheinberger 1997), that neuroscientists require a flexible workflow to coordinate measurement devices, analysis techniques, and experimental conditions with the goals of the experimental study (Sullivan 2009, Haueis 2016), and that researchers need to try different experimental conditions to figure out what is signal and what is noise, and what could be potential confounds that need to be controlled for (Schickore 2019). This broad sense of exploration is a characteristic of experimental activity at all stages of research, regardless of whether scientists label their studies "exploratory" or "confirmatory".

Besides describing the general exploratory nature of neuroscientific experiments, the symposium inquires into three practical and philosophical issues of distinguishing confirmatory and exploratory in experimental biology. The first issue concerns the need of scientists to identify confounding factors to optimize their experimental arrangements (Schickore 2019). Identifying confounders is constitutive of exploratory experiments which aim to determine which parameters produce or modify an experimental effect (Steinle 1997). Yet scientists also identify confounders in the context of testing hypotheses. For instance, testing whether an entity or activity is relevant to a mechanism productive of a phenomenon requires researchers to distinguish experimental manipulations which successfully intervene on an entity or activity, from manipulations which obscure the mechanism's operation. By highlighting the relevance of confounds for hypothesis-driven experiments, the symposium adds to the burgeoning literature on identifying and controlling for errors, confounds and artefacts in experimental practice (Schickore 2019, Sullivan, 2018, Craver and Dan-Cohen 2021, Feest 2022).

The second issue concerns the statistical distinction between "confirmatory" experiments, which test pre-specified alternative hypotheses, and "exploratory" research, which may generate novel hypotheses about an unknown domain. Proponents of the distinction argue that biologists should choose different statistical methods, such as choosing statistical significance levels that reduce false negatives in confirmatory or false positives in exploratory research (Jaeger and Halliday 1998). Others claim that ignoring the distinction can lead researchers to hypothesize after the results are known (Kerr 1998), although the epistemic effects of this practice remain debated (Rubin 2022). So far, however, statistical and philosophical discussions of exploratory and confirmatory research have not been connected. By bringing philosophers and practitioners of experimental neuroscience together, the symposium addresses this lacuna. We inquire whether and how hypothesis testing fits together with the broad exploratory nature of experimentation in the life sciences. Is such experimentation ever "confirmatory" in the statistical sense? Can statistical concepts be used to cover both confirmatory and exploratory experiments, or does the description of flexible and data-driven workflows require different vocabulary?

The third issue concerns the pervasive but multifaceted role of theory in exploratory experimentation. Theoretical claims inform, if not even direct the application of experimental techniques in exploratory research (Franklin 2005, Colaço 2018). Even if they do not aim to test a hypothesis, researchers rely on some body of knowledge to determine which aspects of the system are worth investigating using the technique. However, this insight blurs the distinction between exploratory and confirmatory experiments, questioning whether it captures actual differences in experimental practice. The symposium advances these philosophical debates by introducing new accounts of exploration and exploratory experiments in neuroscience and the biological sciences more generally. These accounts propose novel answers to questions such as whether exploratory and confirmatory experiments lie on a continuum (Elliot 2007) or whether these categories adequately reflect that experimental activity can have a variety of aims, such as testing theoretical claims, developing new concepts and experimental techniques, or identifying and eliminating artefacts.

By exploring different senses of and questions surrounding exploration, the symposium promises a deeper philosophical understanding of experimentation in neuroscience. This understanding is firmly rooted in the symposium participants' own laboratory experience in

various fields of neuroscience and experimental biology (e.g. behavioral, cognitive and systems neuroscience, biochemistry and molecular biology).

References:

Colaço, D. (2018). Rethinking the role of theory in exploratory experimentation. Biology & Philosophy, 33(5), 1-17.

Craver, C. & Dan-Cohen, T. (2021). Experimental Artefacts. BJPS.

Burian, R.M. (1997). Exploratory Experimentation and the Role of Histochemical Techniques in the Work of Jean Brachet, 1938–1952. HPLS19(1), 27–45.

Elliot, K.. 2007. "Varieties of Exploratory Experimentation in Nanotoxicology". HPLS 29, 313-336. Feest, Uljana (2022) Data Quality, Experimental Artifacts, and the Reactivity of the Psychological Subject Matter, EJPS

Franklin, L. (2005). "Exploratory Experiments". Philos. of Sci. 72(5), 888–99.

Gamboa, J.P. (2020). Goltz against cerebral localization: Methodology and experimental practices. Studies HPS C, 84 101304. 2020.

Haueis, P. (2021). Exploratory Concept Formation and the Norms of Description. PSA 2021 Baltimore Presentation —–(2014). Meeting the Brain on its own Terms. Frontiers in Human Neuroscience 8(815) —–(2016). The Life of the Cortical Column. Opening the Domain of Functional Architecture of the Cortex (1955–1981). HPLS38(3). Haufe, C. (2013). Why do Funding Agencies Favor Hypothesis Testing? Studies HPS A 4 (3), 363-374

Jaeger, R.G. & Halliday, T.R. (1998)On Confirmatory versus Exploratory Research. Herpetologica 54, pp. S64-S66.

Kerr, N.L. (1998). HARKing: Hypothesizing after the Results are Known. Personality and Social Psychology Review 2(3), 196-217.

Schickore, J. (2016). "Exploratory Experimentation" as a Probe into the Relation between Historiography and Philosophy of Science. Studies HPS. 55, 20–25.

Schickore, J. (2019). The Structure and Function of Experimental Control in the Life Sciences." Philos. Sci. 86(2), 203–218.

Steinle, F. (1997). "Entering new fields: Exploratory uses of experimentation." Philos. Sci. 64, S65–S74.

Sullivan, J. (2009). The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-Reductionist Models of the Unity of Neuroscience. Synthese 167(511) —–(2020). Model Behaviors for Model Organism Research? Studies HPS C —–Optogenetics, Pluralism and Progress. Philos. Sci. 85:1090-1101 —–(2021) Novel Tool Development and the Dynamics of Control: The Rodent Touchscreen Operant Chamber as a Case Study. PSA 2021 Baltimore Presentation.

Steinle, F. (1997). Entering New Fields: Exploratory Uses of Experimentation. Philos. Sci. 64, S65–S74.

Rheinberger, H.J. (1997). Towards A History of Epistemic Things. Synthesizing Proteins in the Test Tube, 1947—1961. Stanford, CA: Stanford University Press.

Rowbottom, D., & Alexander, R. (2012). The role of hypotheses in biomechanical research. Science in Context 25(2), 247–262.

Rubin, M (2022). The Costs of HARKing. BJPS

Venturelli, A.N. (2021). The Exploratory Dimension of FMRI Experiments. Manuscrito 44(1), 1-36.

## The Rise of Basal Cognition and its Relevance for the Mind and Brain Sciences

### David Harrison, Urte Laukaiyte, Taylor Beck, and Wiktor Rorot

One can learn much from a cursory glance at the history of cognitive science. Nominally composed of six interrelated disciplines, it is well remarked that certain subdisciplines— notably psychology, linguistics and artificial intelligence, and later neuroscience—have had a disproportionate effect on the development and trajectory of the field as a whole. Thus, from the so-called 'cognitive turn' catalysed by Miller's cognitive revolution (Miller, 2003) and Chomsky's 'linguistic turn', through to the development of cognitive neuroscience as a standalone discipline, certain areas of research have been privileged in their influence and dominance in the field of cognitive science. Nevertheless, there appears to be a unifying thread connecting these distinct iterations of cognitive science, and that is the relative neglect of the materiality and material embeddedness of the mind-brain. In our symposium, we would like to suggest that the materiality of the body, brain, mind, and environment has not been taken far enough. That is, we articulate and discuss how and why the emerging field of basal cognition should have relevance in sculpting, shaping, and sketching the future of cognitive science.

