1838

UNIVERSITY OF BELGRADE
FACULTY OF PHILOSOPHY

The Third Conference of East European Network for Philosophy of Science

EENPS 2021

University of Belgrade
Department of Philosophy
Faculty of Philosophy
Belgrade, Serbia
June 9-11, 2021

Conference schedule

Jun 09

10:50-11:00 Opening

11:00-12:00 Keynote talk: Anna Alexandrova (Cambridge): Social Science: A Constructivist Account

12:30-15:00 Three parallel sessions (General, Medicine, Cognitive)

Track 1: General (chair: Dusko Prelevic)	**Track 2: Medicine (chair: Maria Laura Ilardo)**	**Track 3: Cognitive (chair: Borut Trpin)**
Işık Sarıhan: Disagreement and Progress in Philosophy and in Empirical Sciences	Gabriela-Paula Florea: Metacognition in Medicine	Błażej Skrzypulec: Tactile field and the dual nature of touch
Lilia Gurova: The uses of truth: Is there room for reconciliation of factivist and non-factivist accounts of scientific understanding?	Mariusz Maziarz: The meaning of causal claims in biomedicine and its implications for evidence-based medicine	Paweł Zięba: Seeing colours unconsciously
Richard David-Rus: The case is not closed	Johannes Findl and Javier Suárez: Predicting and Understanding COVID-19: the role of the IHME model	Matias Osta-Vélez: Inference and the structure of mental representations
Alik Pelman: Is Confirmation Inductive?	Martin Zach: Modeling in Biomedicine: Extending the notion of models	Marco Facchin: Extended predictive minds: do Markov Blankets really matter?
	Marek Pokropski: Integrating first-person data with neuroscience. A case of migraine aura research	Vanja Subotić: Deep Convolutional Neural Networks and How-Possibly Explanations in Cognitive Neuroscience

15:00-17:00 Lunch break

17:00-19:30 Parallel session (Biology) and Symposium 1

Track 1: Biology (chair: Martin Zach)

Iñigo Ongay de Felipe: Biological organismality and individuality in the Extended Evolutionary Synthesis

Predrag Šustar and Zdenka Brzović: Function Acquisition in Genomics

Mustafa Yavuz: Vitality: A Multi-Layered View

David Villena: Modules as an adaptationist discovery heuristic

Symposium 1: History of formal logic in Eastern Europe

Zuzana Haniková: Vopěnka's Alternative Set Theory within Twentieth Century Mathematics

Walter Dean and Máté Szabó: From Elementary School to Epsilon3, the Early History of Elementary Functions

Constantin C. Brîncuş: The Plurality of Logics, Between Human Thinking and Formal Systems

Jovana Kostić, Katarina Maksimović and Miloš Adžić: From Recursion to Deduction, Two Strands of Modern Logic in Serbia

Jun 10

10:00-12:30 Three parallel sessions (Economics-Social Sciences, Social, Physics)

Track 1: Economics/Social Sciences (chair: Barbara Osimani)	**Track 2: Social (chair: Lukas Bielik)**	**Track 3: Physics (chair: Slobodan Perovic)**
Stevan Rakonjac: Where does causal knowledge in macroeconomics come from?	Milan Urošević: On politics and social science – the subject-object problem in social science and Foucault's engaged epistemology	Marek Woszczek: Quantum contextuality and ontic indefiniteness
Aleksander Ostapiuk: Use and abuse of Weber's methodology by economics	Germán Hevia Martínez: The self-fulfilling prophecy machine. Beyond the science/technology distinction	Cristian López: Time's Direction at the Plank Scale
Jaana Eigi: On expertise and democracy: using philosophy of science when considering Harry Collins and Robert Evans' Owls	Juho Lindholm: Practices as Social Knowledge	Louis Vervoort: The hypothesis of "hidden variables" as a unifying principle in physics
Maria Laura Ilardo: Drug Agencies as Science Diplomats	Raphael Aybar: Thinking through artifacts	Athamos Stradis: The Origins of Observation
	Simon Blessenohl and Deniz Sarikaya: A Norm for Science Advice: Making Beliefs Accurate	Samuele Iaquinto and Claudio Calosi: Quantum Fragmentalism

12:30-14:30 Lunch break

14:30-17:00 Parallel session (General and cognitive) and Symposia 2 and 3

Track 1: General and cognitive (chair: Miljana Milojević)	Symposium 2: Free will and consciousness as a problem of philosophy of science: What can philosophy and science learn from each other?	Symposium 3: Philosophy Meets Psychology of Science
Vladimir Drekalović: Some Philosophical and Mathematical Aspects of the Concept of Infinity	Artem Besedin: Implicit Attitudes: A Challenge to Agency and Moral responsibility	Kaja Damnjanović: Simplicity bias or Ockham's razor?
Matteo De Benedetto: Taming Conceptual Wanderings: Wilson-Structuralism	Damir Čičić: Two Accounts of the Problem of Enhanced Control	Marko Tešić: Explaining away and the propensity interpretation of probability
Vassilis Livanios: Can all natural properties and relations be powers? The case of manifestation-relations	Anton Kuznetsov: Existence of Consciousness and Integrated Information Theory of Consciousness.	Nora Hangel: How do scientists generate scientific claims? Individual, collaborative, and collective accounts from scientific practice
Janko Nešić: Structural Realist Theory of the Self	Sergei Levin and Mirko Farina: How I Learned to Stop Worrying and Love Free Will	Tijana Nikitović: Challenges of scientific mobility facing early-career life scientists
	Andrew Mertsalov: The Relevance of the Interpretations of Quantum Mechanics to the Question of Free Will and Determinism	Vlasta Sikimić: Relationship between political and epistemic values of scientists
	Maria Sekatskaya and Gerhard Schurz: Alternative Possibilities and the Meaning of 'Can'	
	Louis Vervoort: Artificial Consciousness and Superintelligence in Robotics: How to Get There?	

17:30-19:00 EENPS members' meeting

Jun 11

10:00-11:00 Keynote talk: Gerhard Schurz (Düsseldorf): Abduction in Science and Metaphysics

11:30-14:00 Three parallel sessions (General, Explanation, General 2)

Track 1: General (chair: Borut Trpin)	**Track 2: Explanation (chair: Lilia Gurova)**	**Track 3: general 2 (chair: Matteo de Benedetto)**
Alexander Gebharter and Markus Eronen: Quantifying proportionality and the limits of higher-level causation and explanation	Stavros Ioannidis: Reconsidering multi-level mechanistic explanation	Michal Hladky: Neuroscience without Brains: Perspectives on In Silico methods
Jan Sprenger: Causal Attribution and Partial Liability: A Probabilistic Model	Daniel Kostic: Perspectivism and the Veridicality Problem in Non-causal Explanations	Matteo Colombo, Silvia Ivani and Leandra Bucher: Demographic variables predict differential sensitivity to inductive risks
Maria Ferreira Ruiz: Attributing causal specificity	Lukas Zamecnik: Topological explanation in system-theoretical linguistics	Michele Lubrano: Mathematical Proofs: Two Kinds of Unification
Samuel Fletcher: Causal Modeling as Counterfactual Semantics	Lukáš Bielik: Abduction and the Selection Mechanisms	Peeter Müürsepp: Practical Realism as Realism
Barbara Osimani: The (causal) structure of epistemic environments		Takaharu Oda: Berkeley's Pragmatist Theory of Causation in De motu

14:00-16:00 Lunch break

16:00-18:30 Parallel session (General) and symposium 4

Track 1: general/maths/logic (chair: Daniel Kostic)	Symposium 4: Trends in formal philosophy of science
Adam Kubiak: Was Jerzy Neyman a Perspectival Realist?	Jan Sprenger: Formal Methods and Scientific Philosophy
Joaquim Giannotti: Ground for Ontic Structuralists	Borut Trpin: Methodological triangulation and the value of epistemic modesty
Ludovica Conti: Abstraction's Logicality and Invariance	Sandro Radovanović: Using machine learning algorithms to predict the efficiency of experiments in high energy physics
Eric Raidl: Definable Conditionals	Vlasta Sikimić: Benefits and limitations of data-driven approaches in formal philosophy of science
Michael Shenefelt and Heidi White: Why Does Symbolic Logic Emerge During the Industrial Revolution?	

Jun 12

ic.SoAP 2021/EENPS satellite student event (organisers: Michal Hladky, Kyryll Khromov, Federico Donato)
10:00-10:30: Opening
10:45: Maximilian Schlederer: Bridge Laws are Analytic
11:30: Mary Peterson: Ducks, Rabbits, and Progress in Science
12:30: Lunch break
14:00: Sebastian Gil: What is the geometry of nature?
14:45: Maximilian Petrowitsch: Field's Nominalism and Infinite Cardinality
15:30: Coffee break
15:45: Jordan C. Myers: Vicious and Virtuous Selective Scrutiny
16:30: Garrett Credi: The Math Of Sleeping Beauty's Morning
17:15: Closing remarks

Contents

Keynote talks

Social Science: A Constructivist Account

Anna Alexandrova
University of Cambridge

The term ‘social science’ was first coined in the eighteenth century and ever since there has been a debate about what it is and what knowledge can be expected from it. Some accounts of social science call themselves naturalist and draw inspiration from natural sciences. Others are exceptionalist and they emphasise the distinctiveness of social knowledge. As different as they are, these two traditions share a strategy: they position social science with respect to either natural sciences or humanities and then, depending on this initial choice, they proceed to formulate what social scientists should do and how. I shall call this strategy contrastivism and argue that it has been a failure. Instead of positioning social science relative to the supposedly clear categories of natural sciences or humanities, I propose a constructivist account of social science. I develop a version of constructivism according to which social science is any mode of inquiry that serves priorities of any community that undertakes to understand and to improve itself.

Abduction in Science and Metaphysics

Gerhard Schurz
University of Düsseldorf

In this talk I argue for three theses and support them by theoretical reasoning and case studies:

1. The inference of abduction plays a crucial role for the discovery and the justification of theories, both in science and in metaphysics. There are two rationality conditions that distinguish scientific abductions from speculative abductions: achievement of unification and independent testability.

2. Abduction to common causes is of particular importance in science. The justification of metaphysical realism is structurally similar to scientific abductions: external objects are justified as common causes of perceptual experiences.

3. The reliability of common cause abduction is entailed by a principle of (Markov) causality. The latter principle has an abductive justification that is based on statistical phenomena.

Abstracts

Thinking through artifacts

Raphael Aybar

Thinking through Artifacts

Abstract

In this paper, I cast doubts on the view of explicit and common thoughts as tools or models (Sanches de Olivera, Raja and Chemero 2019). In contrast to thoughts, models are public and manipulable. Unlike tools, thoughts cannot be conceptualized in terms of affordances. I propose instead that the notion of *thinking through* things, tools, models or artifacts better characterizes extended forms of cognition. The notion of *thinking through*, as Material engagement theory points out (Malafouris 2018, 2019, 2020), stresses the active role cultural and material things play in cognition. I suggest this view is compatible in fundamental aspects with the Artifactual view of models in the philosophy of science (Knuuttila 2011, 2017), as both treat things and models as public cultural objects. More importantly, both depict a form of cognition that occurs by means of manipulating things and artifacts. This paper has three parts: the first introduces the manipulability of artifacts and their intersubjective character. The second presents and criticizes the view of thoughts as models and tools. The third elaborates the idea of thinking through artifacts in the context of scientific thinking.

This article extends Malafouris' (2018, 2019, 2020) notion of *thinking through* things to the domain of artifacts. Thinking, following Malafouris, is not just thinking *about* something. It is also thinking *with* and *through* things. These emphases make salient the importance of external objects for intelligent behavior. In particular, humans develop their cognitive capacities within cultures by interacting with others and manipulating external artifacts. All of them are extended forms of cognition. The material characteristics of things and artifacts and the forms of interaction are constitutive for cognitive tasks. That is particularly clear from a developmental perspective, where the attention is not on tasks but on the acquisition of cognitive abilities. Consider the importance of exposure to iconic gestures in children's acquisition of language. As Aussems and Kita (2020) point out, "seeing iconic gestures that depict verb referents helps children (a) generalize individual verb meanings to novel events and (b) learn more verbs from the same subcategory."

Gallagher and Ransom (2016) have already approached social interaction in terms of affordances. Social interaction is crucial for the higher-order cognitive capacities, including the capacity to manipulate artifacts since humans learn how to use them by participating in cultural settings. Once humans learn how to use them, they become perceivable opportunities for action (Gibson 1979). The idea of *thinking through artifacts* stresses the active role culturally established artifacts play in human reasoning. Thinking through artifacts has a prominent role in scientific practices. Scientific thinking does not occur merely within scientists' heads. While abstracting, idealizing or hypothesizing, scientists engage with tangible objects, such as models, tools or instruments. It is through using and manipulating them that scientists' reasoning and imagining realize. Scientists' cognitive tasks, such as deducing, inferring or categorizing, are not internal processes detached from the external artifacts scientists think

1

through. For instance, external representations "change the cost structure of the inferential landscape" (Kirsch 2010). Apart from supporting scientific thinking development, scientific artifacts have tangible and abstract aspects exploited in their manipulation, and they constrain the way scientists think about phenomena.

This paper elaborates the notion of thinking through artifacts relying on ideas from Ecological psychology (Gibson 1979), the Artifactual view of models (Knuuttila 2011, 2017), and Material engagement theory (Malafouris 2018, 2019, 2020). This paper has three parts: the first introduces the tangible and intersubjective dimensions of artifacts. The second presents and criticizes an alternative artifactual account of thoughts as tools. The third elaborates the idea of thinking through artifacts paying special attention to scientific thinking based on the Artifactual view of models (Knuuttila 2011, 2017) and Material engagement theory (Malafouris 2018, 2019, 2020).

I. Scientific models as tangible and intersubjectively available objects

Contemporary philosophical approaches to scientific models (e.g., Morrisson and Morgan 1999; Knuuttila 2011, 2017) treat them as mediating instruments or artifacts whose tangible and symbolic aspects enable scientists to think about problems in ways not otherwise possible. For the Artifactual view of models, scientific models are not primarily intended to represent reality. More importantly, the cognitive gain of models, or how they deliver understanding, comes from the processes of constructing and manipulating them. For instance, artificial neural networks (ANNs) models in the domain of cognitive neuroscience, following this perspective, do not provide knowledge by 'representing the brain.' They have many functions in scientific practices: through them, scientists transform raw into optimal data (e.g., Vieira, Pinaya and Mechelli 2017); run computational simulations that validate a model when fitting a summary of empirical data (e.g., Arbib, Plangprasopchok, Bonaiuto and Schuler 2013); or predict individuals' behavior using large data sets of neuromarkers (Rosenberg, Casey and Holmes 2018), to name a few examples.

As cognitive neuroscience aims to explain the neural substrate of cognition, one could believe that ANNs models deliver understanding of the brain by isolating underlying causal or non-causal factors relevant for neural machinery functions, that is, by virtue of sharing critical similarities or structural correspondences with biological neural networks (e.g., Hasson, Nastase and Goldstein 2020). However, ANNs models are used in domains where similarities between models and target systems do not exist. For instance, in food science (Huang, Kangas and Rasco 2007); wine technology (Baykal and Yildirim 2013); predictive analytics for marketing (Artun and Levin 2015); medical science (Patel and Goyal 2007), e.g., they have been used to identify risk factors for mortality associated with COVID-19 (Yu et al. 2020).

ANNs are manipulated and repurposed in different ways in each of the domains above. Any similarity between ANNs and the entities and processes of these domains seems to respond more to design choices than structural correspondences. Naturally, these applications could not be foreseen when the perceptron (see Rosenblatt 1958), the statistical model of the organization of cognitive systems that inspired current ANNs, was invented. However, scientific models are evolving artifacts. While constructing and manipulating them, scientists provide them with new intended uses. Through new intended uses, models transfer to new domains, transcending disciplinary boundaries (see Knuuttila and Loettgers 2014).

The philosophy of scientific modeling literature pays special attention to the processes of constructing and manipulating models (Morrison and Morgan 1999, in Knuuttila 2011). Such processes require models to "have a tangible dimension that can be worked on" (Knuuttila 2011). Modelers, e.g., need a media or vehicle (a computer) with sufficient computing power for training ANNs. They also need semiotic systems, such as programming languages, which are as essential as the vehicles and afford specific forms of intervention, e.g., programming languages and software such as R+ and SPSS are designed for statistics but may be useless for designing games.

The artifactual view treats models as manipulable and intersubjective available objects (Knuuttila 2017). As a case of models' manipulability, consider the multi-agent-based model created by Lewandowsky et al. (2019). This model simulates the spread of climate change deniers' opinions and how it affects the scientific community and the general public. The modelers aimed to explain the disproportionate impact of climate change deniers' different strategies on public opinion. One of these strategies, e.g., is to raise skepticism by creating a chimerical scientific debate, as if scientists were indeed debating the reality of anthropogenic climate change.

The model comprised three kinds of agents: scientists, deniers, and the public. Interestingly, the modelers manipulated those agents in the simulation by treating them as Bayesian agents that update their beliefs by continually evaluating their prior information. Why did they model those agents as Bayesians? Other models can also simulate social interaction, e.g., the Organizational Model for Normative Institutions (Li, Mao, Zeng and Wang 2008), or various game-theoretic models (Xia and LaiLei 2007). The rationale for approaching social agents as Bayesian is simple: Bayesian methods are optimal when dealing with belief-change processes, such as those in which beliefs change in communicative processes. In a Bayesian model, beliefs are represented as hypotheses and biases on climate change as priors with higher weights. Beliefs can be updated by acquiring new information and being maintained because of biases, e.g., an ordinary person overly exposed to deniers' opinions can develop a strong bias towards scientific opinions on climate change.

As stated above, models are intersubjectively available objects. Models neither are tied to a particular function nor, in many cases, domain-specific. Models are cross-disciplinary artifacts that travel and reproduce through disciplinary practices while acquiring new intended uses and extending their employability domains (Knuuttila and Loettgers 2016). Bayesian models' history (McGrayne 2011; see Lukeprog 2011 for a shorter version) illustrates models' intersubjective aspect. Sir Thomas Bayes originally developed methods to deal with uncertainty, motivated perhaps by religious convictions. Around two centuries later, Schlaifer (1959) repurposed Bayesian methods by applying them to business decisions. Later, in 1986, Raftery published an influential paper that motivated sociologists to use Bayesian models. Apart from being employed in these domains, Bayesian models are currently used in cognitive sciences, ecology, genetics, medicine, and many more (Schoot et al. 2021). These applications of Bayesian models have been possible only because Bayesian methods, as artifacts, are public objects available in textbooks and papers (nowadays in programming libraries), retransmitted and repurposed under generations by academic and non-academic actors.

II. Are thoughts models or tools?

The previous section examined the tangible and intersubjective aspects of models. The first implicates that scientists gain knowledge by manipulating models; the second that models are public objects rather than imaginary entities existing in scientists' minds. This section contends that the view of common and explicit thoughts as tools or models is inconsistent with the artifactual view of models because it dismisses these two aspects. In the words of De Oliveira, Raja and Chemero (2019), the view of thoughts as models or tools "is a straightforward application of artifactualism about scientific models to the domain of the mind." This view suggests that common or explicit thoughts such as 'I have to clean the table' are, in fact, tools that help in dealing with present and future scenarios.

De Oliveira, Raja and Chemero (2019) develop a non-representationalist account of explicit and common thoughts, which are instances of higher-order cognitive abilities. Similar attempts have succeeded in describing lower order of cognition in non-representational terms but confronted more difficulties when accounting for "higher-order cognitive abilities such as memory, imagination, reflective judgment and so on" (Gallagher 2017, p. 187, in De Oliveira, Raja and Chemero 2019). Gallagher (2017) refers to this as the 'scaling-up' problem, and the authors suggest that a solution to this problem, i.e., "the right explanation of our higher-order cognitive capabilities," can be found in a "combination of affordance-based virtual action and development within a culture."

How should this combination be understood? Affordances (Gibson 1979) are perceivable opportunities for present or distal action. For example, when I perceive a coffee-machine (an artifact), I do not perceive it just as a thing standing in front of me, I perceive it as an opportunity for preparing me a coffee right now (present) or later (distal). Tools and artifacts are conspicuous objects because their meanings are perceivable to their users. When seeing the coffee-machine, I do not need to remember what it offers me but perceive that directly. How are the affordances of tools and artifacts related to human development within a culture? De Oliveira, Raja and Chemero (2019) answer this question arguing that humans become acquainted with their world, one comprising social actors, norms, tools and artifacts, by participating in *cultural settings*:

> Several theorists have claimed that human development and participation in cultural settings, including narrative practices and skill learning, lead to new kinds of abilities to tell stories and use cultural artifacts that enable us to have thoughts about things that are distant in space and time (Gallagher and Hutto 2008; Rietveld and Kiverstein 2014, in De Oliveira, Raja and Chemero 2019).

Participating in cultural settings is a step needed to develop higher-order cognitive abilities. Tools and artifacts' affordances are perceivable only by habitual participants of those settings. Importantly, social environments' material and cultural differences determine and sometimes thwart opportunities for action (Gallagher 2020). Suppose a person is planning to learn German, but nobody speaks this language in her social circle, and neither are German courses in the city where she lives. Her environment's material characteristics might thwart her plan but not necessarily, as she still could learn via e-learning platforms. The purposeful design of these artifacts enables their users to develop their language skills. Creators design them to be spaces for interaction, in which users communicate with tutors and other students. Users also interact with bots acting as instructors and offering guidance. Finally, users have access to materials such as video lessons and exercises. As artifacts, these platforms offer spaces for social interaction aimed at the development of a cognitive ability.

De Oliveira, Raja and Chemero (2019) treat common and explicit thoughts as instances of affordance-based action. They suggest that when people have explicit thoughts, such as "I have to clean the table," they are, in fact, building tools that help them in dealing with future scenarios. They describe thoughts as models or tools based on the distinction between models *of* and models *for* in philosophy of science (Keller 2000). This distinction suggests that models are not merely copies, representations or images of some real thing. Instead, models are primarily for some action, in the sense that "they are designed to be manipulated in specific ways and to accomplish certain tasks" (De Oliveira, Raja and Chemero 2019). Crucially, models are not in principle tied to one specific function. Bayesian models, e.g., are for investigating perception, improving decision-making processes, or investigating the dynamics of belief-change in social simulations (e.g., Lewandoski et al., 2019). Apart from being for some action, models are for someone. As tools, they "are used by someone to do something in some context" (De Oliveira, Raja and Chemero 2019).

Transferring the distinction above, the authors suggest that thoughts are *for* something rather than *of* something. The thought 'I need a coffee,' which happens right after I perceive the coffee-machine, is not a thought *about* the coffee machine. It is about the action I intend to do

5

and may be *for* providing it with a temporal location. In that sense, thoughts are not representations *of* the things they are about, but rather "tools, instruments and artifacts that some agent creates and uses." As tools, they serve for accomplishing common tasks, such as "simplifying complex problems, predicting future experiences, explaining past experiences, categorizing experiences, confirming or disconfirming expectations, enabling adaptivity, generating understanding, and more" (De Oliveira, Raja and Chemero 2019).
Though models and thoughts are similar in the sense that they are *for* something and some users, unlike models, common or explicit thoughts are not "public objects that can be interacted with by different users, thoughts are typically private creations of individuals" (De Oliveira, Raja and Chemero 2019). This difference is strong enough to undermine the analogy's value because models, as cultural artifacts, are not tied to a particular or context relative function. In contrast to explicit and common thoughts, models reproduce over time, sometimes generations, as the example of Bayesian models in the previous section illustrated. Acknowledging this critical difference, the authors insist that what matters the most is that thoughts are usable for someone who knows how to use them. This formulation relates thoughts with tools, which is the analogy that the authors strictly commit with.
The problem of this second analogy is that thoughts and tools have an essential difference. Can thoughts be perceived as opportunities for action? That does not seem to be the case, as common thoughts mostly occur after perceiving objects or are about them. They may verbalize or conceptualize an affordance without being a tool in virtue of that. Besides, unlike tools, explicit and common thoughts are not perceived as being for something in virtue of their external appearances, and sometimes neither need to be verbalized, i.e., to have an external appearance, in the first place. The appearance of a tool, in contrast, can be an indicator of its functions. Another difference is that tools have an intentional design, which gives them their appearances. In short, to be successful, the analogy between thoughts and tools must undermine the significance of the material and perceivable properties and the purposeful design of tools, which is a price not willing to pay.

III. Thinking through artifacts

An artifactual account of thinking does not need to conflate thought with tools or models. It just needs to explain how various forms of thinking proceed by manipulating tools and artifacts. In the following, I sketch an alternative artifactual view of thinking also based on Material engagement theory (MET). I propose that certain types of thinking, e.g., scientific reasoning, can be best understood as thinking through artifacts. According to MET (Malafouris 2018), "human beings are not merely embedded in a rich and changing universe of things," but their cognitive and social life are "genuinely mediated and often constituted by them." For instance, by engaging with the things surrounding us, we develop our capacities to feel and know our bodily senses (Malafouris 2018). MET puts into question the passive role cognitive scientists attribute to things in cognition. In Malafouris' (2020) words,

> We are used to thinking about things as inert and passive. Moreover, thinking is usually understood as something we do about things in the absence of things. Material-engagement theory proposes a radically alternative claim: that human mental life (cognition and affect) is a process genuinely mediated and often constituted by things. The presence of the simplest artifact has the potential to alter the relationships between humans and their environments. New artifacts create novel relations and understandings of the world. New materialities bring about new modes of acting and thinking (Knappett and Malafouris 2008; Latour 1992).

For MET, thinking does not solely happen inside our heads and is not primarily representational. Thinking is not just about or of something but also thinking with or through things. Studies of the 2-4-6 rule discovery task illustrate the idea of thinking through something. In the task, participants have to find the rule that governs the arrangement of these numbers. An experiment showed that participants performing the task with external representations were more likely to discover the rule. In that sense, a graphical representation fosters the development of a cognitive task (Vallée-Tourangeau and Payton 2008).

Moreover, MET depicts thinking as transactional because it involves trade-offs between the elements of a whole organism-environment system (Iliopoulos 2019). Participants finding the 2-4-6 rule with external representations are not simply offloading their cognitive efforts to external representations; they think of the problem distinctively compared to participants who do not think through them. MET and, in particular, its idea of thinking through stress the constitutive role of the physical substrate of thinking (Malafouris 2018).

MET is consistent in many respects with the artifactual view of models. For MET, processes of thinking do not occur inside our brains but in interaction with external things. Likewise, for the artifactual account, scientific models are external artifacts that offer specific ways of thinking about phenomena by manipulating them. Besides, their material properties possibilitate specific forms of thinking. For instance, thinking phenomena as displaying systemic behaviors require models that analogically display such behavior, as Kirsch (2010) illustrates in the following,

> To compute the behavior of an *n* body system, such as our solar system, our best hope is to construct a small analog version of that system—an orrery—then run the model, and read off the result. Using this analog process, we can compute a function (to a reasonable degree of approximation) that we have no other reliable way [...]. If this interaction is essentially physical—if, for instance, it relies on physical equilibria, or mechanical compliance, or friction—there may be no reliable way of running an internal simulation. We need the cognitive amplification that exploiting physical models provide. We would need to rely on the parallel processing, the physical

7

> interaction, and the intrinsic unpredictability of those analog systems. There is nothing in our brains (or minds) like that.

Both MET and the artifactual view emphasize the active role things and artifacts play in cognitive processes. For both, non-scientific and scientific thinking proceed by means of manipulating tangible things and artifacts, which are vehicles of cognition, respectively. Thinking, MET suggests, is an interactive process involving an interplay between what the mind does to things, e.g., providing them with forms in pottery making, a case of object production (Malafouris 2018), and what material things afford to the mind. From a different perspective, Menary (2007) exemplifies the role of external vehicles when thinking and writing an article:

> There is, of course, an attenuated sense in which I can compose an article in my head. The likelihood of retaining much of the argument and structure, would, however, become very limited. Making revisions and corrections would be almost impossible, for example: trying out ideas and then deleting them. By contrast, becoming integrated with external tools and representations transforms my cognitive capacity to compose a philosophy paper. Importantly, there are things I can do with pen, paper, or word processor that I cannot do in my head. Stable and enduring external written sentences allow for manipulations, transformations, reorderings, comparisons and deletions of text that are not available to neural processes.

Menary's account shows the importance of external tools in the process of writing. As he notices, written sentences allow for various interventions that cannot be possible by thinking of an article inside the head. These external vehicles are "available for further manipulations such as restructuring, revising and re-drafting" (Menary 2007). Similarly, in scientific practices, external tools do not merely enhance, substitute or support reasoning; they allow scientists to think about objects in specific ways not otherwise possible. This happens because scientists exploit the material and symbolic aspects of scientific models, as Knuutila (2017) illustrates in the following:

> In the case of physical three-dimensional models, concrete media plays a more direct epistemic role. The material dimension of physical models does not merely function as an external scaffolding for cognitive and communicative functions. It also allows scientists to draw inferences that are based on the material features of the model. But the material features of the model also embody a symbolic, conceptual dimension. For example, the Phillips–Newlyn hydraulic model as a physical three-dimensional object embodies and renders visible economic ideas such as the principle of effective demand and the conceptualization of economy in terms of stocks and flows. The way water pools and flows in the containers and the tubes of the model takes upon it a symbolic significance.