Briefly, basal cognition is a new area of research that assesses the cognitive and computational capacities of 'basal' organisms: where this is typically taken to mean organisms that lack nervous systems or where nervous systems are present in only a minimal sense (e.g., in comb jellies). Standard model organisms consist of acellular or cellular slime moulds, colonial bacterial organisations, basal animals such as Cnidaria or Placozoa, and, in some cases, artificially engineered, 'soft' robotic systems capable of minimal forms of cognitive capacities. The idea here is to understand how cognition, mind, and intelligence manifest and relate to the metabolic, homeostatic, and physiological imperatives required for thriving in a precarious environment. Crucially, the absence of neurons in the case of Placozoa, as well as in the case of simple neural organisms as Hydra and other Cnidaria, the absence of a centralized nervous system is telling (Ginsburg & Jablonka, 2021). This is because the goal of basal cognition research is to decentralise the importance of neurons and nervous systems in manifesting cognitive capacities such as learning (Dexter et. al., 2019), goal-directedness (Levin, 2019; 2020), and prototypical instances of 'minimal' subjectivity (Godfrey-Smith, 2016a; 2016b). Indeed, as one prominent researcher in the area remarks, "capacities usually assigned to [organisms] with nervous systems, such as integrating spatiotemporal information, memory, and ability to pursue specific outcomes via selection from a number of possible behaviours evolved from far older pre-neural origins" (Levin, 2019: 2, emphasis added). Moreover, renowned neuroscientist Samuel Gershman (Gershman et. al., 2021: 1) has reconsidered the evidence for complex learning in unicellular organisms, suggesting that it has "the potential to profoundly reshape our understanding of learning in multicellular organisms".

One perspective comes from psychiatric disease, which offers a case study on basal cognition. Bipolar disorder, in particular, is a portrait of metabolism's impact on thought: a hereditary disorder of neuronal excitability, manifest in changes in drive and appetite — for sleep, food, sex, novelty — as well as mood shifts, racing thoughts and cognitive changes. Molecular findings about the cells of bipolar patients offer clues to what metabolic changes

may modulate dimensions of thought (Mishra et al., 2021; Shen et al., 2020; Hoffman, et al, 2018; Mertens et al., 2015; O'Shea et al., 2015; etc.).

Intriguingly, basal cognition has already influenced the field of soft robotics: a subdiscipline of robotics and artificial intelligence that places a premium on 'soft', 'elastic', and 'vulnerable' dimensions of embodiment, motility, and mentation (Bongard & Levin, 2021; Man & Damasio, 2019). Because researchers in the field of basal cognition associate mind with the physiological demands for survival, there also tends to be an emphasis on the 'precarious' nature of our biological embodiment (Froese 2016). Picking up on a similar theme, Man and Damasio (2019) have recently articulated a novel position vis-à-vis more traditional artificial intelligence and robotics approaches to Artificial General Intelligence (AGI). Thus, they write, "We propose the design and construction of a new class of machines organised according to the principles of life regulation, or homeostasis. These machines must have physical constructions—bodies—that must be maintained within a narrow range of viability states and thus share some essential traits with all living systems" (2019: 446). In addition to the implications basal cognition might have for the mind and brain sciences, it appears that it could motivate-via the related discipline of soft robotics-alternative approaches to more 'traditional' paths towards AGI. Indeed, this is most evident in the study of 'embodied computation', which lessens the cognitive load required of centralised processors by outsourcing to the materiality of the body itself. It is proposed, then, that the emerging field of basal cognition has ramifications and significance for some constituent disciplines of cognitive science.

Thus, our symposium proposes to chart, map, and discourse this exciting body of literature and assess how (indeed, whether) basal cognition research should permeate, influence, and affect the mind and brain sciences: from neuroscience to artificial intelligence. The underlying theme will centre around explicating basal cognition, its distinctive claims, potential problems, and the significance it may hold for other disciplines. Overall, we hope that our symposium will suggest some initial answers to questions such as: What lessons ought the cognitive and neurosciences draw from the claims found in basal cognition research? If the basis of mind and cognition is more 'somatic' and 'physiological' than commonly supposed, what does this mean for more traditional accounts of mentation and intellection? Do notions of distributed or embodied computations or representations have tractable relevance for neuroscience? What are the ramifications of these views for practical artificial intelligence and the eventual goal of AGI? Can this body of research provide any methodological direction for practitioners in the cognitive sciences? Basal cognition has opened many paths for investigation, and these are just some of the questions we hope to engage in our symposium.

## On the use of racial categories in medicine across geographic and national contexts

### David Ludwig, Joanna Malinowska, Zinhle Mncube, and Phila Msimang

Racial disparities in health are numerous in societies across the world. The scientific value of racial categories in biomedical research and clinical practice continues to be a subject of heated debate amongst philosophers of science. Critics employ a host of metaphysical, epistemological, and ethical arguments to argue that we ought to eliminate the practice of collecting and reporting of race data in biomedical research. Some critics argue that such 'race-based' research will erroneously reify race as a biological category. Other critics argue that using race in this way obscures heterogeneity in disease incidence amongst populations lumped under a single racial category (Valles, 2012). Yet other critics argue that race lacks value as a proxy for the genetic factors that play a causal role in health disease (Cooper et al., 2003).

Recently, philosophers of science and science studies scholars have argued against any strict demarcation of the biological and the social when it comes to race. These scholars point to emerging developments in biomedicine which provide mechanisms through which social racial hierarchies can become 'embodied' and effect biological and physiological change, contributing to disparities in health outcomes (Lorusso & Bacchini, 2015; Kalewold, 2020). Some scholars argue that in addition to tracking the effects of racism, for example, racial categories are useful in biomedical research to explore hypotheses about genetic causes of health disparities about races (Burchard et al., 2003; Andreasen, 2008; Spencer, 2018).

A current flashpoint in the debate about race and medicine is the practice of "race correction". Several risk calculators in nephrology, pulmonology, cardiology, and obstetrics, amongst others, use race correction factors to measure medical risk based on an individual patient's race. Scholars arguing for the abolition of this practice claim that this "adjustment" is based on essentialist and often racist notions about races (Vyas et al., 2020). They argue that race correction engenders negative consequences for patients.

However, implicitly, and explicitly, much of the philosophical debate about the use of race in biomedicine is limited to the United States (US) context. When Spencer (2018) argues that there is a racial classification that is useful in medicine, he explicitly limits his arguments to races as defined by the US Office of Management of Budget (OMB). When critics argue that race is a statistically unreliable variable, it is because the OMB's guidelines leave room for inconsistencies in identifying races. Finally, when scholars argue that race is useful to track the effects of racism, it's racism as it plays out in the US for US 'races'. However, racial categories are constructed, deployed, and shape medical categories in context-dependent ways (Ludwig, 2019). The use of racial categories in medicine across global and national contexts, and its implications, has received little attention in philosophy of science. This is a problem not only because the use of racial categories in medicine presents complex epistemological, ethical, and political issues that directly affect the care of individual patients across the world. In addition, philosophy of science is perfectly primed with the conceptual tools to provide a helpful starting point to tease these issues apart.

This symposium seeks to remedy this situation by bringing new, epistemically productive perspectives on this topic to the BSPS. By addressing the conceptual, epistemic, and political dimensions of the ways in which race enters into medicine from a global perspective, the symposium advances and shows the limitations of current debates on the scientific and clinical value of race.

#### The conceptual implications of varying racial ontologies

The variation in racial schemas across the world raises a host of interesting conceptual questions, including whether we should embrace pluralism about racial ontologies, how we can take account of the social context in our definitions of race, and when and where racial ontologies or racialised thinking are in effect. These questions gain particular significance when race comes to play a role in guiding practices in medicine, biomedicine, and health policy. The contribution by David Ludwig provides a conceptual framework for navigating the zones in which differing racial ontologies come into contact and shape one another. Even when racial categories are not explicitly invoked, the contribution of Joanna Malinowska and Tomasz Żuradzki shows that when ethnicity is used there is nevertheless the tendency towards the biologization of race in genomic research across 41 countries. Understanding the ways in which racial ontologies are maintained and negotiated in spheres of research and "trading zones" such as the interface between science and policy is not only necessary for a fuller picture of how race operates in its social context, but provides critical conceptual grounding for making sense of the range of ways in which racial schemas are utilised in medicine and biomedicine globally.