Finally, the shift from internalistic to externalistic views of thinking can serve to reassess some empirical approaches to scientific thinking. For instance, empirical studies on hypothetical thinking tend to argue that human biases implicated in this form of reasoning result from the "design limitations of our cognitive machinery" (Ball 2020). In this context, Simon's (1982) notion of bounded rationality suggests that human biases are not irrational, but proceeds in the way it does because our cognitive machinery has evolved to meet adaptive constraints (Ball 2020). The idea of thinking through artifacts extends the notion of bounded rationality, as it stresses that material engagements are permanently changing "not just through history and evolution but even within a single lifetime" (Malafouris 2019). In other words, biases in thinking could be actively shaped by thinking through artifacts. For instance, while using concepts and invoking theoretical entities, such as structures or systems, many philosophers and scientists become more and more convinced that such concepts and entities are more than human imagination products and have an actual existence. They are partially on the right track: concepts and theoretical entities exist as artifacts; they are intersubjective objects that people can use to think through.

Conclusion

In this paper, I raised skepticism towards the view of thoughts as models and tools. This view fails to explain the different tangible and intersubjective dimensions involved in the use of tools and models. Then, I developed an artifactual account of thinking based on material engagement theory (MET). In this account, scientific thinking can be best understood as thinking through artifacts. Both MET and the artifactual view emphasize the role of external things in thinking. The artifactual view stresses that modelers obtain their cognitive gain by manipulating the models' tangible material and symbolic aspects. Thinking through models, for instance, occurs when scientists make inferences by manipulating the material and symbolic properties of models.

References

Arbib, Michael, Anon Plangprasopchok, James Bonaiuto, and Robert E Schuler. 2013. "A neuroinformatics of brain modeling and its implementation in the brain operation database BODB." *Neuroinformatics* 12(1): 5–26. https://doi.org/10.1007/s12021-013-9209-y.

Aussems, Suzanne, and Sotaro Kita. 2021. "Seeing iconic gesture promotes first- and second-order verb generalization in preschoolers." *Child Development* 92(1): 124–41. https://doi.org/10.1111/cdev.13392.

Artun, Ömer, and Dominique Levin. 2015. *Predictive marketing*. New Jersey: John Wiley & Sons, Inc.

Ball, Linden J. 2020. "Hypothetical Thinking." In *The Cambridge Handbook of the Imagination,* edited by Anna Abraham, 514–28. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108580298.

Bayes, Thomas. 1763. "An essay towards solving a problem in the doctrine of chances." *Philosophical Transactions of the Royal Society of London* 53: 370–418. https://doi.org/10.1098/rstl.1763.0053.

Baykal, Halil, and Hatice Kalkan Yildirim. 2013. "Application of artificial neural networks (ANNs) in wine technology." *Critical Reviews in Food Science and Nutrition* 53(5): 415–21. https://doi.org/10.1080/10408398.2010.540359.

Gallagher, Shaun. 2017. *Enactivist interventions: Rethinking the mind*. Oxford: Oxford University Press.

Gallagher, Shaun. 2020. *Action and Interaction*. Oxford: Oxford University Press.

Gallagher, Shaun, and Daniel Hutto. 2008. "Understanding others through primary interaction and narrative practice." In *The shared mind: Perspectives on intersubjectivity*, edited by J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen, 17–38. Amsterdam: John Benjamins.

Gallagher, Shaun, and T. G. Ransom. 2016. "Artifacting minds: Material engagement theory and joint action." In *Embodiment in evolution and culture*, edited by C. Tewes, 337–351. Berlin: De Gruyter.

Gibson, James J. 1979/2015. *The Ecological Approach to Visual Perception*. New York: Taylor & Francis.

Glorot, Xavier, Antoine Bordes, and Yoshua Bengio. 2011. "Deep Sparse Rectifier Neural Networks." In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, edited by Geoffrey Gordon, David Dunson, and Miroslav Dudík, 315–23. Fort Lauderdale: JMLR Workshop and Conference Proceedings 15.

Hasson, Uri, Samuel A. Nastase, and Ariel Goldstein. 2020. "Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks." *Neuron* 105(3): 416–34. https://doi.org/10.1016/j.neuron.2019.12.002.

Huang, Yiqun, Lars J. Kangas, and Barbara A. Rasco. 2007. "Applications of artificial neural networks (ANNs) in food science." *Critical Reviews in Food Science and Nutrition* 47(2): 113–26. https://doi.org/10.1080/10408390600626453.

Iliopoulos, Antonis. 2019. "Material engagement theory and its philosophical ties to pragmatism." *Phenomenology and the Cognitive Sciences* 18(1): 39–63. https://doi.org/10.1007/s11097-018-9596-5.

Keller, Evelyn Fox. 2000. "Models of and models for: Theory and practice in contemporary biology." *Philosophy of Science* 67. https://doi.org/10.1086/392810.

Kirsh, David. 2010. "Thinking with External Representations." *AI & Society* 25(4): 441–54. https://doi.org/10.1007/s00146-010-0272-8.

Knappett, Carl, and Lambros Malafouris, eds. 2008. *Material Agency: Towards a non-anthropocentric approach*. New York: Springer.

Knuuttila, Tarja. 2005. "Models, representation, and mediation." *Philosophy of Science* 72(5): 1260–71.

Knuuttila, Tarja. 2011. "Modelling and representing: An artefactual approach to model-based representation." *Studies in History and Philosophy of Science Part A* 42(2): 262–71. https://doi.org/10.1016/j.shpsa.2010.11.034.

Knuuttila, Tarja. 2017. "Imagination extended and embedded: Artifactual versus fictional accounts of models." *Synthese*. https://doi.org/10.1007/s11229-017-1545-2.

Knuuttila, Tarja, and Andrea Loettgers. 2014. "Magnets, spins, and neurons: The dissemination of model templates across disciplines." *The Monist* 97(3): 280–300. https://doi.org/10.5840/monist201497319.

Knuuttila, Tarja and Andrea Loettgers. 2016. "Model templates within and between disciplines: From magnets to gases—and socio-economic systems." *European Journal for Philosophy of Science* 6(3): 377–400. https://doi.org/10.1007/s13194-016-0145-1.

Knuuttila, Tarja, and Andrea Loettgers. 2017. “Modelling as indirect representation? The Lotka-Volterra model revisited.” *British Journal for the Philosophy of Science* 68(4): 1007–36. https://doi.org/10.1093/bjps/axv055.

Lewandowsky, Stephan, Toby D. Pilditch, Jens K. Madsen, Naomi Oreskes, and James S. Risbey. 2019. “Influence and seepage: An evidence-resistant minority can affect public opinion and scientific belief formation.” *Cognition* 188: 124–39. https://doi.org/10.1016/j.cognition.2019.01.011.

Latour, Bruno. 1992. “Where are the missing masses? The sociology of a few mundane artefacts.” In *Shaping technology/building society: Studies in sociotechnical change*, edited by W. E. Bijker, and J. Law, 225-59. Cambridge: The MIT Press.

Li, Xiaochen, Wenji Mao, Daniel Zeng, and Fei-Yue Wang. 2008. “Agent-based social simulation and modeling in social computing.” In *Intelligence and Security Informatics*, edited by Christopher C. Yang, Hsinchun Chen, Michael Chau, Kuiyu Chang, Sheau-Dong Lang, Patrick S Chen, Raymond Hsieh, et al., 401–12. Berlin, Heidelberg: Springer.

Lukeprog. 2011. “A history of Bayes’ theorem.” *Lesswrong* (blog). August 29, 2011. https://www.lesswrong.com/posts/RTt59BtFLqQbsSiqd/a-history-of-bayes-theorem.

Malafouris, Lambros. 2020. “Thinking as ‘thinging’: Psychology with things.” *Current Directions in Psychological Science* 29(1): 3–8. https://doi.org/10.1177/0963721419873349.

Malafouris, Lambros. 2019. “Mind and material engagement.” *Phenomenology and the Cognitive Sciences* 18(1): 1–17. https://doi.org/10.1007/s11097-018-9606-7.

Malafouris, Lambros. 2018. “Bringing things to mind: 4Es and material engagement.” In *The Oxford Handbook of 4E Cognition*, edited by Albert Newen, Leon De Bruin, and Shaun Gallagher, 755–72. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.40.

McGrayne, Sharon Bertsch. 2011. *The theory that would not die: How Bayes’ rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New Haven: Yale University Press.

Menary, Richard. 2007. “Writing as thinking.” *Language Sciences* 29(5): 621-632.

Morrison, Margaret, and Mary S. Morgan. 1999. “Models as mediating instruments.” In *Models as Mediators. Perspectives on Natural and Social Science*, edited by Margaret Morrison, and Mary S. Morgan, 10-37. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511660108.003.

Patel, Jigneshkumar, and Ramesh Goyal. 2007. “Applications of artificial neural networks in medical science.” *Current Clinical Pharmacology* 2(3): 217–26. https://doi.org/10.2174/157488407781668811.

Raftery, Adrian E. 1986. “Choosing models for cross-classifications.” *American Sociological Review* 51(1): 145–46.

Rietveld, Erik, and Julian Kiverstein. 2014. “A rich landscape of affordances.” *Ecological Psychology* 26(4): 325–352.

Rosenberg, Monica D., B. J. Casey, and Avram J. Holmes. 2018. “Prediction complements explanation in understanding the developing brain.” *Nature Communications* 9(1): 1–13. https://doi.org/10.1038/s41467-018-02887-9.

Rosenblatt, Frank. 1958. “The Perceptron: A probabilistic model for information storage and organization in the brain.” *Psychological Review* 65(6): 386–408.

Sanches de Oliveira, Guilherme, Vicente Raja, and Anthony Chemero. 2019. “Radical embodied cognitive science and ‘real cognition.’” *Synthese*. https://doi.org/10.1007/s11229-019-02475-4.

Schlaifer, Robert. 1959. *Probability and Statistics for Business Decisions. An introduction to managerial economics under uncertainty*. New York: McGraw-Hill.

Schoot, Rens van de, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G Tadesse, et al. 2021. "Bayesian statistics and modelling." *Nature Reviews Methods Primers* 1(1). https://doi.org/10.1038/s43586-020-00001-2.

Simon, H. A. 1982. *Models of bounded rationality*. Cambridge: The MIT Press.

Vallée-Tourangeau, F. and T. Payton. 2008. "Graphical representation fosters discovery in the 2-4-6 task." *Quarterly Journal of Experimental Psychology* 61(4): 625–640.

Vieira, Sandra, Walter H.L. Pinaya, and Andrea Mechelli. 2017. "Using Deep Learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications." *Neuroscience and Biobehavioral Reviews* 74(A): 58–75. https://doi.org/10.1016/j.neubiorev.2017.01.002.

Xia, ZhengYou, and LaiLei Huang. 2007. "Emergence of Social Rumor: Modeling, analysis, and simulations." In *Computational Science – ICCS 2007*, edited by Yong Shi, Geert Dick van Albada, Jack Dongarra, and Peter M A Sloot, 90–97. Berlin, Heidelberg: Springer.

Yu, Yuetian, Cheng Zhu, Luyu Yang, Hui Dong, Ruilan Wang, Hongying Ni, Erzhen Chen, and Zhongheng Zhang. 2020. "Identification of risk factors for mortality associated with COVID-19." *PeerJ*. https://doi.org/10.7717/peerj.9885.

Abduction and the Selection Mechanisms

Lukáš Bielik

a) General Philosophy of Science

Title: *Abduction and the Selection Mechanisms*

Keywords: Abductive inference, the context of discovery, the context of justification, the context of hypothesis-application, explanation, the selection mechanisms;

Extended Abstract:

There seems to be a theoretical consensus that abductive inference enhances various cognitive processes in science. Philosophers usually highlight two prominent contexts of using abduction as an inferential device (see e.g. Aliseda 2006; Douven 2017, Schurz 2008): a) the context of discovery (or alternatively, the context of hypothesis-generation); and b) the context of justification (or the context of confirmation; see also Niiniluoto 2018 for a further distinction between the confirmatory use and the acceptance use of abduction). Leaving aside other aims that abductive inference may help to attain (such as computational task solving), I pay close attention to a methodologically different context of abductive inference: c) the context of hypothesis-application. It is this third context where abductive inference plays the role of a genuine explanation of particular fact (similar to that of standard models of explanation such as Hempel's DN model or Woodward's causal model) or where it provides a diagnosis for a certain state of affairs (Josephson & Josephson 1996).

Depending on the context in which an abductive inference is operating, I propose to distinguish three broad kinds of abductive inference in an explicitly formal way, i.e. in terms of selection mechanism or selection principle. I introduce three different (ideal) models of selection mechanism corresponding to the three contexts of abduction and I define the selection principle (for a given context) as a partial function ***S*** defined on the following triple of arguments: 1. a (non-empty) set of explanatory hypotheses; 2. a background ***B*** of an agent (or a group of agents) ***A*** comprising their doxastic and epistemic propositional elements on the one hand, and (possibly an empty set of) theoretical virtues on the other hand; and 3. a (new) piece of evidence ***E*** to be explained. If a given selection principle operates on an admissible n-tuple of arguments, it yields at most one explanatory hypothesis (or its relevant content-part) as its function-value.

Having the notion of selection principle at our disposal, it is possible to elucidate explicitly, how both, the logical form of abductive inference and an accompanying set of methodological factors differ with respect to these three contexts in a unifying framework. In particular, the notion of selection principle enables us to model and represent all relevant elements of abductive inferential process within the context of discovery, the context of justification, and the context of hypothesis-application. The selection mechanisms differ in those contexts with respect to the arguments of the selection principle as well as to the function-values (if any) they yield. For instance, when considering the context of justification it is usually assumed that some theoretical virtues play the role of preference-weighting for a pool of (available) explanatory hypotheses with respect to a given evidence (see Harman 1965; Thagard 1978). Theoretical virtues reflect (in a given context) how good explanation a given hypothesis provides for evidence ***E***. As a result the hypothesis which fares better than the alternatives reaches a confirmatory approval. However, in the context of hypothesis-application, the selection effect of theoretical virtues is minimal. Instead, the relevant pieces of evidential information supplied by background ***B*** affect the selection of one of the already confirmed (justified) hypotheses, or its relevant content-part, as an adequate explanation (or a final piece of explanans) for ***E***. The aim of the selection mechanism in this context is neither to use explanatory considerations to confirm a hypothesis, nor to discover a potential hypothesis for further tests. Rather it is to use already confirmed hypotheses to select the best proper explanation for a particular fact.

Finally, I provide a further motivation for identifying inference to the best explanation ("IBE") with abduction in the context of justification. In particular, I show that at least some widely-discussed objections against IBE in the literature (such as van Fraassen's (1989) argument from a bad lot) which have been also widely discussed (e.g. Lipton 2004; Okasha 2000; Psillos 2004 and others) are not relevant to other forms of abduction: those in the context of discovery and the context of hypothesis application. Hence, our proposal of modelling different kinds of abduction by different kinds of selection principles helps us to distinguish the objections that are limited to IBE as a special form of abduction from those that affect abduction in general.

References (a reduced version)

Aliseda, A. (2006): *Abductive Reasoning.* Springer.

Douven, I. (2017): Abduction. In: Zalta, E. N. (ed.): *The Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/entries/abduction.

Harman, G. (1965): Inference to the Best Explanation. *Philosophical Review* 74, No. 1, 88-95.

Josephson, J. & S. Josephson (eds.) (1996): *Abductive Inference. Computation, Philosophy, Technology.*

Lipton, P. (2004): *Inference to the Best Explanation*. London: Routledge.

Magnani, L. (2001): *Abduction, Reason, and Science: Processes of Discovery and Explanation.* New York: Kluwer & Plenum.

McCain, K. & T. Poston (eds.) (2017): *Best Explanations. New Essays on Inference to the Best Explanation.* Oxford: Oxford University Press.

Niiniluoto, I. (2018): *Truth-Seeking by Abduction.* Springer Nature Switzerland.

Okasha, S. (2000): Van Fraassen's Critique of Inference to the Best Explanation. *Studies in History and Philosophy of Science* 31, 691-710.

Poston, T. (2014): *Reason and Explanation*. Basingstoke: Palgrave Macmillan.

Psillos, S. (2004): Inference to the Best Explanation and Bayesianism. In: F. Stadler (ed.): *Induction and Deduction in the Sciences*. Dordrecht: Kluwer, 83-91.

Schurz, G. (2008): Patterns of Abduction. *Synthese* 164, 201-234.

Smokler, H. (1968): Conflicting Conceptions of Confirmation. *The Journal of Philosophy* 65, 300-312.

Thagard, P. (1978): The Best Explanation: Criteria for theory choice. *The Journal of Philosophy* 75, 76-92.

van Fraassen, B. (1989): *Laws and Symmetry*. Oxford: Oxford University Press.

a) General Philosophy of Science

Title: *Abduction and the Selection Mechanisms*

Keywords: Abductive inference, the context of discovery, the context of justification, the context of hypothesis-application, explanation, the selection mechanisms;

Short Abstract:

There are two prominent uses of abduction usually distinguished with respect to the aim to be attained: a) the context of discovery; and b) the context of justification. Putting other potential uses of abduction aside, I pay close attention to another methodologically different context of abductive inference: c) the context of hypothesis-application. It is this third context where abductive inference plays the role of a genuine explanation of particular facts. I define the notion of a selection principle in terms of which it is possible to explicitly distinguish the three different contexts of abduction. Finally, I discuss the relevance of traditional objections against Inference to the best explanation with respect to those contexts.

A Norm for Science Advice: Making Beliefs Accurate

Simon Blessenohl and Deniz Sarikaya

a) General Philosophy of Science

Title:
A Norm for Science Advice: Making Beliefs Accurate

Key words:
Scientific Policy Advice, Uncertainty, Values in Science, General Philosophy of Science

Short Abstract: (96 Words)
Politicians rely on scientists to inform their policy-making. When may scientists make claims such as `X is toxic' even though their evidence does not conclusively show that X is toxic? One view says that scientists ought to make their uncertainty explicit. But politicians might ignore or even be confused by language indicating uncertainty. Another view says that scientists ought to take ethical consequences of policies into account. This view clashes with the proper role of scientists in democratic decision-making. We propose a view overcoming these issues: scientists should say what maximizes the accuracy of politicians' credences.

Long Abstract: (674 Words)
How should scientists communicate their findings when they advise politicians? One view holds that scientists should say what they have a high credence in. For example, they should not assert `X is toxic' if they only have a credence of 0.7 that X is toxic. Rather, they should make their uncertainty explicit to say something weaker that they have a high credence in, such as `it is likely that X is toxic'. Another view holds that scientists should say what they expect to have the best policy consequences. For example, if scientists know that politicians will enact climate policies only if scientists do not make their uncertainty explicit, and the scientists think that climate policies are desirable, then they should not make their uncertainty explicit. We explore a third view, according to which scientists should say what they expect to make the politicians' credences most accurate. That is, if a scientist has a credence of 0.7 in X being toxic, then she should say whatever brings the politician's credence close to 0.7. If this requires not making uncertainty explicit or saying things she takes to bring about suboptimal policy consequences, so be it. For ease of reference, let us state the three views as three norms for scientists advising politicians.

(Honesty)	Advising scientists ought to say what they have a high credence in.
(Policy)	Advising scientists ought to say what maximizes the expected value of the policy consequences of what they say.
(Addressee)	Advising scientists ought to say what maximizes the expected accuracy of their addressee's credences in the target propositions.

This talk explores the advantages, problems, and implications of (Addressee). First, we outline potential advantages of (Addressee) over its two alternatives. In particular, (Addressee) does not permit scientists to skew their advice based on their moral assessment of policies. This is

an advantage over (Policy) because such influence would undermine procedural values of democratic decision-making. It also constitutes a defense of the value-free ideal, which is sometimes attacked on the basis of a norm such as (Policy). Also, (Addressee) does not require scientists to say what they have a high credence in even if that is counterproductive to induce more accurate credences in the addressee. This is an advantage over (Honesty) because, in such cases, it seems permissible to say what one has a low credence in, if that makes the politician's credences more accurate. We then turn to a central problem of (Addressee), that it seems to require vicious communication strategies if those happen to maximize the expected accuracy of the politician's credences. Finally, we assess prominent examples of science advice in the light of (Addressee).

Just as many other activities, science advice might be governed by different kinds of norms which employ different senses of `ought'. Maybe there is a moral norm, which tells us what scientists morally ought to say, and an epistemic norm, which tells us what scientists epistemically ought to say. Then, the above norms might not be mutually exclusive: they might simply employ different senses of `ought'. But we read all these norms in a specific sense of `ought': the moral sense. What we are after is a moral norm of science advice, not an epistemic norm.

The three views are simplified versions of more plausible views. In particular, the above views say that advising scientists should always optimize for a particular metric. For instance, (Honesty) says that scientists should always say what they have a high credence in. This is very implausible. In a case in which the scientist would be killed if she said what she had a high credence in, it seems false that she morally ought to say what she has a high credence in. But instead of making the views more complicated by introducing qualifications to deal with such extreme cases, we hereby restrict the scope of our discussion to everyday cases of science advice. We are interested in what should guide scientists when they communicate to politician's in standard, non-pathological situations: honesty, policy, or accuracy.

This is joint work with **Anonymized**.

Demographic variables predict differential sensitivity to inductive risks

Matteo Colombo, Silvia Ivani and Leandra Bucher

Demographic variables predict differential sensitivity to inductive risks

Extended Abstract
Scientists and policy makers commit numerous errors. Some are trivial; some are serious for only some communities; some are catastrophic for anybody. One way to understand these mistakes is in terms of the distinction borrowed from statistical decision theory between Type-I and Type-II errors. Type-I errors are "errors of commission," whereas type-II errors are "errors of omission." There is the risk of making a type-I error, where, for example, a new, actually unsafe medical treatment is introduced to the market as safe. And there is the risk of making a type-II error, where a new, actually safe treatment is withheld from the market as unsafe.

The risk of committing type-I or type-II errors constitutes an *inductive risk*, which philosophers define as the risk of error in accepting or rejecting hypotheses (Hempel 1965, 92). Appealing to the notion of an inductive risk, some philosophers of science have recently argued that ethical, social, political, and economic values (*non-epistemic* values, Douglas 2000) play an ineliminable, often legitimate, role in scientists' and policy makers' decision making under uncertainty (Rudner 1953, Douglas 2000).

The argument from inductive risk begins by noting that there are often significant social costs associated with making type-I or type-II errors (Rudner 1953). The relative seriousness of a type-I vs. type-II error can be assessed based on their social, political or economic consequences. Because the tradeoffs involved in weighing the costs and benefits of these consequences cannot be resolved without bringing non-epistemic considerations to bear on the relevant standards of evidence required to accept or reject a hypothesis, it is false that non-epistemic values should never influence scientists' inquiry and decisions (Douglas 2000).

A substantial body of work in philosophy of science has examined scientists' and policy makers' decision making in the face of inductive risks. Much of its focus has been on how the argument from inductive risk should be understood, its consequences for understanding the role of non-epistemic values in science, and on sources of scientific error in several historical case studies (e.g. Douglas 2000; Staley 2017).

One question this research leaves unanswered is how to balance type-I and type-II errors. In particular, it remains unclear whether or not anything general could be said about how to trade off the risk of making a type-I error and the risk of making a type-II error.

Our aim in this paper is to address this question empirically, focusing on how lay people judge policy making decisions based on uncertain scientific evidence. Specifically, our aim is to begin to clarify whether or not people are generally biased against possible type-I vs. type-II errors.

Although it's obvious that there is no single, generally correct way to make this tradeoff, several scientists and methodologists often suggest that type-I errors are more serious than type-II errors. Some statistics textbooks suggest that type-I errors are generally more serious than type-II errors, making an analogy with legal trials. Just like it is generally more desirable to wrongly let a guilty defendant free than to wrongly convict an innocent, it would generally be better to commit a type-II error and reject a true scientific hypothesis than a type-I error and accept a false scientific hypothesis (see e.g. Nickerson 2000, 244; Feinberg 1971; Wasserman 2013). Neyman (1942) explains this bias against type-I errors as a consequence of the methodological choice in hypothesis testing of choosing as the null hypothesis the hypothesis with the worst consequences of type-I errors. He says: "a more or less general convention was adopted to consider as the hypothesis tested [the null hypothesis] the one by which the errors of the first kind are of greater importance than

those of the second" (Neyman 1942, 304), where "of greater importance" means with the worst consequences. Echoing Neyman, it's been pointed out that researchers in psychology have been guided by "a strong bias against the making of a Type I error" in the selection of confidence levels (Nickerson 2000, 243). This bias against type-1 errors could be described tendency as a form of conservatism inspired by the Occam's razor principle; the idea here is that: "The approach of biasing against Type I error is intended to be conservative in the sense of beginning with an assumption of no difference and giving up that assumption only on receipt of strong evidence that it is false" (ibid.).

Based on these suggestions, we set out to test the hypothesis that people are generally biased against type-I errors. Specifically, we hypothesized that, across different domains of inquiry and policy making, the risk of type I error predicts more reliably than the risk of type-II error people's levels of confidence about the goodness of a policy decision taken on the basis of uncertain scientific evidence.

To test this hypothesis, we conducted two experiments. Experiment 1 focused on inductive risks involving possible type-I errors. Experiment 2 considered inductive risks involving possible type-II errors. Overall, our results are *inconsistent* with the null hypothesis that people are not differentially sensitive to possible type-I and type-II errors. Our results are instead consistent with the idea that people are systematically biased against type-I errors across a variety of policy domains.

In order to explore this difference between different types of inductive risks, we identified associations between our experimental participants' judgements and their demographics and political values. Our main findings here are that women were more uncertain than men about rejecting a result, and non-conservatives were overall more uncertain than conservatives about any policy decision.

These results are philosophically and psychologically interesting, because they bear out the suggestion oft made in the philosophical and statistical literature that avoiding type-I errors is of greater importance than avoiding type-II errors, and also because they suggest an intricate relationship between reasoning under uncertainty, personal values and endorsement of public policies. While our paper demonstrates that people's demographics and political values can predict their judgments about policy decisions in the face of inductive risks, it leaves open the question of what role citizens' values, needs, and expectations should play in answering urgent policy issues.

Abstraction's Logicality and Invariance

Ludovica Conti

Abstraction's Logicality and Invariance

Section f): Formal Philosophy of Science
Keywords: Abstraction principles, Logicality, Invariance.

1 Abstract

In this talk, I aim at discussing a logicality criterion suggested in the abstractionist debate. I briefly recap the criteria usually adopted in this debate and focus on a particularly weak criterion, which has been recently proposed to prove the (alleged) logicality of the so-called indefinite expressions. My aim is to show that such criterion is too weak to guarantee the logicality of the abstraction operators, but it is very useful for spelling out the semantic properties of abstraction.

2 Abstraction, Logicality and Invariance

Abstractionist theories are composed by a logical core augment with an abstraction principle, of form: $@_R\alpha = @_R\beta \leftrightarrow R(\alpha, \beta)$ – which introduces and rules an operator term-forming ($@_R$) as a new symbol of the language. Then, the logicality of such theories plainly depends on the logicality of the abstraction principles. The issue of their logicality originally was rised into the seminal abstractionist program, Frege's Logicism. The inconsistency of this project (i.e. a theory equivalent to second-order logic augmented with Basic Law V) seemed to determine the inconsistency and, then (in a classical logic) the non-logicality of Basic Law V and – *a fortiori* – of any other abstraction principle[1].

Recently, the issue of the logicality has been resumed regarding the consistent abstraction principles, in order to clarify that conclusion in light of the intervening studies about logicality. Briefly, a standard account of logicality has been provided, in semantical terms, by means of the Tarskian notions of invariance under permutation and isomorphism (cfr. [5]). In order to apply this criterion to abstraction principles, we can specify at least three different subjects to be examined: the whole abstraction principle, the abstraction relation or the abstraction operator.

Regarding the abstraction principle, the more informative criterion consists of *contextual invariance*: an abstraction principle AP is *contextually invariant* if and only if , for any abstraction function f_R: $D_2 \rightarrow D_1$ and permutation π, $\pi(f_R)$ satisfies AP whenever f_R does (cfr. [1]). I argue that this criterion is not adequate to state the logicality of a principle. I suggest two argument in support of this hypothesis. Firstly, it under-determines the choice between principles that are mutually inconsistent (like Hume's Principle and Nuisance's principle). Secondly, such criterion appears to be – both formally and conceptually dependent on the fulfilment of constraints concerning the abstraction relation: it is provably implied[2] by the

[1] We will describe a relation between the abstraction principles based on the *finess* of their equivalence relations. Cfr. [1].