The epistemic and ethical challenges of the use of race in medicine around the world

This symposium also aims to highlight and explore the ways that racial categories are being and should be deployed in medicine in global contexts, as well as how this shapes knowledge production, health policy, and the clinical encounter. Phila Msimang argues that heterogeneity within racial groups in South Africa, both in terms of genetic variation and social and economic environment, provides a distinct challenge for the application of race categories in medicine in this context. Furthermore, the symposium will draw attention to the continuities and discontinuities between particular non-US contexts and the use of racial categories in the US: exploring this is crucial for understanding the extent to which the UScentric philosophical literature on the question of race in medicine can be usefully exported to other contexts. Zinhle Mncube and Azita Chellappoo draw on two case studies, from South Africa and India, to argue that race-based corrections in health measurement travel across borders while the underlying explanations for racial difference and racial ontologies differ. This creates distinct epistemic, ethical, and political challenges that are thus far under discussed in the race in medicine literature.

#### References:

Andreasen, R. O. (2008). The concept of race in medicine. In M. Ruse (Ed.), The Oxford handbook of philosophy of biology Oxford: Oxford University Press.

Burchard, E.G. et al. (2003), The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice. New England Journal of Medicine, 348, 1170–1175.

Cooper, R. S., Kaufman, J. S., & Ward, R. (2003). Race and genomics. The New England Journal of Medicine, 348(12), 1166–1170.

Kalewold, K. H. (2020). Race and Medicine in light of the new mechanistic philosophy of science. Biology and Philosophy, 34(4): 1–22.

Lorusso, L., & Bacchini F. (2015) A reconsideration of the role of self-identified races in epidemiology and biomedical research. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 52:56–64

Ludwig, D. (2019). How race travels: relating local and global ontologies of race. Philosophical Studies, 176(10), 2729–2750.

Spencer, Q.N.J, (2018) A racial classification for medical genetics. Philosophical Studies, 175(5):1013–1037.

Valles, S.A. (2012) Heterogeneity of risk within racial groups, a challenge for public health programs. Preventive Medicine, 55(5):405–408.

Vyas, D.A., Eisenstein, L.G., & Jones, D.S. (2020). Hidden in plain sight— reconsidering the use of race correction in clinical algorithms. New England Medical Journal, <u>https://doi.org/</u>10.1056/NEJMms2004740.

Yudell, M., Roberts, D., DeSalle, R., & Tishkoff, S. (2016). Taking race out of human genetics. Science, 351(6273): 564–565.

## **Understanding Causal Inference**

### David Papineau, Thomas Blanchard, and Toby Friend

The current pandemic has focused attention on the techniques used by epidemiologists and other non-experimental scientists to infer causal hypotheses from correlational data. These techniques raise a range of questions. A first issue is how they work. In response, a number of philosophers have now analysed these "Bayesian network" techniques and regimented the principles of inference involved (Spirtes et al 1993, Pearl 2000, Woodward 2003, Williamson 2005). Still, even after this regimentation, philosophical challenges remain. Given that causation transcends correlation, what makes it possible to infer the former from the latter? What is it about causation that allows such inferences to proceed?

Ideally, we might hope that an analysis of causation will explain why causal structures display themselves in the correlational patterns that non-experimental scientists use to identify them. Yet none of the mainstream accounts of causation—counterfactual, process, dispositional, regularity—cast any light at all on why the correlational techniques work.

By contrast, probabilistic and interventionist theories of causation (Reichenbach 1956, Suppes 1970, Spohn 2001, Glymour 2003, Woodward 2003, Schutz and Gebharter 2016) posit a constitutive connection between causation and correlational structures. But this arguably makes the relation too tight—correlational structures can be misleading about causal links—and moreover has trouble explaining single-case causation.

This symposium proposes to explore a rather older approach to causal inference, one that appeals to systems of structural equations satisfying independence requirements on their exogenous terms. The ideas that these equational structures underlie causal relationships goes back to H.A. Simon in the 1950s, and more recently has been explored by Judea Pearl, but has generally been eclipsed by the more recent focus on Bayesian networks, and has received little attention from philosophers. The attraction of the structural equations approach is that it allows us to view the correlational patterns as evidence for causal relationships rather than constitutive of them. It also promises to explain how facts of single-case causation relate to the kinds of causal hypotheses inferred using Bayesian network techniques.

Causation is a central category in the sciences and has long been subject to philosophical scrutiny. Yet the rise of correlational causal inference techniques over the last century has been little attended to by philosophers concerned with the nature of causation. If the structural equations approach can help to bridge the gap between the inference techniques of practising scientists and the philosophical understanding of causation, this will mark a significant advance for the philosophy of science.

David Papineau's talk will argue that systems of structural equations with independent exogenous terms underlie all causal relationships, and will aim to show how the central principles of causal inference follow from this. In particular he will show that the causal Markov condition is guaranteed by the nature of causation, while the faithfulness condition can be relied on as a reliable rule of thumb.

Thomas Blanchard will ask whether this account of the causal inference principles is any improvement on that offered by proponents of the "interventionist" approach to causation,

and will further question whether the independence of exogenous terms has the power uniquely to pin down causation relations.

Toby Friend will argue that there are features of causation that cannot be accommodated by an account of causation in terms of structural equations, and also that this account is illsuited to accommodate the transmission of conserved quantities and explanatory asymmetries in dynamic explanations.

## Aetiology of the Replication Crisis

### Edouary Machery, Sophia Crüwell, and Felipe Romero

The replication crisis describes an ongoing phenomenon, particularly in the social andmedical sciences, in which there is a high frequency of unsuccessful replications which is surprising to the field in question. Large scale replication failures are seen as problematic because they appear to undermine trust in science, particularly in sciences that enjoy media attention and tend to be highly relevant to policy decisions, such as medicine and psychology. In response to this perceived crisis, metascience or meta-research has emerged as a new field, working mostly empirically to identify and investigate underlying causes, and to develop and evaluate solutions (Hardwicke et al., 2020). Many possible causes behind individual and field-wide replication failures have been proposed. Commonly considered causes – both distal and proximal – include cognitive biases (Bishop, 2020; Machery, 2021), misuse of statistical methods or use of questionable research practises (Simmons, Nelson, & Simonsohn, 2011), lack of experimenter skill (Baumeister, 2016), publication bias (Ferguson and Heene, 2012), and perverse incentives (Higginson & Munafò, 2016).

This crisis and related matters are clearly timely and important topics in the sciences, but have also become cutting edge issues in current philosophy of science. Over the last few years, philosophers of science have increasingly engaged with these ongoing discussions. They have tried to clarify, for instance, the concept of replication by investigating questions such as: what is a replication (Machery, 2020), what is its role in science (Fletcher, 2021), and who should do replication labour (Romero, 2018)? Philosophers of science have also focused on underlying problems of the crisis, for example arguing that replicability issues are inevitable given the reward structure of science (Heesen, 2018); that a high rate of replication failure is compatible with good science (Bird, 2020); or that the replication crisis revolves not around replication but questionable research practices (Feest, 2019). While the topic is becoming more popular in philosophy of science, there remains a need for further philosophical treatment of issues surrounding the replication crisis. Key to understanding the replication crisis, whether from a metascientific or a philosophical point of view, is to understand its aetiology; this is also an important prerequisite for finding solutions to the perceived crisis that treat not only its symptoms but also any underlying problems or causes.

In this symposium, we thus want to focus generally on topics surrounding the aetiology of the replication crisis. "Aetiology" is here understood quite broadly as concerned with the causes of replication failure, both on the level of individual replications and on a system-wide level, i.e., the replication crisis. The upshots of our discussion of this topic should be applicable to replication crises in different fields, but it is important to note that our shared focus is on psychology in particular. To briefly summarise our symposium: Machery puts a popularly proposed cause of the replication crisis – perverse incentives – under the microscope to argue that it is less an explanation than a facade; Crüwell makes use of metascientific evidence to argue that the replication crisis in psychology is not a crisis of low replicability but a crisis due to inappropriate inference; Romero looks towards a possible future path to investigating the aetiology of the replication crisis by examining the emergent empirical field of metascience.

More specifically, Edouard Machery's talk focuses on a specific candidate explanation of poor replicability, namely the "perverse-incentives hypothesis". He examines formal models of scientific incentive structures to show that they are implausible and transparent, and lack robustness. Machery concludes that we do not currently have good reason to assume that perverse incentives are a central cause of poor replicability. Sophia Crüwell takes empirical evidence from the replication crisis and puts this into the context of a Bayesian framework to argue that the underlying problem is one of inference not replication. Crüwell argues that, once we adapt our inferences, the large-scale replication failures seen in e.g. social psychology will not be surprising anymore. Felipe Romero finishes the symposium by shining a critical light on metascience as an emerging field, exploring its conceptual origins and challenging the increasingly popular view of metascience as an authoritative field to study the causes and solutions to the replication crisis.

We hope that our symposium provides interesting insights into the replication crisis and its aetiology as a crucial developing topic in both contemporary science and philosophy of science. At the BSPS 2021 Annual Conference, only one talk was explicitly concerned with this subject matter, and there was no dedicated replication crisis symposium. Bringing our three talks together in one specialised symposium will provide a platform for further discussion based on our different perspectives. Our symposium convenes a group of researchers that is diverse in several dimensions: our speakers range from very early-career

to more senior researchers; we work in the Netherlands, the UK and the US; and our academic work and backgrounds represent various approaches in philosophy of science, including empirically-informed philosophy, analytic philosophy, but also metascience, psychology and the cognitive sciences.

#### References:

Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners,

losers, and recommendations. Journal of Experimental Social Psychology, 66, 153-158.

Bird, A. (2020). Understanding the replication crisis as a base rate fallacy. The British Journal for the Philosophy of Science.

Bishop, D. V. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. Quarterly Journal of Experimental Psychology, 73(1), 1-19.

Feest, U. (2019). Why replication is overrated. Philosophy of Science, 86(5), 895-905.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. Perspectives on Psychological Science, 7(6), 555-561.

Fletcher, S. C. (2021). The role of replication in psychological science. European Journal for Philosophy of Science, 11(1), 1-19.

Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S. N., & Ioannidis, J. P. (2020). Calibrating the scientific ecosystem through meta-research. Annual Review of Statistics and Its Application, 7, 11-37.

Heesen, R. (2018). Why the reward structure of science makes reproducibility problems inevitable. The journal of philosophy, 115(12), 661-674.

Higginson, A. D., & Munafò, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. PLoS Biology, 14(11), e2000995.

Machery, E. (2021). A mistaken confidence in data. European Journal for Philosophy of Science, 11(2), 1-17.

Machery, E. (2020). What is a replication?. Philosophy of Science, 87(4), 545-567.

Romero, F. (2018). Who should do replication labor?. Advances in Methods and Practices in Psychological Science, 1(4), 516-537.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science, 22(11), 1359-1366.

### **Quantum Probability Beyond Quantum Measurement**

#### Jean Baccelli, Sam Fletcher, Ben Feintzeig, and Jer Steeger

There is wide agreement on the values of the probabilities that quantum mechanics predicts for the outcomes of experiments, but the interpretation of these probabilities is a source of much controversy. What are these probabilities supposed to represent, exactly? Many believe quantum probabilities represent something objective—for example, relative frequencies of measurement outcomes. Others have proposed that quantum probabilities measure only agents' degrees of personal belief that certain events will occur and might differ between agents with different personal beliefs. We call personalists those who think quantum probabilities represent only agents' degrees of belief, owing to their emphasis on personal opinion; we call dualists those who allow for two sorts of probability: personal beliefs and "correct" probabilities, identifying the beliefs that any agent ought to have. Should probabilities in quantum mechanics be interpreted along personalist or dualist lines?

Discussions of probability in quantum theory often assume a particular approach to the infamous quantum measurement problem, that the way standard quantum theory treats measurement is at best incomplete and at worst inconsistent. In order to explain how quantum systems interact with measuring devices to produce outcomes in a complete and consistent way, researchers have developed different interpretations or modifications of quantum theory. For example, QBists (deriving from "Quantum Bayesianism," as personalism is often called "Subjective Bayesianism"; Fuchs and Schack, 2017) assert that quantum mechanics does not directly describe the behavior of physical systems, but merely agents' degrees of belief about outcomes of measurements upon them. QBists reinterpret the state of a physical system as a collection of predictions about future measurements through results like the quantum de Finetti theorem (Caves et al., 2002), but many of their arguments are closely tied to their austere approach to measurement. In contrast, the Everettian or many-worlds interpretation depicts a universe that "branches" at every interaction so that every possible outcome to a measurement occurs in its own "world." An approach due to Deutsch (1999) and Wallace (2003, 2007, 2012) provides Everettians with a dualist understanding of probabilities, but it too ostensibly rests on controversial assumptions about quantum measurement. Is it possible to interpret quantum probability without committing to some particular solution to the measurement problem?

We propose that there are several compelling justifications for dualism that do not commit to a particular solution to the quantum measurement problem. In this symposium, we discuss two measurement-neutral ways to justify the dualist's objective probabilities: applying the method of arbitrary functions and grounding them in state-space symmetries. Then we discuss one measurement-neutral way to link these chances to subjective probabilities, using new perspectives from decision theory. Our arguments are significant for three key reasons. First, while some physicists and philosophers ascribe to personalism, many formal epistemologists and philosophers of science believe dualism is a cogent alternative worthy of investigation (Howson and Urbach, 2006; Lewis, 1980, 1994; Pettigrew, 2012). Second, dualism seems the near-consensus view among physicists outside quantum foundations; it appears tacitly assumed in textbook presentations of quantum theory. (See, for example, Binney and Skinner, 2014, §6.3; Bransden and Joachain, 2000, §14.1; Merzbacher, 1998, §15.5; Sakurai and Napolitano, 2010, §3.4; Sudbery, 1986, §5.1; or Townsend, 2012, §5.7.) Third, the importance of an interpretation of quantum probability that does not depend on a particular solution to the measurement problem is already recognized in the literature. Recently, researchers have extended personalism about probability beyond QBism by recasting some key results in terms of "generalized probabilistic theories" (Holik et al., 2016), which do not depend on a particular interpretation of quantum measurement (or even the formal structure of quantum theory). But this work introduces an asymmetry to the status of personalism and dualism in the literature on quantum probability, an imbalance that we seek to redress.

This symposium presents work yielded by the first half of a two-year, NSF-funded project, "Quantum Epistemology Beyond Quantum Measurement" (Grant #2043089). The project is led by Ben Feintzeig (co-PI), Sam Fletcher (co-PI), and Jer Steeger (Postdoctoral Scholar); the paperspresented below are works by each of these three authors. The paper by Fletcher is a collaboration with Jean Baccelli, a non-presenting contributor.

#### Paper 1: Quantum probability via the method of arbitrary functions,

In quantum mechanics, the so-called Born rule assigns probabilities to outcomes of measurements on a physical system. Many have attempted to explain this probability assignment from underlying interpretive or foundational assumptions. For example, the Duetsch-Wallace approach (Deutsch, 1999; Wallace, 2007, 2012) uses a decision-theoretic argument to derive the Born rule within Everettian quantum mechanics, and the approach of Durr, Goldstein, and Zangh`i (1992) employsan argument based on the concept of typicality (along with the law of large numbers) to derive the Born rule within Bohmian mechanics. I consider an approach to deriving the Born rule that does not depend on the interpretive assumptions of any particular solution to the measurement problem, and instead draws on a resource from classical probability theory known as the "method of arbitrary functions."