[2] Cfr. Antonelli 2010, proposition 9: "Suppose R is weakly invariant and D_2 is π-closed; then the principle Ab_R is contextually invariant."

weakest form of invariance[3] of the abstraction relation; furthermore, a careful review of the syntactical structure of the abstraction principles shows that abstraction relation is the real "engine" of the abstraction principle.

Regarding the abstraction relation, we can distinguish, at least, four kind of invariance: *weak invariance*, *double invariance*, *internal invariance* and *double weak invariance* (cfr. [1], [3], [4].). I briefly describe their mutual relations and emphasise that, regarding abstraction relations, a very relevant meaning of logicality is provided in terms of *internal invariance*: an equivalence relation $R(X,Y)$ is *internally invariant* if and only if, for any domain D and permutation $\pi : X \cup Y \rightarrow D$, $R(X,Y)$ if and only if $R(\pi[X], \pi[Y])$.

As anticipated *weak invariance* of the abstraction relation is sufficient to have a *contextually invariant* abstraction principle but none of these criteria coincide or imply the invariance of the abstraction operator.

3 Abstraction operator

Regarding the abstraction operator, logicality is usually spelled out in terms of *objectual invariance*: an abstraction operator @ is *objectual invariant* if and only if for any domain D and permutation π, $\pi(@_R) = @\mathrm{R}$ – namely, $@_R(\pi(X)) = \pi(x) \leftrightarrow @_R(X) = x$. Such criterion fails precisely in case of operator related to weak invariant relation (cfr. [1]).

In order to overcome such unfortunate mismatch between abstraction operator and abstraction relation (then, abstraction principle), a different criterion has been proposed: it is the result of a generalisation of the Tarskian usual criterions, in accordance with the alleged indefinite meaning ofthe abstraction operators – spelled out, in its turn, in terms of *arbitrary* reference[4].

Such weakened criterion of invariance – which we will call *weak invariance* – proposed for the *arbitrary* interpretation of the abstraction operator[5] consists in a generalised version of the Tarskian isomorphism invariance and turns out to be satisfied, at least on some domains, by all the abstraction operators that index classes of the partitions obtained by *weakly invariant* equivalence relation[6], then, *a fortiori*, by the abstraction relations that exhibit higher form of invariance and by the *contextually invariant* abstraction principles. If we accepted such criterion of logicality, we would be able to classify many abstraction operators as logical symbols.

My first aim consists in further clarify this (apparently) new criterion – derived (in informal terms) from the standard criterion of isomorphism invariance only by substituting the canonical notion of reference with the arbitrary one. I will formally prove that *weak invariance* of the arbitrary denotation of an abstraction operator is nothing but the contextually invariance of the abstraction principle respect to all the possible non-arbitrary denotations of the same operator. By means of a supervalutational semantics, we can define all the possible precisifications of the operator on a certain domain as the ordered pair comprising the domain and the choice of a possible (non-arbitrary) denotation of the abstraction operator. Then, we can prove that the arbitrary denotation of the operator is *weakly invariant* on a domain D if and only if every precisification on D makes the principle *contextually invariant*.

Now, we can focus on some objections precisely pointing against the adoption of such criterion as a hallmark of logicality. Firstly, we will explore a general source of concern with

---

[3] A relation R is *weakly invariant* if and only if, for any permutation π, R(X,Y) if and only if R($\pi[X], \pi[Y]$)

[4] It was proposed in [7] and in [2].

[5] Such criterion is available in [2]: given an isomorphism i from a domain D, let i^+ such that for every set γ of objects from D, $i^+(\gamma) = \{i(x) : x \in \gamma\}$. Then, an expression ϕ is invariant just in case, for all domains D, D' and bijections i from D to D', the denotation of ϕ on $D(\phi^D)$ is such that $i^+(\phi^D) = \phi^{D'}$.

[6] This result is proved in [7].

the weakness of the criterion. We could note that the notion of arbitrary reference actually involves the idea of automorphism and ideally that of isomorphism[7]; then, we could object that the weakened criterion is satisfied because it – being applied only to arbitrarily interpreted expressions – "anticipates" some typical *desiderata* of logicality in corresponding semantical assumptions.

Secondly, we will analyse three objections concerning the arguments that should support the adoption of such criterion. The logicality – as fulfilment of the weak invariance – of the abstraction operators indeed, is introduced by analogy with other variable-binding operators, like ι, ϵ and η[8] – which, for brevity, we will call "choice operators". Nevertheless, we should note some relevant differences between these expression and the abstraction operators: firstly, (second-order) abstraction operators – differently from the choice operators – are not *full*, namely they turn out to be empty whether evaluated in some domains; secondly, logicality (weak invariance) of abstraction operators do not coincide – differently from the logicality of the choice operators – with their purely logical definability[9]; thirdly, while the logicality of choice operator seems to formalise a property of the whole class of similar expressions, then of the intuitive notion of choice, on the contrary, the logicality of abstraction principles seems to regard only second-order abstraction principles, by excluding, e.g., any first-order abstraction operator, hence failing to capture the notion of abstraction in itself.

However, the weakness of such criterion, while appearing to underdetermine the actual logicality of the abstraction function, turns out to identify the crucial meaning of an *undemanding* or *deflationist* interpretation of abstraction (cfr. [1], [7]) – by reducing function symbols to devices for selecting first-order representatives of equivalence classes. For this reason, I suggest to consider such weakened criterion as the formalisation of the necessary semantical condition of the second-order abstraction. I will support this hypothesis by the evidence that, while other logicality criteria are satisfied also by inconsistent abstraction principles[10], such criterion regarding abstraction operator is satisfied by all and only the consistent ones.

References

[1] Antonelli, G. A. (2010). Notions of invariance for abstraction principles. Philosophia Mathematica, 18(3), 276-292.

[2] Boccuni, F., Woods, J. (2018). Structuralist neologicism. Philosophia Mathematica, 28(3), 296-316.

[3] Cook, R. (2016).Abstraction and Four Kinds of Invariance (Or: What's So Logical About Counting), Philosophia Mathematica, 25,1, 3–25.

[4] Fine, K. (2002). The limits of abstraction. Clarendon Press.

[5] Tarski, A. (1956). The concept of truth in formalized languages. Logic, semantics, metamathematics, 2(152-278), 7.

[6] Tennant, N. (1980). On ϵ and η. Analysis 40 (1): 5.

[7] The class that constitutes the arbitrary reference of a term ("the set of candidates denotations") could be informally considered as the set of all the images of its canonical denotation under any automorphism (i.e. permutation). Furthermore, the usual denotation of an arbitrary expression is defined (cr. [7]) as a function that assigns, *to every domain D*, the set of candidates denotations and turns out to be automorphism invariant if and only if, for any domain D, it is invariant under automorphism on D. Then, the images mentioned above (which naively constitute "the set of candidates denotations") are implicitly considered on every domain and should constitute a set invariant under any isomorphism.

[8] Cfr. [7].

[9] Cfr. [6].

[10] For example, the criterion of *weak invariance* of the equivalence relation is satisfied by the relation of order-isomorphism included in ordinal abstraction – which is inconsistent.

[7] Woods, J. (2014). Logical indefinites. Logique et Analyse, 277-307.

The case is not closed

Richard David-Rus

The case is not close

In my talk I'll argue for the plausibility of inquiring a specific type of scientific understanding, the one without explanation (UwE). In a short retrospective introduction, I will track the distinction between the explanatory and the non-explanatory understanding in the debate on understanding and recall the single articulated account on this type of understanding - Lipton's one. Next I will proceed by discussing the challenge that Lipton's examples might rise to any general theory of understanding. I'll argue that Kelp's construal (2015) that sees Lipton's examples as exposing the overintellectualisation exhibited by the existing theories of understanding does not exhaust the real challenge. This is rather to be identified in the need to account for such cases doing justice to their particularity – as cases of understanding *without explanation*. Neither does Kelp's theory of understanding offer a proper solution to Lipton's examples since it does not have the necessary resolution to adequately account for such forms of understanding esp. in scientific contexts. This might be due to a methodological mistake - using an epistemological approach in scientific contexts - as I will argue by supporting it with Dellsen's remarks (2018).

Lipton's examples are especially a threat for the explanationists (as Kelp calls them) theories of understanding and here we find also the strongest rejection of his approach. I will proceed next to reject the best articulated critique of Lipton's account advanced by Khalifa (2013, 2017). I'll concentrate on two main steps: on his reconstruction of Lipton's argument and his 'bigger than strategy' i.e. the strategy he intends to show that the explanatory understanding is bigger than the non-explanatory one. I will show that in his reconstruction introduces un unjustified existential reading that assumes the existence of a correct explanation behind any sort of UwE. This is obvious in the reconstruction of the basic assumption i.e. Lipton's Assumption (LA) for Khalifa. LA refers to Lipton's proposal to identify UwE with different benefits of an explanation, benefits that could be gained through other means while Khalifa's reconstruction reads out the necessary existence of an explanation. This reading misconstrues Lipton's argument and intention and facilitates Khalifa's unfair critique.

Khalifa's 'bigger than' strategy does not hold either in the paper version of his critique (2013) nor in the more disguised form in his book (2017). In the paper version his argument invokes measuring the degrees of understanding through the sets of answers to w-questions (what-if-things-would-be different) that an explanation generates. Nevertheless the strategy fails since the inclusion relation between such sets does not hold without extra assumptions. Such an implausible assumption in case of UwE through possible explanation stipulates that knowledge of actual explanation should imply also knowledge of all the other possible explanations. I'll use Lipton's fictitious example of the rigged boxing match to illustrate the situation. Moreover if we call the efficiency and productivity criteria of understanding (de Regt, Gijsbers 2017) we end having a greater understanding in case of a possible explanation than from the actual one in the discussed case.

In the book version, the RTO – the right track objection, the first and most important from the three objections that Khalifa formulates against UwE, embodies the bigger than strategy. According to RTO UwE forms are only stations on the track to the final explanation. The critique is based on a narrow construal of scientific inquiry by which all possible explanations are regimented in providing understanding or proto-understanding towards the goal of a final correct explanation. This ignores such major roles of possible explanations to redirect inquiry, to change tracks. By characterizing such UwE forms as proto-understanding, i.e. understanding provided by grasping elements that have just indirect explanatory roles, Khalifa's strategy becomes exposed to the danger of underdetermination. This is due to the fact that the same piece of background information can play indirect role in many different explanations.

I will end my presentation with a short sketch of some positive reasons that could support a further working agenda for the inquiry into UwE forms. Among them I will mention the initial motivation for inquiring scientific understanding to liberate it from an approach on explanation so that an inquiry into UwE will bring this project to its ultimate consequences. Another is linked to a better contribution to the inquiry of scientific practice in the frame of the philosophy of science in practice approach facilitating this way an understanding of the scientific knowledge production and dynamics. Moreover, the differences between understanding forms promoted in different domains of scientific inquiry as the social vs natural sciences might be illuminated through such an investigation. Not last the connection of the UwE forms with the explanatory sort of understanding should be an important item on the agenda.

References

de Regt H.W., Gijsbers V. (2016), *How False Theories Can Yield Genuine Understanding* [in:] *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science,* St Grimm, C. Baumberger, S. Ammon (eds.), New York and London: Routledge, 50-76.

Dellsen F. (2018), *Understanding in Epistemology and Philosophy of Science: A Complicated Relationship*, "Metascience" 27, 195–198. https://doi.org/10.1007/s11016-018-0293-3

Khalifa K. (2013), *The Role of Explanation in Understanding*, "British Journal for the Philosophy of Science" 64 (1), 161–187. https://doi.org/10.1093/bjps/axr057

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

Kelp C. (2015), *Understanding Phenomena*, "Synthese" 192(12): 3799–3816. https://doi.org/10.1007/s11229-014-0616-x

Lipton P. (2009), *Understanding without Explanation* [in:] *Scientific Understanding: Philosophical Perspectives*, H. W. de Regt, S. Leonelli, K. Eigner (eds.), Pittsburgh: University of Pittsburgh Press, 43–63.

Taming Conceptual Wanderings: Wilson-Structuralism

Matteo De Benedetto

Taming Conceptual Wanderings:
Wilson-Structuralism

ABSTRACT

Mark Wilson (Wilson, 2006, 2017) presents a highly original account of conceptual behavior that challenges many received views about concepts, reference, and conceptual change in analytic philosophy. His main point of contention with traditional understandings of concepts is the overestimation of semantic finality and conceptual mastery. In a detailed analysis of several conceptual histories ranging from everyday language to classical mechanics, Wilson shows how behind many apparently stable macroscopic predicates such as 'weight' or 'hard' lies a very complex web of localized patches of usage. Wilson conceptualizes this patchwork structure of local conceptual usages in terms of patches and facades, i.e. pluralistic and partially indeterminate meta-conceptual units that substitute traditional philosophical understandings of theories.

Despite the vast praise of Wilson's work, few attempts have been made to give a precise semantic reconstruction of his framework. In this work, I will show how a modified version of the structuralist view of scientific theories (Sneed, 1979; Stegmüller, 1976; Balzer et al., 1987) is able to rationally reconstruct Wilson's framework of patches and facades. More specifically, I will first show some surprising connections between two central notions of the structuralist reconstruction of scientific theories, i.e. theory-elements and theory nets, and Wilson's patches and facades. Moreover, we will see that the Structuralist framework, when adequately modified to eliminate its hierarchical understanding of scientific theories, is able to offer a precise semantic reconstruction of Wilson's ideas.

More precisely, I will present a modified structuralist framework, i.e. what I will call *Wilson-Structuralism*, in which I will keep the coarse-grained organization of Structuralism, but I will drastically change its representation of how the different law-like statements of a given scientific theory are organized. I will change the definition of a theory-net and the related specialization relation. I will not require theory-elements of the same theory-net to be related by subset inclusion of models, constraints, and links, but only by the non-empty intersection between these components. This change will allow the modified theory-nets to enjoy the 'multi-valuedness' needed to adequately represent several wanderings phenomena described by Wilson.

I will then argue that Theory-Elements in Wilson-Structuralism explicate Wilson's patches, while the modified Theory-Nets play the role of Wilson's facades. In order to support my claim, I will demonstrate how several wandering

phenomena described by Wilson can be represented in an anti-realist way in Wilson-Structuralism as particular set-theoretic relationships between components of different theory-elements. I will also further strengthen my case by showing how one of Wilson's main case studies of the wandering behavior of scientific terms, i.e. viscous fluids forces in classical mechanics, can be adequately reconstructed within Wilson-Structuralism.

References

Balzer, W., Moulines, C.U., and Sneed, J. (1987): *An Architectonic for Science, The structuralist Program.* D. Reidel Publishing Company, Dordrecht.