The method of arbitrary functions is an attempt to show how "objective" or "universal" probability distributions are determined by the dynamics of physical systems. The idea is that for a given dynamics, but for arbitrary distributions describing uncertainty over the initial conditions, the system evolves in some appropriate limit or approximation (e.g., long times) to the same final distribution. For example, Poincar'e (1896) showed that for arbitrary initial distributions, the classical dynamics describing a roulette wheel lead to equal probability for the outcomes as long as the initial velocity of the wheel is large enough. I will apply this method of arbitrary functions to some simple quantum mechanical systems, with the aim of deriving the Born rule from arbitrary initial distributions in measurement contexts. Along the way, I will consider two approximations: the limit of long times and the classical limit.

In this talk, my goal is to report on progress in this research program of applying the method of arbitrary functions to quantum mechanics. I will cover toy quantum models, as well as dynamical models of measurement setups. Then I will consider how obtained and conjectured results relate to other derivations of the Born rule within interpretations of quantum mechanics. And finally, I will argue that the method of arbitrary functions can fill a gap in a proposed solution to the measurement problem due to Landsman and Reuvers (2013).

#### Paper 2: One world is (probably) just as good as many, Jer Steeger

The aforementioned Deutsch-Wallace approach consists of roughly two steps: first, it argues that objective probabilities are grounded in symmetries of the quantum state space; then, it represents these probabilities as an agent's rational preferences over actions. This talk focuses on the first half of this approach. Wallace (2003; 2012) argues that it requires an Everettian approach to measurement. I find that this argument is unsound. I demonstrate a counter-example by applying one version of the Deutsch-Wallace theorem to the de Broglie-Bohm pilot wave theory. This talk is based on the paper of the same name, forthcoming in Synthese.

I focus onWallace's non-decision-theoretic derivation of objective probability values, which I call the symmetry theorem (2012, pp. 148–156). A key premise in this theorem is state supervenience, the assumption that chances supervene on the wave function  $\Psi$ . Roughly, Wallace suggests that Everettians ought to justify state supervenience by noting that  $\Psi$  captures all that exists. He then asserts that particles' configurations in pilot wave theories violate the premise by breaking symmetries in  $\Psi$ . But the precise values of configurations q in Bohm's theory are, indeed, "hidden" from agents—at least as a practical limitation on their ability to measure a system (Bohm, 1952a,b). Bohm's theory does not take an agent to observe precise particle configurations q directly, but rather indirectly and approximately via the particle's selection of a particular branch of the overall wavefunction  $\Psi$  of the system and the measuring device (Barrett, 2019). Thus, an agent only everhas approximate knowledge of q before measurement—an epistemic fact that I call q-ignorance.

Wallace uses David Lewis's (1980) principal principle to understand objective probability. Say an agent believes in a theory assigning chance values to events; then, the principal principle instructs an agent to set their personal probability for those events in accord with the chances, all else being equal. On Wallace's operational reading, only reliable background information can affect these assignments. However, with chance so understood, qignorance secures state supervenience. The Bohmian is free to adopt Wallace's operational approach to chance, and so they can use the symmetry theorem to derive Born-rule chances.

But is this justification just as good as the Everettian's? I think so, at least on Wallace's "Lewisian" semantics for many-worlds. On this view, system-talk corresponds to a particular branch of  $\Psi$  (2012, Ch. 7). To secure state supervenience for the Lewisian, Wallace appeals to self-ignorance, the assumption that an agent cannot reliably know the branch-identity fixing their future path in a branching event (2012, p. 150). But, I argue, this assertion is only plausible due to environmental decoherence, the process responsible for generating  $\Psi$ 's branching structure. Decoherence places a limit on an agent's reliable knowledge via a principle that I will call decoherence exclusivity. From this principle, self-ignorance follows. But this principle also implies state supervenience directly, rendering appeals to further physical or metaphysical details superfluous. Indeed, the Bohmian can use the principle, too, and it provides a good explanation of q-ignorance as a bonus. This argument yields a precise sense in which one world is (probably) just as good as many.

#### Paper 3: Quantum Dynamic Consistency

This talk focuses on the second step of the Deutsch-Wallace approach: the representation of quantum probabilities as an agent's preferences over actions. This step uses Lewis's principal principle to translate objective probabilities into the rational beliefs of agents. Then, following decision theory (Ramsey, 1931; Savage, 1954; von Neumann and Morgenstern, 1953), it represents the rational beliefs of agents—what probabilities they ascribe to various events—as an abstraction from their preferences over actions. Saying that an agent assigns a probability more than 1/2 to an event, for example, means that they would prefer betting on that event happening over betting on it not happening. It then argues that if an agent accepts the Everettian interpretation of quantum mechanics, their personal probabilities for the outcomes of quantum mechanical experiments will be exactly as the Born rule prescribes. Thus, the Deutsch-Wallace approach recovers chance by the degrees of belief that followers of the Everettian interpretation adopt, where the degrees of belief are interpreted as pragmatic guides to decision-making.

Like many matters in the interpretation of quantum theory, whether the Deutsch-Wallace approach is successful is a matter of controversy. Many criticisms have focused on whether the decision-theoretic axioms employed are really dictates of rationality (e.g., Dizadji-Bahmani, 2015; Jansson, 2016; Mandolesi, 2019), particularly since they are not the standard assumptions of decision theory (Jeffrey, 1965; Savage, 1954; von Neumann and Morgenstern, 1953). Instead, they are seemingly bespoke axioms particular to quantum theory, as are the details of how they constrain an agent's preference. Wallace (2012) has suggested that his "diachronic" approach to such dynamic decision making might be an innovation that even classical decision theory might fruitfully adopt. His particular approach contains assumptions that the agent is indifferent between states of affairs that differ by certain symmetries in the expression of the quantum state as it changes over time.

Despite these interesting claims to methodological novelty, however, the Deutsch-Wallace approach has not received significant attention from decision theorists. Yet through their lens, the approach in fact bears strong similarities with an older approach to the representation of degrees of belief, one little studied by philosophers but known to economists as the "dynamic consistency" approach to decision making (Epstein and Le Breton, 1993; Ghirardato, 2002; Hammond, 1988;Hammond and Zank, 2014; Machina, 1989). It complements the more traditional static approach, epitomized by Savage (1954). Like in the Deutsch-Wallace approach, the dynamic consistency literature seeks to show how an agent's preferences, constrained by considering certain multi-step decision problems equivalent by virtue of some "symmetries" or independence principles those problems obey, give rise to belief represented by a probability function. Yet Deutsch and Wallace, unaware of this literature (which we have confirmed in personal correspondence with Wallace), have developed their approach without the insights from the preexisting work on dynamic consistency.

We reconstruct the decision-theoretic Deutsch-Wallace theorem using the tools of the dynamic consistency literature. This clarifies which aspects of the Everettian interpretation and of rationality might be truly necessary for it.

#### References

Barrett, J. A. (2019). The Conceptual Foundations of Quantum Mechanics. Oxford University Press, Oxford.

Binney, J. and Skinner, D. (2014). The Physics of Quantum Mechanics. Oxford University Press,Oxford.

Bohm, D. (1952a). A suggested interpretation of the quantum theory in terms of "hidden" variables. I. Physical Review, 85(2):166–179.

Bohm, D. (1952b). A suggested interpretation of the quantum theory in terms of "hidden" variables. II. Physical Review, 85(2):180.

Bransden, B. H. and Joachain, C. J. (2000). Quantum Mechanics. Pearson Education Limited, Harlow, 2nd edition.

Caves, C. M., Fuchs, C. A., and Schack, R. (2002). Unknown quantum states: the quantum de Finetti representation. Journal of Mathematical Physics, 43(9):4537–4559.

Deutsch, D. (1999). Quantum theory of probability and decisions. Proceedings: Mathematical, Physical and Engineering Sciences, 455(1988):3129–3137.

Dizadji-Bahmani, F. (2015). The probability problem in Everettian quantum mechanics persists. British Journal for the Philosophy of Science, 66(2):257–283.

D<sup>"</sup>urr, D., Goldstein, S., and Zangh`ı, N. (1992). Quantum equilibrium and the origin of absolute uncertainty. Journal of Statistical Physics, 67(5-6):843–907.

Epstein, L. G. and Le Breton, M. (1993). Dynamically consistent beliefs must be Bayesian. Journal of Economic Theory, 61(1):1–22.

Fuchs, C. A. and Schack, R. (2017). Notwithstanding Bohr, the reasons for QBism. Mind andMatter, 15(2):245–300.

Ghirardato, P. (2002). Revisiting Savage in a conditional world. Economic Theory, 20:83–92.

Hammond, P. J. (1988). Consequentialist foundations for expected utility. Theory and Decision, 25(1):25–78.

Hammond, P. J. and Zank, H. (2014). Rationality and dynamic consistency under risk and uncertainty. In Machina, M. J. and Viscusi, W. K., editors, Handbook of the Economics of Risk and Uncertainty, volume 1, pages 41–97. Elsevier.

Holik, F., Fortin, S., Bosyk, G., and Plastino, A. (2016). On the Interpretation of Probabilities in Generalized Probabilistic Models, page 194–205. Springer.

Howson, C. and Urbach, P. (2006). Scientific Reasoning: The Bayesian Approach. Open Court, Chicago, 3 edition.

Jansson, L. (2016). Everettian quantum mechanics and physical probability: Against the principle of "state supervenience". Studies in History and Philosophy of Modern Physics, 53:45–53.

Jeffrey, R. C. (1965). The Logic of Decision. McGraw-Hill, New York.

Landsman, N. P. and Reuvers, R. (2013). A flea on Schrödinger's cat. Foundations of Physics, 43:373–407.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey, R. C., editor, Studies in Inductive Logic and Probability, volume 2, pages 263–294. University of California Press, Berkeley.

Lewis, D. (1994). Humean supervenience debugged. Mind, 103(412):473–490.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. Journal of Economic Literature, 27:1622–1688.

Mandolesi, A. L. G. (2019). Analysis of Wallace's proof of the Born rule in Everettian quantum mechanics II: Concepts and axioms. Foundations of Physics, 49(1):24–52.

Merzbacher, E. (1998). Quantum Mechanics. John Wiley & Sons, New York, 3rd edition.

Pettigrew, R. (2012). Accuracy, chance, and the principal principle. Philosophical Review, 121(2):241–275.

Poincare, H. (1896). Calcul des probabilit'es. Gauthier-Villars, Paris.

Ramsey, F. P. (1931). Truth and probability. In Braithwaite, R. B., editor, Foundations of Mathematics and Other Essays, pages 156–198. Kegan, Paul, Trench, Trubner, & Co., London. Originally written in 1926.

Sakurai, J. J. and Napolitano, J. (2010). Modern Quantum Mechanics. Addison-Wesley Publishing Company, Reading, 2nd edition.

Savage, L. J. (1954). The Foundations of Statistics. Wiley.

Sudbery, A. (1986). Quantum mechanics and the particles of nature: An outline for mathematicians. Cambridge University Press, Cambridge.

Townsend, J. S. (2012). A Modern Approach to Quantum Mechanics. University Science Books, 2 edition. Times cited: 316.

von Neumann, J. and Morgenstern, O. (1953). Theory of Games and Economic Behavior. Princeton University Press, Princeton.

Wallace, D. (2003). Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 34(3):415–439.

Wallace, D. (2007). Quantum probability from subjective likelihood: improving on Deutsch's proof of the probability rule. Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics, 38(2):311–332. Times cited: 114.

Wallace, D. (2012). The Emergent Multiverse: Quantum Theory According to the Everett Interpretation. Oxford University Press.

# Reliability in science: how is it established, and how is it lost?

#### Kirsten Walsh, Stephan Guttinger, and Peter Vickers

Science is seen by many as our most reliable source of knowledge about the world. Even though it can go wrong in the short term, in the long term it is assumed to produce, every now and then, evidence and insights that we can rely on.

However, we don't have a clear picture of how scientists arrive at evidence, empirical claims, and theoretical frameworks that are trustworthy and that can stand the test of time. What are the hallmarks of a reliable experimental arrangement or technique? How and why is a product of an error-laden process identified as a reliable result? And how can non-scientists assess whether a scientific theory is reliable? Is the process of answering these questions ultimately a matter of collecting more empirical evidence? Or is the consensus among scientists shaped by non-epistemic factors? If the latter, then is this a problematic aspect of science? These are the questions that this symposium will explore.

Questions about the reliability of scientific output have formed the focal point of extensive work in philosophy, history and sociology of science, but they have gained renewed attention recently because of the worry that science is in the middle of a "replication crisis": individual experiments and whole studies, including those published in leading journals such as Nature, Cell, or Science, cannot be reproduced by other research groups. This, to many, indicates that things are not right in contemporary science. Science seems to have lost some of its reliable character.

One of the key factors behind this problem has been identified as the rise of so-called "questionable research practices" (QRPs). These include the potentially flawed use of statistical tools, the absence of "correct" experimental controls, or the selective publishing of results. The idea is that a deterioration of the research process has led to a deterioration of the product. One solution to the crisis that several authors have proposed is therefore to reduce the prevalence of QRPs. What this solution illustrates is a specific vision of what reliability in science consists of and how it can therefore be assessed and managed: reliability lies in the quality and the structure of individual experimental processes.

Whilst there certainly is a close connection between process and product, the above approach faces several challenges. One of these comes to the fore in the very word "questionable": there is a deep uncertainty about the actual quality of the practices of interest. One the one hand, QRPs are part of the canon of accepted methods. That is why the peer review process often lets them pass. On the other hand, at least some members of the relevant research community have serious doubts about these practices and their reliability. This uncertainty is reflected in the fact that there is no consensus on how the label "questionable" should be applied. From discussions over p-values to debates over the quality of an antibody or the correct controls to be implemented in an experiment – the reliability of the materials and methods used in science is often a matter of ongoing debate within the community. But how exactly are these debates settled? Are they settled by accumulating more empirical data, as some have argued is the case when researchers decide whether something is an artefact or not (Hudson 1999; Craver and Dan-Cohen

2021)? Or are there other factors at work, such as sociological factors (e.g., Collins 1985; Rasmussen 1993)? And if the latter is the case, are these elements problematic?

Ultimately, the negotiations that take place in these spaces of uncertainty will always be complex and multifaceted. The aim of this symposium is to take a closer look at the question of reliability in science and explore in more depth how it is established and lost. To this end the above questions will be approached from three different angles.

Kirsten Walsh will turn to the history of experimentation. Looking at Newton's correspondence with the Jesuits, she will analyse how the debate about the reliability of Newton's prism experiments unfolded. Her analysis will show that, faced with a "replication crisis", Newton had to balance his attitude that others must simply 'try the experiment; and judge' with his interest in managing community perception of his work. What comes to light in Newton's strategy is a complex mixture of worries about empirical evidence, public opinion, and community membership.

Stephan Guttinger will look at a case study from the molecular life sciences to analyse how an experimental system comes to be labelled as "reliable" (and how it might lose that label again). More specifically, he will look at a method called "Far Western blot" (FWB), a variation of the well-known Western blot. The FWB is interesting because it was designed to do something that blotting techniques were never meant to do, namely, to study proteinprotein interactions. This expansion into functional analysis was also part of the problem for this assay, as most researchers saw it as error-prone; it was not deemed to be a reliable process for making the inferences it was designed for. But the story of the FWB might have taken a turn when a new class of proteins was discovered. Analysing the history and the fate of this assay, Guttinger will argue that the label "reliable" is assigned based on a complex negotiation between evidential support, metaphysical assumptions, and the power of embedded practices and norms.