Sneed, J. (1979): *The Logical Structure of Mathematical Physics.* D. Reidel Publishing Company, Dordrecht.

Stegmüller, W. (1976): *The Structure and Dynamics of Theories.* Springer-Verlag, Heidelberg.

Wilson, M. (2006): *Wandering Significance: An Essay on Conceptual Behavior.* Clarendon Press, Oxford.

Wilson, M. (2017): *Physics Avoidance: And Other Essays in Conceptual Strategy.* Oxford University Press, Oxford.

Some Philosophical and Mathematical Aspects of the Concept of Infinity

Vladimir Drekalović

Some Philosophical and Mathematical Aspects of the Concept of Infinity

The concept of infinity can be found in various fields of spiritual activity, both scientific and non-scientific. In various contexts, it appears, for example, in philosophy, theology, mathematics, physics, etc. This paper aims to expose, analyze, and compare some meanings of the term that can be found in philosophical-theological and mathematical contexts. Among other things, we will show that understanding the concept of infinity is inseparable from the conception of finitude and is based on it, not only in the negative[1] but also in the positive meaning. Namely, although in its original philosophical/ancient sense, infinity is understood as a negation of finity, it can be shown that the modern mathematical understanding of the concept of infinity proposed by Cantor in a hierarchical sense is defined precisely on the analogy with the notion of finitude.

The second thing we want to do is analyze the attempt to establish a kind of hierarchy between, on the one hand, the mathematical and, on the other, the philosophical-theological notion of infinity. More specifically, it is an attempt to show that the first term is necessary for the full understanding and explanation of the second without such a relationship to exist in reverse.[2] We will show that this attitude is problematic because examples can be found that show that without an intuitive background of philosophical-theological character it is not possible to understand all manifestations of the concept of infinity in a mathematical context.[3] In this regard, we will put forward arguments in favour of a non-competitive and non-hierarchical understanding of terms in the two fields. We will prove that it is more sensible to speak of seeking differences and similarities in these terms and in connection with their common historical roots, rather than establishing their mutual order by the criterion of significance.

The term infinity is used in mathematics in different contexts. However, we could summarize all these uses under three general situations. The first is when you

[1] Bombieri (2011, p. 55).
[2] Oppy (2011, pp. 235-36).
[3] Cardinal is defined as the size of a set … This is the definition that St. Anselm used with the idea of God. (Usó-Doménech et al. (2016)). Similarly, the intuition underlying the fact that the inerval (0,1) and R are equipotent can be found in some philosophical-theological considerations about the "equipotentiality" of the micro and macro world, the individual soul and God (see Njegoš (2013)).

refer to infinity as the "number" of objects of a set. In the second case, we speak of infinity as the boundary of the set. In the third case, these are the so-called infinite-transfinite numbers. We aim to determine whether some of these three mathematical understandings of infinity can be recognized in philosophical/theological theories as well. Classical theologians, including scholastics, who of course could not have understood Cantor's formal theory, were not unanimous in this regard. For example, according to Aristotle's *Physics*, Thomas Aquinas attributes mathematical/quantitative infinity solely to matter, therefore, not to God.[4] Augustine's understanding of God's conception of infinity, on the other hand, is related precisely to the mathematical concept of infinity in terms of the cardinal number of sets.[5] In addition to the classical, there are a number of contemporary theologians who, in connection with the modern mathematical theory of cardinality, also speak of infinity as an attribute of God, finding in this place also one of the three types of mathematical infinity mentioned above.[6] Our aim is to show that Cantor, through his notion of "absolute infinity", the highest entity in the mathematical hierarchy of infinity, tried to construct a formal image of the absolute being that is part of philosophical-theological theories.

Literature:

Aquinas, T (1981). *Summa Theologiae* (Translated by Fathers of the English Dominican Province), London: Christian Classics.

Augustine (1966). *De civitate Dei* XII,19=*The City of God Against the Pagans*, vol IV (trans. P. Levine), Cambridge, MA: Harvard University Press.

Bombieri, E. (2011). The Mathematical Infinity. In M. Heller and W. H. Woodin (Eds.), *Infinity - New Research Frontiers* (pp. 55-75). Cambridge: Cambridge University Press.

Njegoš, P., P. (2013). *The Ray of the Microcosm*, Belgrade: Svet knjige.

Oppy, G. (2011). God and infinity: Directions for future research. In M. Heller and W. H. Woodin (Eds.), *Infinity - New Research Frontiers* (pp. 233-254). Cambridge: Cambridge University Press.

[4] Aquinas (1981, p.I q.7 a.3).
[5] Augustine (1966, pp 89-91).
[6] For example Pannenberg (1988, p. 28) and Tapp (2011, p. 95).

Pannenberg, W. (1988). *Metaphysik und Gottesgedanke*, Göttingen: Vandenhoeck & Ruprecht.

Tapp (2011), Infinity in Mathematics and Theology. *Theology and Science*, 9 (1), 91-100.

Usó-Doménech, JL., Selva, J.A.N. & Requena, M.B. (2016). Mathematical, philosophical and semantic considerations on infinity (I): General concepts. *Foundations of Science*, 21 (4), 615-630.

On expertise and democracy: using philosophy of science when considering Harry Collins and Robert Evans' Owls

Jaana Eigi

(Note: this is a re-submission of an unchanged abstract accepted for EENPS2020)

e) History, Philosophy and Social Studies of Science

On expertise and democracy: using philosophy of science when considering Harry Collins and Robert Evans' Owls

Keyword: elective modernism, expertise, values, representation

Harry Collins and Robert Evans have proposed a new institution for improving the use of scientific expertise in democracy: the Owls. The role of the Owls is to grade scientific consensus on issues of public relevance. In the presentation, I argue that some philosophers of science are already doing the work assigned to the Owls. At the same time, however, these philosophers', and others', arguments suggest that the Owls cannot operate in quite the same way Collins and Evans describe. I conclude that discussion of Collins and Evans' elective modernism can benefit from work in philosophy of science.

Collins and Evans (2017) offer a vision for expertise in democratic society that attempts to avoid both populism and technocracy. One of its guiding principles is the principle that policy-makers may always choose to reject scientific consensus but they must never misrepresent it. For example, policy-makers may choose not to act despite scientific advice to the contrary. However, when doing so in a situation of strong consensus in the relevant expert community, they must not justify the lack of action by claiming that more research is needed. Elective modernism also proposes a novel expert institution: the Owls. The role of the Owls is to characterise the state of consensus in the relevant expert community. Importantly, this does not mean identifying the truth of the matter – only the state of the debate in the relevant community. This knowledge can then provide an input for decision-making in accordance with the principle described above.

The Owls have to combine expertise in the relevant field with disinterestedness. Collins and Evans suggest that this combination is to be found in some reflexive natural scientists (the few who do not allow their own view of the issue to influence consensus grading) and some knowledgeable social scientists (the few who have expert knowledge of the relevant natural science). Finding sufficiently many of such experts, however, remains an important problem.

In the first part of my argument, I suggest that some philosophers of science already contribute to the task of grading consensus in specific scientific communities. For example, Philip Kitcher's (1997) work exposing problems in research on sex and race differences aims to show that our confidence in this research should be very low. Another example is Jacob Stegenga's (2017) analysis of the approval of new medicines that argues that our confidence in the appropriateness of approval should be very low as well.

The existence of such philosophical research suggests that there is an additional pool of candidates to join the Owls – an important discovery, given Collins and Evans' concern about suitable candidates. Even more importantly, it gives some plausibility to the institution of the Owls that may initially seem to be a part of an ideal theory with no connection to real practices.

I thus suggest that some work in philosophy of science may give support to Collins and Evans' proposals. As the second step of my argument, however, I argue that the work by many philosophers of science, including Kitcher and Stegenga, shows that the Owls cannot operate in quite the same way Collins and Evans envisage. In particular, I focus on two issues: democratic representation and discussion of values.

An element of the approach to expertise in elective modernism is the idea that experts do not represent ordinary citizens – they only represent their own epistemic communities. This also applies to the Owls as an expert community. The issues of democratic representation are to be addressed outside of expert communities – for example, by establishing citizen juries where ordinary citizens are represented. I suggest that this black-boxing of expert communities as non-representative is problematic in the light of philosophical work on values in science. One example is the argument from inductive risk that shows how non-epistemic values play an inevitable role in the evaluation of the sufficiency of evidence for accepting a hypothesis (see, e.g., Douglas 2009). Once the role of values is recognised, however, the question concerning the values involved in decisions and practices of a particular scientific community becomes important. In other words, it becomes important to ask what values are represented in the respective community – in this case, the Owls. Thus, the Owls should not be treated as simply non-representative.

Another crucial element of elective modernism is the insistence that intrinsic politics – which inevitably characterises expert communities – must not be made extrinsic. In particular, it means that expert disagreements are to be settled in accordance with the community's epistemic values and without referring to the consequences of the acceptance of particular conclusions. However, this requirement is similarly problematic in the light of the argument from inductive risk. The argument shows that the amount of evidence required for accepting a hypothesis reflects the seriousness with which one sees the consequences of a possible error. Several versions of the argument from inductive risk, including Douglas's, argue that appropriate values should in fact play a role in setting the standards of evidence – and, importantly, researchers should be open and transparent about these values. Again, this applies to the Owls as an expert community as well. Thus, the Owls should not be expected to keep the values involved hidden.

To conclude, I have argued that philosophy of science can be fruitfully brought into contact with elective modernism, helping to see points of agreement and points of disagreement when thinking about expertise and democracy.

References

Collins, Harry M. and Robert Evans (2017). *Why Democracies Need Science*. Cambridge and Malden (Mass): Polity Press.

Douglas, Heather (2009). *Science, Policy, and the Value-Free Ideal.* Pittsburgh: University of Pittsburgh Press.

Kitcher, Philip (1997). "An Argument about Free Inquiry". *Noûs* 31(3), 279–306.

Stegenga, Jacob (2017). "Drug Regulation and the Inductive Risk Calculus". In: Elliott, Kevin C. and Ted Richards (eds). *Exploring Inductive Risk: Case Studies of Values in Science*. New York: Oxford University Press, 17–36.

Extended predictive minds: do Markov Blankets really matter?

Marco Facchin

Extended predictive minds: do Markov Blankets really matter?

Section C: philosophy of cognitive and behavioral sciences

Keywords: Extended Mind, Free-energy, Markov Blankets

Short abstract

Many philosophers, inspired by Fristont's free-energy principle, have tried to adjudicate whether the extended mind thesis is true by restorting to Markov Blankets (MBs). Here, I claim that these attempts are misguided. I will argue that trying to directly determine whether the mind extends by using MBs is either question begging or circular, depending on the interpretation of MBs one endorses. Moreover, I will argue that simply framing the debate over the extended mind using MBs has nasty theoretical consequences, as is forces us to accept a way to identify mental constituents that is grossly extensionally inadequate

[97 words]

Long abstract [with references]

According to the extended mind thesis, a subject's mental machinery can sometimes include, alongside neurons, the environmental props and tools the agent manipulates while solving a cognitive task [1]. The recent rise to popularity of Friston's free-energy principle [2] reshaped the debate surrounding the extended mind thesis, suggesting that we should identify (cognitive) systems through formal boundaries known as Markov Blankets (MBs) [3,4]. Here, I wish to argue that MBs do not adjudicate, nor help us adjudicate, the debate over the extended mind thesis.

To start, I will provide a succinct introduction to the free-energy principle. In doing so, I will adhere to the *most prominent* reading of the free-energy principle, according to which free-energy minimization is a tool for biological self-organization [5] and MBs are the functional boundaries separating organisms from their environments [6].

Having done so, I will argue that, if such an understanding of MBs is in place, then resorting to MBs to adjudicate the debate over the extended mind is simply question begging.

This is because, according to the proposed interpretation, MBs are boundaries of organisms. However, the extended mind thesis just is the claim that the material constituents of the mind can escape such boundaries [1]. Hence, appealing to MBs to adjudicate whether the extended mind thesis is true is simply question begging.

I will then consider a second interpretation of MBs, according to which MBs bound biological systems (if not systems in general) at each scale of organization [7], conceding that such an interpretation of MBs does not directly begs the question against the extended mind. However, I will argue that even if MBs bounds systems in general and are thus “multiple and nested”, we cannot determine whether the physical machinery of the mind extends by looking at its MB. This is because, given how MBs are formally defined [e.g. 8], identifying a MB *presupposes* having identified the system it bounds beforehand. Hence, in order to identify the MB of the material constituents of the mind, we must *already* have identified said constituents. Hence, looking at the MB of the mind to see whether it extends looks viciously circular.

I will then consider the claim that MBs are useful *framing devices* to adjudicate the truth of the extended mind thesis [e.g. 3, 9], and argue that it is false.

To do so, I will examine two prominent MB-based approaches to the extended mind debate, one [9] defending and the other [3] attacking the extended mind; showing that, as a result of their reliance on MBs, both approaches are forced to draw the boundaries of the mind in a doubtlessly *extensionally inadequate* way.

This is because the framing offered by MBs forces one to identify the cognitive machinery with the machinery playing a role in free-energy minimization *overtime*, *on average and in the long run*. However, many “non-cognitive” items play a stable role in free-energy minimization. For instance, clothes allow us to minimize free-energy over time by keeping our bodily temperature around the expected 37°, but surely clothes are not part of our cognitive system. Moreover, the "overtime, on average and in the long run" is extensionally inadequate because not all the material constituents of the mind play such a temporally extended role in free-energy minimization. For example, during childhood many synapses die as a result of a process of neuronal selection [10], and so do not, in the long run, contribute to free-energy minimization. But surely these synapses *were* part of our cognitive system, when properly attached to the rest of the brain. The fact that in the future certain synapses will be pruned does not prevent them from playing a cognitive role *now*; and so does not prevent them from being partial constituents of a subject's mind.

I will then conclude suggesting that the challenge I raised to MBs might be more profound than my argument shows. This is because, strictly speaking, MBs are only formal properties of graphical models [11], and it is thus far from clear whether they can play the "mind bounding" role philosophers inspired by the free-energy principle wish them to play.

[870 words, references included]

References

[1] Clark, A. (2008). *Supersizing the Mind*. OUP.

[2] Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *TICS*, 13(7), 293-301.

[3] Hohwy, J. (2016). The self-evidencing brain. *Nous*, 50(2), 259-285.

[4] Clark, A. (2017). How to knit your own markov blanket. *Philosophy and Predictive Processing*, DOI: 10.15502/9783958573031

[5] Hohwy, J. (2020). Self-supervision, normativity, and the free-energy principle. *Synthese*, 10.1007/s11229-020-02622-2

[6] Friston, K. (2013). Life as we know it. *J. R. Soc. Interface*, DOI: 10.1098/rsif.2013.0475

[7] Kirchhoff, M. *et al*. (2018). The markov blankets of life. *J. R. Soc. Interface*, DOI: 10.1098/rsif.2017.0792

[8] Koski, T., & Noble, J. (2011). *Bayesian Networks: an Introduction*. Wiley and Sons.

[9] Kirchhoff, M., & Kiverstein, J. (2018). How to determine the boundaries of the mind: a markov blanket proposal. *Synthese*, DOI: 10.1007/s11229-019-02370-y

[10] Changeux, J. P. (1985). *Neuronal Man*. New York: Pantheon Books.

[11] Bruineberg, J. *et al.* (2020). The emperor's new Markov Blankets. *Preprint*. Accessed at: http://philsci-archive.pitt.edu/18467/

Attributing causal specificity

Maria Ferreira Ruiz

a) General Philosophy of Science

b) Philosophy of Natural Science

Attributing causal specificity

Key words: Causal parity thesis – causal specificity – causal selection – genetic causation – fine-grained control

Abstract

Critics of causal parity have defended an argument that invokes causal specificity as a property that grounds objective distinctions between genetic and nongenetic factors. I expose difficulties with this concept that threaten the intended objective nature of specificity-based distinctions. Such issues concern the contrast between *specific* and *nonspecific* (or "dial-like" and "switch-like") cases. It is unclear whether switch-like cases should be deemed either minimally specific (as suggested) or rather maximally specific. This is so because, I show, the idea of specificity encompasses two components that are logically disconnected, and intuitions vary depending on which one is given more weight.

Extended abstract

Causal specificity is drawing considerable attention from philosophers of biology and philosophers of causation. It became the rationale for rejecting (and later endorsing) a thesis of causal parity of developmental factors (Griffiths and Gray 1994). This literature assumes that attributing specificity to causal relations is at least in principle a straightforward (if not systematic) task. However, the parity debate in philosophy of biology seems is at a point where it is not more biological details that will release it. Indeed, it is drawing from biological details that some invoke specificity to make the case that developmental factors are not on a par, as certain factors are (more) specific, and others to make the case that non-genetic factors are highly specific, too (e.g. Waters 2007; Weber 2017a, 2017b; Griffiths et al. 2015).

My aim is to take a step back to reassess the value and limitations of specificity for the dispute. This is not carried out on the basis of further analysis of the overly discussed cases of DNA vs. the polymerase, or vs. alternative splicing. Rather, I identify *prior* complications with the notion of specificity itself. Even when recent views depict specificity as admitting of degrees, attributions of specificity still seem to rely on a somewhat high contrast between specific and nonspecific relations, the latter being often construed as “switch-like” in a popular radio analogy (Woodward 2010). My contribution reveals an irreducible “dual” nature of causal specificity, which is really about two components that are not mutually entailed. Assuming the relevant relations are causal, these are:

REPERTOIRE	*many* possibilities exist for each relatum, and
CONNECTEDNESS	the possibilities on both sides are connected *in a special way* (e.g. bijective).

While there is partial recognition that specific relations may differ roughly along these two aspects, the extant literature fails to notice the mutual independence of components and how they pull in different directions. I clarify this situation, and show how it affects attributions of specificity. Notably, intuitions about switch-like cases hang on the relative weight given to the components. These cases can be deemed *minimally* specific, as score low on repertoire, or *maximally* specific, because the function is bijective, which is paradoxical.

Ultimately, I contend, these are not minor issues, as they directly affect the very purposes for which specificity is often invoked: *in general*, to compare/distinguish amongst causal factors; *in particular*, to settle the causal parity dispute in philosophy of biology. Possibilities from here are sketched.

Griffiths, P. and Gray, R. (1994). “Developmental Systems and Evolutionary Explanation”. *Journal of Philosophy*, 91, pp. 277-304.

Griffiths, P., et al. (2015). “Measuring Causal Specificity”. *Philosophy of Science*, 82(4):529-555.

Waters, K. (2007). “Causes that make a difference”. *The Journal of Philosophy*, CIV(11):551-579.

Weber, M. (2017a). “Causal Selection vs Causal Parity in Biology: Relevant Counterfactuals and Biologically Normal Interventions.” In K. Waters, M. Travisano, and J. Woodward (Eds), *Causation in Biology and Philosophy*, Minneapolis: University of Minnesota Press.

Weber, M. (2017b). “Discussion Note: Which Kind of Causal Specificity Matters Biologically?”. *Philosophy of Science*, 84(3):574-585.

Woodward, J. (2010). “Causation in biology: Stability, specificity, and the choice of levels of explanation”. *Biology and Philosophy* 25:287-318.

Predicting and Understanding COVID-19: the role of the IHME model

Johannes Findl and Javier Suárez

3. Philosophy of the life sciences

Prediction, understanding and COVID-19: Early epidemiological models of COVID-19 shed light on the relationship between prediction and understanding

In an unprecedented manner, the COVID-19 pandemic has caused rapidly growing rates of viral infections all over the world, threatening the lives of many people and putting hospitals in acute danger of becoming overwhelmed. This emergency situation has accelerated scientific research, and many resources have been dedicated to understanding the pandemic, one key aspect being how it would spread. By developing epidemiological COVID-19 models that predict the rates of infections, mortality, and hospitalizations for scenarios in which countermeasures such as social distancing policies are introduced or removed at a time, scientists have sought to provide an urgently needed basis for political decision-making.

COVID-19 models can be distinguished by three different types: *statistical models* that derive their estimations from a regression analysis that fits a curve to empirical data such as the number of infections or deaths, *mechanistic models* that simulate disease transmission between (groups of) persons on the basis of empirical data such as the virus's spread and the onset of disease symptoms, and *hybrid models* that combine both approaches.

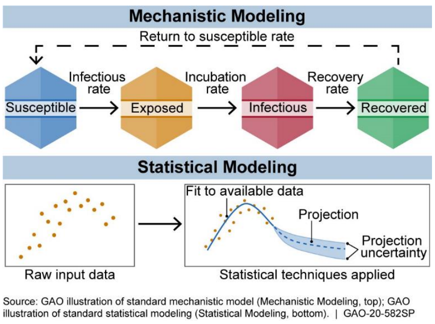

Figure 1. There are two general types of infectious disease models: mechanistic models, which use scientific understanding of disease

transmission and dynamics, and statistical models, whose predictions rely only on patterns in the data (GAO 2020).

Since little was known about the spread of the disease at the beginning of COVID-19, predictions were mostly generated with statistical models that did not include causal-mechanistic knowledge. While initially criticized for the large discrepancy between predicted and actual deaths, scientists have further developed those early statistical models and have eventually released increasingly accurate versions of them. Through continuous updates and modifications, some of these models have been improved up to the point where they can be said to have acquired genuinely predictive capacity, becoming indispensable tools for understanding how COVID-19 would spread in different scenarios. Arguably, the improvement also suggests that by and by, epidemiologists have achieved a better understanding of the main variables determining the spread of the disease.

This raises important philosophical questions about how purely statistical models can yield understanding, and what the relationship between prediction and understanding in such models is. In this paper, we investigate them by analysing how the statistical model from the Institute of Health Metrics and Evaluation (IHME model) was developed and modified between March and April 2020. The IHME model projects the mortality and hospitalization rates resulting from COVID-19 and has been used by US policy makers from early on in the pandemic.

We believe that by way of studying the details of the IHME model building process over time, we will be in a good position to analyse how the concepts of prediction and understanding interact and affect each other in the epidemiological practice related to COVID-19, which may also give us valuable insights into the more general nature of this relationship.

The first result of our analysis is that a key epistemic virtue of the IHME model is its ability to generate regularity patterns through predictions, thereby enabling scientists to understand how COVID-19 would spread in different scenarios. More precisely, we observe that the model creates regularity patterns by combing a simple *technical framework* that depicts a mathematical function (i.e., a Gaussian error function) with a series of *assumptions* about the set of variables that would affect the results of the technical framework (i.e., the introduction or the removal of countermeasures) and, in a sense, affect the phenomenon (i.e., the death and hospitalization rates). Understanding, we argue, results from considering and comparing the model's visualizations. As we show, this is in line with de Regt's intelligibility requirement, according to which scientists can acquire understanding only if they recognize qualitative consequences of their theory. We claim that the IHME's model visualizations should be thought of as regularity patterns that apply to possible scenarios in which countermeasures are either introduced or removed at a time. (**Figure 2**).

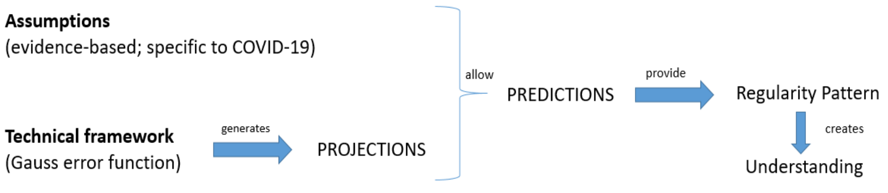

Figure 2. Schematic representation of the generation of predictions and their relation to understanding in the IHME model.

A second result of our analysis is that predictions have another important epistemic function over and above providing the regularity patterns that makes understanding of the phenomenon feasible. Focusing on how the IHME scientists developed their model between March and April 2020, we find that when comparing the model's predictions with the actual evidence, the scientists were prompted to reconsider their starting assumptions. By doing so, they eventually understood which of them were correct, and which ones were mistaken. Concretely, our analysis shows that certain important local conditions that severely affected the death rates from COVID-19 in different countries had not been included in early versions of the IHME model (such as population density, urban area vs. countryside, or level of compliance with the mobility restrictions) and thus were incorporated through a number of updates. This observation confirms our hypothesis that predictions facilitate understanding of a phenomenon by pinpointing the *error and success* in the model building process. That is, scientists use the model's predictions in order to test the assumptions and the technical framework that had generated them. This crucial step allows them to see where the model gets the phenomenon right, and where it needs to be improved.

Finally, we show that in the case of the IHME model's development, the concepts of understanding and prediction are intimately linked in a dynamic and dialectical way which results from the generation of a regularity pattern that is then compared with actual data. As **Figure 3** illustrates, at the first stage, the model generates predictions that provide a regularity pattern that allows scientists to understand the phenomenon (in our case, COVID-19-derived mortality rate). At the second stage, though, the predictions work *backwards*: they are contrasted with the evidence and, if scientists observe large divergence, they are forced to reshape their model by modifying the assumptions upon which it relies. This second step may require altering some of the prediction-generating assumptions, as it happened in the IHME model's development. Importantly, after this second stage has taken place, *the predictions are themselves altered*, i.e., they are freshly produced. This is because the change in the assumptions reshapes the model, hence the prediction-generating process starts again, and, in that vein, the whole assumption-prediction-understanding process is taken to the next stage.

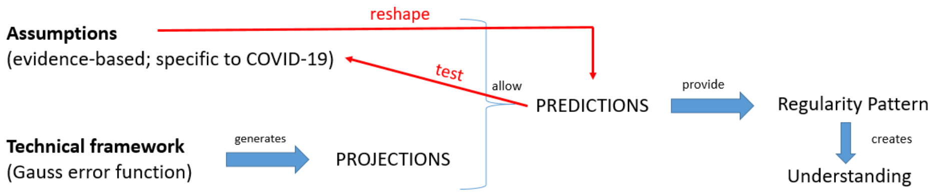

Figure 3. Schematic representation of the role of predictions in updating the IHME model. Notice that the movement is dialectical, as the relationship between the predictions and the assumptions works back and forth.

In sum, our work emphasizes an important step in the philosophical comprehension of purely statistical epidemiological modelling and how the philosophical concepts of scientific understanding and predictions are related with it.

Selection of references:

Broadbent, A (2013). Philosophy of Epidemiology. Palvgrave Macmillan.

de Regt, H. W. (2017). Understanding Scientific Understanding. Oxford University Press.

Douglas, H. E. (2009). Reintroducing prediction to explanation. Philosophy of Science, 76(4), 444–463.

IHME COVID-19 Forecasting Team. (2020). Modeling COVID-19 scenarios for the United States. Nature Medicine. https://doi.org/10.1038/s41591-020-1132-9

Causal Modeling as Counterfactual Semantics

Samuel Fletcher

Causal Modeling as Counterfactual Semantics

Over the last few decades, causal models have become indispensable in statistics and the social and biological sciences. The utility of causal models for guiding intervention in the world has even led to a resurgence in popularity of manipulability theories of causation, of which Woodward's (2003) is the most prominent. That account states that, roughly, X is a (direct) cause of Y when there is a possible intervention on X that, ceteris paribus, changes Y. (The ceteris are made paribus through interventions, too.) However, Woodward defines interventions in terms of causal relations, which only permits an implicit definition of causal relations and interventions together. The existence of causal models shows that these definitions are consistent, but because they are not explicit, they are not taken to provide a reduction of causal concepts to ones about intervention.

Woodward defends the utility of his theory in the face of this circularity, indicating how its effects are not necessarily epistemically pernicious. Nonetheless, reducing causal relations to others better understood or explicated would still be desirable, all things considered. My present goal is such a reduction, not just for Woodward's interventionist definition of causation, but for any of the plurality of causal notions definable with causal models, e.g., type and token, interventionist and counterfactual, etc. To do so, I show how interventions and the semantics for interventionist counterfactual conditionals—the main concepts used to define of causal concept with causal models—can be understood in terms of a variation on the Stalnaker-Lewis (S-L) semantics for counterfactual conditionals, adapted for application in science. (For simplicity, I focus here just on deterministic causal models.)

The S-L semantics make use of the collection of possible worlds and a similarity ordering thereon (or some other equivalent structure), relativized to each world. Lewis controversially proposed that the similarity ordering at a world should prioritize first widespread matching of laws of nature, then the size of the spatiotemporal region of matching particular facts, then localized matching of laws of nature, then localized particular facts. Then, “if X were the case, Y would be the case” is true at a world w just when at all the worlds in which X is true that are most similar to w are also worlds in which Y is true.

I propose to keep the structural features of the semantics while changing the relata of the similarity relation and how that relation is determined. First, I replace worlds with causal models. A causal model is a triple (V, S, A), where V is a set of variables, S is a collection of structural equations for V, and A is a value assignment to the variables in V. Variables specify a range of values of a particular property, e.g., whether a switch is “on” or “off”, or the number of people in a particular community. A structural equation expresses the value of one variable in V in terms of a function of the others, e.g., X=f(Y, Z) is a structural equation for the variable X in terms of variables Y and Z. There is at most one structural equation for any variable in V. The assignment A maps each variable in V to one of its possible values, as long as that mapping is compatible with the structural equations, e.g., X = "on" if X denotes the unimpeded position of a switch. In contrast to some authors developing the causal modeling framework (e.g., Pearl 2009), I do not take the structural equations of S to represent unanalyzed, primitive causal mechanisms. They merely place constraints on the possible assignments A. This is important to enable an explicit definition of causal notions in terms of causal models.