Peter Vickers will consider an alternative way of identifying reliability in science, having to do with 'scientific consensus' (inspired by Oreskes 2019). This approach avoids many examples, such as those at the heart of the replication crisis, because the majority of such examples don't enjoy a strong scientific consensus. However, it invites other examples, where a 'scientific consensus' seemingly did get overturned. Vickers' talk sets out to identify the consensuses that are (super-) reliable, distinguishing them from consensuses that are potentially unstable. The approach is used to motivate a new theory of 'scientific facts'.

#### References:

Collins, H. (1985). Changing order: Replication and induction in scientific practice. University of Chicago Press.

Craver, C.F., and Dan-Cohen, T. (2021) Experimental Artifacts. British Journal for the Philosophy of Science.

Hudson, R. G. (1999). Mesosomes: a study in the nature of experimental reasoning. Philosophy of Science, 66(2), 289-309.

Oreskes, N. (2019). Why trust science? Princeton University Press.

Rasmussen, N. (1993). Facts, artifacts, and mesosomes: Practicing epistemology with the electron microscope. Studies in History and Philosophy of Science Part A, 24(2), 227-265.

# Assisted Discoveries, Understanding and the Aims of Science

#### Milena Ivanova, Mike Stuart, and Rune Nyrup

In the last years we have seen an increased interest in the aims of science. Does science aim at increasing truth, empirical adequacy, understanding or some other epistemic good? This debate was historically motivated by the role that highly abstracted and idealised models playin advancing scientific discovery (Bangu 2017; Baumberger 2019; Boon 2009; Boumans 2009; Boyd 2020; De Regt, Leonelli, and Eigner 2009; De Regt 2015; 2017; Dellsén 2020; Elgin 1993; Elgin 2017; Potochnik 2017; Khalifa 2013, Khalifa 2017). This symposium aims to explore the implications for the aims of science literature of a different recent development, namely the increased use of machine learning and other forms of artificial intelligence (AI) in scientific discovery. By discussing the kinds of epistemic goods machines can help us deliver in science, we will address a number of questions concerning the nature of assisted discoveries and its relation to scientific understanding. We investigate three main themes.

First, will assisted discoveries continue to have the same value—or the same kind of value as human discoveries? Or might they attain significance for different reasons? Human discoveries have for a long time been construed as a product of human genius and fine aesthetic sensibility, with many scientists compared to artists and celebrated for their beautiful experiments, inventions and creative and imaginative thinking. But what happens when a machine makes a discovery? Are these discoveries to be seen also as aesthetically pleasing and a product of creativity? We explore a number of assisted discoveries and the reaction of the scientific community to them to highlight several questions that need to be addressed. We argue that while assisted discoveries exhibit some of the features of human discoveries, there seems to be resistance to the idea that these discoveries are beautiful or the machine should be seen as creative, even though it leads to something highly novel and valuable. Such arguments are closely related to concerns about explainability in AI, i.e., the fact that we often only have a limited sense of how or why a complex algorithm generated a particular output. This is often framed as an ethical issue, but research communities also assign negative aesthetic value to the absence of explainability in the discovery. We consider features that could be construed as aesthetically pleasing about scientific discoveries and explore whether assisted discoveries can have any of these types of aesthetic value.

Second, insofar as assisted discoveries do generate understanding, to whom should we ascribe that understanding? If we take understanding to be some kind of mental state, the only plausible candidate would seem to be individual scientists. However, if understanding is construed in more pragmatic terms, as an ability to manipulate a system, this might be had by groups of scientists, such as research a lab, even if no individual scientist possesses it. If we take the latter option seriously, what does it imply about pragmatic understanding when it comes to discoveries made using AI? Who has pragmatic understanding in these cases: the scientists, the machines, or their combination?

Third, how should the development of algorithms for assisted discovery and the philosophical study of epistemic goods such as understanding relate to one another? As mentioned, while assisted discoveries certainly deliver outcomes of great value, many

researchers are dissatisfied with having knowledge yielded by 'black boxes', with limited understanding of how and why an algorithm contributed to the discovery. To address this concern, AI researchers have begun to develop methods for creating so-called explainable AI (XAI). But what kinds of explanation are we looking for here? Can we simply introduce existing philosophical accounts of explanation and understanding into this field of research, or do we need to transform them to capture the epistemic and normative concerns that XAI research seeks to address? We discuss the shortcomings of existing measures of explanation in XAI and the role philosophical models of explanation have played so far within the field. We consider an alternative, more promising approach, focused on identifying what types of understanding different researchers and groups are after in a given context and how AI systems can be designed to support those kinds of understanding. This in turn raises questions concerning whose values should be prioritised in contexts where different explanatory interests come into conflict and—more generally—what role non-epistemic values should play in XAI research.

Overall, then, exploring the philosophical implications of assisted discovery promises not just to yield insights about one of the most significant recent developments in how scientific research is structured; it also provides a new lens through which to consider broader questions about the nature of understanding and the aims of science.

#### References:

Bangu, Sorin. 2017. "Scientific Explanation and Understanding: Unificationism Reconsidered." European Journal for Philosophy of Science 7 (1): 103–26. https://doi.org/10.1007/s13194-016-0148-y.

Baumberger, Christoph. 2019. "Explicating Objectual Understanding: Taking Degrees Seriously." Journal for General Philosophy of Science 50 (3): 367–88. https://doi.org/10.1007/s10838-019-09474-6.

Boon, Mieke. 2009. "Understanding in the Engineering Sciences:: Interpretive Structures." In Scientific Understanding: Philosophical Perspectives, edited by Henk de Regt, Sabina Leonelli, and Kai Eigner, 249–70. <u>https://doi.org/10.2307/j.ctt9qh59s.16</u>.

Boumans, Marcel. 2009. Understanding in Economics: Grey Box Models. Edited by Henk Regt. Boyd, Kenneth. 2020. "Environmental Luck and the Structure of Understanding." Episteme 17 (1): 73–87. https://doi.org/10.1017/epi.2018.18.

De Regt, Henk. 2015. "Scientific Understanding: Truth or Dare?" Synthese 192 (12): 3781-97. https://doi.org/10.1007/s11229-014-0538-7.

----. 2017. Understanding Scientific Understanding. New York: Oxford University Press.

De Regt, Henk, Sabina Leonelli, and Kai Eigner, eds. 2009. Scientific Understanding: Philosophical Perspectives. University of Pittsburgh Press.

Dellsén, Finnur. 2020. "Beyond Explanation: Understanding as Dependency Modelling." The British Journal for the Philosophy of Science. https://doi.org/10.1093/bjps/axy058.

Elgin, Catherine. 1993. "Understanding Art and Science." Synthese 95: 13–28.

Elgin, Catherine Z. 2017. True Enough. Cambridge, MA: MIT Press.

Khalifa, K. 2013. "The Role of Explanation in Understanding." The British Journal for the Philosophy of Science 64 (1): 161–87. https://doi.org/10.1093/bjps/axr057.

Khalifa, Kareem. 2017. Understanding, Explanation, and Scientific Knowledge. New York: Cambridge University Press.

Potochnik, Angela. 2017. Idealization and the Aims of Science. University of Chicago Press. https://doi.org/10.7208/chicago/9780226507194.001.0001.

## New Mechanism and Alternative Causal Concepts in Biology: Processes and Pathways

### Olesya Bondarenko, Tyler Brunet, John Dupre, Giovanni Boniolo, and Raffaella Campaner

The new mechanist philosophy, or New Mechanism (NM)–most notably represented by Glennan (1996, 2017), Machamer, Darden and Craver (2000), and Bechtel (2011)–is a family of philosophical views concerning the nature of causation. Roughly, these views share the idea that the phenomena of causation in target systems should be broken down into components and operations on those components, or into entities and activities, that are connected in such a way that they have a beginning, an end, and display a kind of regularity in intermittent connections. Moreover, NM holds that such properly organized entities and activities – in short, individual mechanisms – can explain how phenomena are produced. The new mechanist philosophy has been of special interest in molecular biosciences (molecular biology and the -omics in particular), since these sciences often aim to identify mechanisms.

However, the broad scope of phenomena investigated in the biological sciences are not always so easily fit into the framework of NM; sometimes a mechanistic approach to such phenomena seems Procrustean. In particular the NM has encountered criticism over its applicability to a number of biological cases: over the long periods of time investigated in evolutionary biology; on the cusp of biological research where it encounters phenomena from the social sciences; and, ironically perhaps, in some of the complex and dynamic pathways and processes of molecular biology that NM was in part initially developed to account for.

This has spurred an interest in investigating alternative causal and explanatory concepts that may be of use within biology and its philosophy. There are a number of other philosophically grounded metaphysical accounts of causation, with more or less direct connections to biology, such as counterfactual theories, interventionism and causal process theory. In this session we focus more narrowly on a cluster of causal concepts that we believe are particularly important in biological theory and practice, including pathways and processes. Pathways and processes differ from mechanisms in at least a few interesting respects: they do not necessarily come with an analysis of phenomena into components and operations (entities and activities), they may not display the same sort of regularity and indeed may be fundamentally dynamic or unstable, and they may occur over timescales and in analytic contexts unusual of paradigm mechanisms.

Even though this symposium is partly motivated by the limitations of NM in specific research contexts, this does not necessarily mean that it should be supplanted by a different causal framework. In fact, we believe it is sometimes possible for the causal concepts under discussion to either co-exist or to be integrated in one way or another. One of our aims, therefore, is to examine these various possible relationships between mechanism and its alternatives such as process and pathway.

Each of the papers in the symposium addresses the relationship between mechanisms and processes/pathways from a different theoretical angle. We begin with the metaphysics and

progress to a discussion of in-practice and epistemological issues. John Dupré, one of the architects of the recent turn to processual concepts in the philosophy of biology, will discuss the idea of a process-based account of causation, with some consequences for the mechanistic (and nomothetic) understanding of causation. In another paper, Tyler Brunet will discuss the possibility of reconciliation or compromise between our metaphysical accounts of mechanistic and processual causation, focusing specifically on examples of mechanisms from molecular biology.

A paper by Giovanni Boniolo and Raffaella Campaner argues in favour of a processual epistemology that focuses on the notion of a molecular pathway. The authors also offer a logical framework for investigating pathways in biology. Finally, Olesya Bondarenko, drawing on examples from integrative research at the interface of biology and social science, examines whether the causal concept of mechanism ought to be replaced in this context with the concept of a pathway or a cascade, both recently elaborated by Lauren Ross. Even though, as the author claims, biosocial research would benefit from reorienting it away from mechanisms and toward these alternatives, there is still a place for mechanisms within a pluralistic framework.

## Function in the light of frequency-dependent selection

#### Samir Okasha, Tim Lewens, and Jonathan Birch

How to make sense of functional statements is a classic topic in the philosophy of biology. Such statements are ubiquitous in biology, as when we say that the function of a cactus's spines is to deter herbivores, or the function of bird-of-paradise's dance is to attract mates. Other natural sciences have little use for functional statements. Geologists do not talk about the function of glaciers nor chemists about the function of covalent bonds. Why then is biology different and how should function-talk be understood? This question is the point of departure for the large literature on biological function that emerged in the 1960s and continues to burgeon today. One well-known idea is that the function of a biological trait is determined by facts about evolutionary history. On this view, what makes it true that the function of the bird-of-paradise's dance is to attract mates (rather than to attract predators, for example) is that it is because the dance serves to attract mates that natural selection led the dance to evolve in the first place. This is the "etiological" or "selected effect" (SE) theory of biological

function, versions of which are defended by philosophers including Millikan (1984), Neander (2017), Garson (2019) and others. The SE theory has a number of advantages. It offers a clear basis for the distinction between a trait's function and its "unintended side effects", such as the bird-of-paradise's dance attracting predators as well as mates. It respects the fact that citing an item's function often serves to explain why that item exists. It yields a neat account of the analogy between biological functions and the functions of man-made artifacts. It justifies the view that functional statements in biology are both scientifically respectable and fully objective, resting on a secure naturalistic basis. And finally, the SE theory makes good sense of much actual biological usage, particularly in evolutionary contexts. For these reasons, the SE theory of function has achieved the status of nearorthodoxy in philosophy of biology. Even its critics typically allow that the SE theory gives the correct account of some biological function statements (often while arguing that the ahistorical "causal role" theory is needed to account for others). Moreover, the SE theory has found application beyond the philosophy of biology, as it underpins most versions of teleosemantics, the project of trying to naturalize representational content by reducing it to biological function (e.g. Shea 2019). In an important recent paper, J. Christie (et al.) (2022) argue that the SE theory of function runs into a trouble because it is implicitly premised on an idealized form of natural selection that ignores many real-world complexities. Equating a biological trait's "selected effect" with the causal explanation of why the trait evolved, which they see as the signature feature of the SE theory, only works in the simplest cases, Christie (et al.) argue, where the environment is constant and evolutionary change involves a simple monotone increase in the frequency of the "fittest" trait, driven by natural selection. But under more complex evolutionary dynamics, arising from factors such as environmental stochasticity or frequency-dependent selection, the SE theory does not work, Christie (et al.) argue. This has negative consequences for the teleosemantic project, they claim, since the mode of natural selection operative in the evolution of signalling is unlikely to be of the simple sort.

The Christie (et al.) paper throws up three important issues. Firstly, how if at all should we understand the notion of biological function under frequency-dependent selection?

(Selection is frequency-dependent when the relative selective value of a trait, vis-à-vis other traits, depends on how common it is in a population.) This important issue has been strangely overlooked in the literature, considering how common frequency-dependence is in nature and how central it is to evolutionary theory. Secondly, if it is true that standard formulations of the SE theory of function presuppose a simple form of natural selection, what are the prospects of generalizing that theory to the case of frequency-dependence? Thirdly, is the teleosemantic project compromised by its reliance on an underlying theory of biological function that lacks generality? Our proposed workshop will address these questions. The three participants are well-placed to make progress on them. Samir Okasha has previously examined how the notion of adaptation should be understood under complex evolutionary dynamics (Okasha 2018). Tim Lewens has written extensively on biological function and has critiqued the SE theory (Lewens 2004, 2007). Jonathan Birch has written on social evolution, where frequency-dependent selection is central, and on teleosemantics and signalling (Birch 2014, 2017). The workshop will be highly focused, with all three participants responding to each other's arguments. Given the centrality of the topic, we anticipate that the workshop will be of interest to a large number of BSPS attendees.

References:

Birch, J. (2014) "Propositional Content in Signalling Systems", Philosophical Studies 171, 493–512. Birch, J. (2017) The Philosophy of Social Evolution, Oxford University Press.

Christie, J. R., Brusse, C., Bourrat, P., Takacs, P. and Griffiths, P. E. (2022) "Are Biological Traits Explained by their 'Selected Effect' Functions?", Australasian Philosophical Review (forthcoming).

Garson, J. (2019) What Biological Functions Are and Why They Matter, Cambridge University Press.

Lewens, T. (2004) Organisms and Artifacts: Design in Nature and Elsewhere, MIT Press.

Lewens, T. (2007) "Functions" in M. Matthen and C. Stephens (eds.) Handbook of Philosophy of Biology, Elsevier, 537-559.

Millikan. R. (1984) Language, Thought and Other Biological Categories, MIT Press. Neander, K. (2017) A Mark of the Mental: In Defense of Informational Teleosemantics, MIT Press.

Okasha, S. (2018) Agents and Goals in Evolution, Oxford University Press.

Shea, N. (2019) Representation in Cognitive Science, Oxford University Press.