The second change I make to the S-L framework is to jettison Lewis's ordering based on miracles and matters of fact for one more bespoke for causal models. (There are nonetheless some qualitative similarities, as will be apparent.) The following broad outline of the new ordering will suffice for present purposes. Consider a causal model (V, S, A). First, any causal model (V', S', A') with V=V' is more similar to (V, S, A) than any with V≠V': such a (V', S', A') is more similar to (V, S, A) to the extent that V' approximates V. Second, any causal model (V', S', A') with V=V' and S=S' is more similar to (V, S, A) than any with V=V' and S≠S': such a (V, S', A') is more similar to (V, S, A) to the extent that S' approximates S. Third, any causal model (V, S, A) is more similar to itself than any with V=V', S=S', and A≠A': any (V, S, A') is more similar to (V, S, A) to the extent that A' approximates A. With this second change, one can evaluate counterfactual conditionals concerning (in both the antecedent and consequent) arbitrary syntactic combinations of variable assignments in the usual S-L way.

This is enough to model counterfactual definitions of causation that rely on causal models, but not for interventionist definitions of causation. For according to the latter definitions, in order to evaluate a counterfactual conditional, "if X were the case, Y would be the case," one must find not just causal models in which X is the case, but ones in there has been an *intervention* to *make* X the case.

Fortunately, one can define an intervention as a modal operator—much like Pearl's (2009) *do* calculus—applying to particular variable assignments, one which is true in a causal model just when its variable assignment is true and the variable in question is exogenous. (A variable is exogenous in a causal model just when there is no structural equation for it.) Interventionist counterfactuals then involve different antecedents—e.g., "if *do*(X) were the case, Y would be the case"—than non-interventionist ones. To evaluate it at a model (V, S, A), one must find the model most similar to (V, S, A) in which X holds and is exogenous. The above similarity ordering selects exactly the model arising from a surgical intervention on (V, S, A), as that notion is usually defined in the causal modeling literature. Thus S-L semantics models interventionist counterfactuals.

Metacognition in Medicine (Extended Abstract)

Gabriela-Paula Florea

Section: c) Philosophy of Cognitive and Behavioral Sciences

Title: Metacognition in Medicine

Keywords: metacognition, mental representation, biological response, chronic disease, the physician-patient relationship

The purpose of this paper is to find the right understanding of metacognition's meaning and role, applied to clinical medicine. The main definition of metacognition in literature comes from psychology and philosophy of mind and pertains to 'thinking about thinking'. Yet, other perspectives bring fuzziness to the subject. To render the matrix of metacognition with its multiple dimensions, I divide this paper into two sections. The first section (Proust's et al. contribution) brings upfront the philosophical perspective on metacognition. The second section (the necessity of introducing metacognition into medical practice) defends the role which metacognition could play in bringing a solution for the medical-care-crisis: the lack of communication between the physician and the patient because of the technological innovations that changed the way physicians look at the subject, the patient, and at the object, the disease. In the end, I present a personal perspective about the points discussed in the previous sections.

If we place the entire metacognitive research work that has been done so far on all registries under the idea of the web of thoughts, we can create a complex mental architecture spiderweb-like. Trying to describe metacognition accordingly to what the researchers have been said by now, one always starts from the original definition, namely 'knowing about knowing' after which various epistemological lines give different directions to this concept (developmental psychology, cognitive neuroscience, education etc). Starting from the core definition, one can draw radial lines towards each epistemological. Because of the metacognitive process' complexity and also the general applicability that it has in different fields, in some points, they intertwine with each other. The question derived from this, after so many thoughts of 'thinking about thinking', is how to connect them? And which of them can be taken into consideration in a field like clinical medicine where intensive cognitive processes are taking place in patient's mind and doctor's mind during their encounter, to find its utility in medical practice not only in medical education?

Bringing metacognition as a practical concept within the medical act would mean introducing new elements of cognitive skills as instruments that need to be implemented by those involved in a medical situation, namely the physician and the patient. The doctor would benefit by understanding how she would retrieve memorised information[1] and how to help the patient internalize metacognitive knowledge by having *collaborative conversations*[2] with the knowledgeable doctor through metacognitive linguistic input. The patient would be part of her own treating process without standing on a side, waiting for outside help only. Having in mind Flavell's model, proposing to monitor 'cognitive enterprises' that are implying different phenomena[3], one can argue about a segment of metacognitive knowledge that can be activated intentionally or unintentionally. Without relying on the unintentional activation, the doctor can trigger this response in what Sodian et al. call 'collaborative conversations', as mentioned above. This intentional or unintentional activation "may and probably often does influence the course of the cognitive enterprise without itself entering consciousness"[4]. This leads us to the idea that many patients experience this. This activation can lead to a 'conscious experience' calling it 'metacognitive experience'. Mental processes based on metacognitive knowledge "can have a number of concrete and important effects on the cognitive enterprises"[5] on children as well as on adults. Activating intentional metacognitive knowledge and bringing it to the level of metacognitive experience may lead the patient to positive effects on her healing process. I endorse this assertion by making an appeal to Benedetti's results in the pieces of evidence derived from neuroscience[6], by describing processes inside the patient's and physician's brain activity during their encounter.

As its main outcome, this paper aims to contribute towards a theoretical framework for future empirical research of how physician's and patient's metacognitive abilities could improve the evolution of the patient's disease and the prevention of physician's burnout.

[1] Eichbaum, Quentin G. 2014. "Thinking about thinking and emotion: the metacognitive approach to the medical humanities that integrates the humanities with the basic and clinical sciences." *The Permanente Journal* 64-75.

[2] Joelle Proust, Michael J. Beran, Johannes Brandl, Josef Perner. 2012. *Foundations of Metacognition.* Oxford University Press; p.128

[3] Metacognitive knowledge, goals/tasks and actions/strategies.

[4] Flavell, John H. 1979. "Metacognition and Cognitive Monitoring. A New Area of Cognitive—Developmental Inquiry." *American Psychological Association* (American Psychological Association,) Vol. 34 (No. 10): 906-911; p.908

[5] *Ibidem.*

[6] Benedetti, Fabrizio. 2011. *The Patient's Brain. The neuroscience behind the doctor-patient relationship.* Oxford University Press.

Quantifying proportionality and the limits of higher-level causation and explanation

Alexander Gebharter and Markus Eronen

Section: a) General Philosophy of Science

Title: Quantifying proportionality and the limits of higher-level causation and explanation

Short Abstract: Supporters of the autonomy of higher-level causation (or explanation) often appeal to proportionality, arguing that higher-level causes are more proportional than their lower-level realizers. Recently, measures based on information theory and causal modeling have been proposed which allow to shed new light on proportionality and the related notion of specificity. In this paper we apply ideas from this literature to the issue of higher vs. lower-level causation (and explanation). Surprisingly, proportionality turns out to be irrelevant for the question of whether higher-level causes (or explanations) can be autonomous while specificity is a way more informative notion for this purpose.

Extended Abstract: The status of higher-level causation is of key importance for many philosophical and scientific issues, including explanation in the special sciences, mechanistic explanation, emergence, non-reductive physicalism, and mental causation. The question is whether higher-level properties can in some sense be causally autonomous or indispensable with regard to their lower-level realizers, or whether genuine causation can only take place at lower (physical) levels. Sometimes the issue is also formulated in terms of whether causal explanations are better or preferable: Should we always prefer lower-level causal explanations when they are available, or are there situations or structures where higher-level causal explanations are better? On one side, emergentists and non-reductive physicalists (or non-reductively minded philosophers or scientists in general) argue that higher-level causes are at least sometimes autonomous, indispensable, or preferable. On the other side, reductionists argue that lower-level causes (or explanations) always trump higher-level causes (or explanations), and that appeals to higher levels are only needed for pragmatic reasons.

One important strategy for non-reductionists, going back at least to Yablo (1992), has been to appeal to proportionality to argue that higher-level causes can be autonomous (see, e.g., List & Menzies, 2009; Menzies & List, 2010; Schoemaker, 2000; Zhong, 2014), or provide better or equally good explanations than lower-level causes (see, e.g., McLaughlin, 2007; Weslake, 2013; Woodward, 2010, 2018). The idea is roughly that the causes should be somehow commensurate to their effects. The standard example is from Yablo's (1992): A pigeon is conditioned to peck at red objects. When presented with a scarlet ball, it pecks at the ball. Was the pecking caused by the ball being scarlet, or by the ball being red? Intuitively, it seems the ball being scarlet is too fine-grained as a cause, and that the redness of the ball is a more proportionate cause. Similarly, it can be argued that higher-level causes (e.g., mental states) are more proportionate than the lower-level causes (e.g., neural states) that realize them.

Although the idea of proportionality is quite intuitive, it has turned out to be difficult to spell out in a way that is clear and consistent. Many authors have also argued that it is a problematic or ill-defined notion that is not helpful for defending higher-level causation (see, e.g., Bontly, 2005; Franklin-Hall, 2016; McDonnell, 2017). However, in recent years advances in the causal modeling literature have led to opportunities to precisely define proportionality and the related notion of specificity. Several authors have proposed to characterize proportionality or specificity based on information theory (see, e.g., Bourrat, 2019; Griffiths et al., 2015; Pocheville, Griffiths, & Stolz, 2017). In this talk, we draw from this literature and apply ideas about how to quantify proportionality and specificity to the philosophical debate about higher-level causation. Based on these measures, we analyze causal structures involving higher- and lower-level causes, and draw several interesting conclusions:

(1) There are indeed cases where higher-level properties are more causally proportional than their lower- level realizers, but

(2) how common they are is a purely empirical question.

(3) Proportionality is irrelevant for the autonomy of higher-level causation (or explanation). It is not a good indicator for the causal powers or causal influence of a variable, and cannot be used to support the autonomy of higher-level causes.

(4) Higher-level causal explanations cannot provide information about the effects that goes beyond the information provided by the lower-level realizers, but can still be autonomous in the weak sense that they are not worse than explanations provided by the lower-level realizers.

In this talk, we assume an interventionist approach to causation. This approach has two strands. The more philosophical strand, developed by Woodward (2003), analyzes causation in terms of interventions. The basic idea of this approach is that a variable C is causally relevant for another variable E if there are possible interventions on C that would lead to a change in E. Interventions can be roughly understood as ideal experimental manipulations that change C without influencing any other causes of E (for details, see ibid., sec. 3.1.3 and 3.1.4). In our analysis, we apply the more general and more formal framework of causal Bayes nets (Pearl, 2000; Spirtes, Glymour, & Scheines, 2000), where causal structures are represented as causal graphs that are based on conditional independence relationships between variables.

Similarly, when we rely on the interventionist approach to explanation and explanatory power. According to this approach, explanatory power is a matter of providing answers to what-if-things-had-been-different questions, also known as w-questions (Woodward, 2003; Woodward & Hitchcock, 2003; Hitchcock & Woodward, 2003). More precisely, a generalization is explanatory insofar as it can answer w-questions, and explanation A is better or more powerful than explanation B if it can answer a wider range of w-questions. To take a classic example, the ideal gas law is invariant under a certain range of interventions, but van der Waal's force law is invariant under all of those interventions plus a range of further interventions. For this reason, van der Waal's force law can provide answers to a broader range of w-questions and consequently has more explanatory power. We acknowledge that there are alternative accounts and further dimensions of explanatory power (Ylikoski & Kuorikoski, 2010), but in this talk we focus on the ability to answer w-questions, as it is a widely held and relatively well-defined account.

The structure of the talk is as follows. After a short and more general introduction (part 1), we introduce the notions of proportionality and specificity in a more detailed way in part 2. In part 3, we discuss how to quantify proportionality and specificity based on recent literature and motivate the specific measures we will use in the remainder of the talk. In part 4, we finally apply these measures to the issue of higher-level causation and explanation.

Ground for Ontic Structuralists

Joaquim Giannotti

General Philosophy of Science

Ground for Ontic Structuralists

Under the banner of *ontic structuralism*, diverse approaches in the philosophy of science and the philosophy of physics gather. Despite specific differences, these views share two theses, which can be called respectively the *Fundamentality Thesis* and the *Priority Thesis*. The Fundamentality Thesis states that the structures of a theory in question—relations, extrinsic properties, and symmetries described by the relevant formalism of the theory—are fundamental. The Priority Thesis states that these structures are prior in an ontological sense to objects. In more precise terms, these can be formulated as follows.

Fundamentality Thesis. All fundamental physical entities are structures.
Priority Thesis. Fundamental structures are prior to putative physical objects if these exist.

A preliminary requirement for assessing the tenability of any version of ontic structuralism is the elucidation of the Fundamentality Thesis and the Priority Thesis. Candidate notions to make sense of these such as supervenience and ontological dependence failed to accomplish this aim. My aim is to show that *grounding* is not just a better candidate, but it also captures the metaphysical commitments of some standard ontic structuralist approaches.

To begin with, grounding is a form of non-causal metaphysical determination with explanatory import. It captures the idea that some entities obtain *because* or *in virtue of* other ones. In precise terms, where uppercase letters denote suitable entities, A grounds B if and only if A metaphysically determines B. On the orthodoxy view, grounding has an intimate tie with fundamentality and priority that is captured in two principles: **Ungroundedness** and **MFT**.

Ungroundedness. An entity X is fundamental if and only if there is no entity Y that grounds X.
MFT. If an entity X partially or fully grounds an entity Y, then X is more fundamental than Y.

According to these principles, the fundamental entities are those that are ungrounded, and the *grounds* are more fundamental than the *groundees*. Grounding orthodoxy offers a promising package deal for ontic structuralism. I propose to reinterpret the Fundamentality Thesis a claim about the ungroundedness of structures:

Ungroundedness Thesis. A structure is fundamental if and only if there is nothing else that grounds its identity or existence.

In a similar vein, I propose to construe the Priority Thesis in terms of grounding. Here the idea is that the priority of structures over objects can be understood adequately understood in the sense of the former being more fundamental than the latter. Accordingly, I reformulate the Priority Thesis as a *Grounding Thesis* as follows:

Grounding Thesis. The existence or identity of each putative fundamental physical object is partially or fully grounded in that of some fundamental structure.

The main advantage of reinterpreting the Fundamentality Thesis and the Priority Thesis respectively as the **Ungroundedness Thesis** and the **Grounding Thesis** is that we can elucidate both theses in a unified way. Therefore, a grounding version of ontic structuralism would be able to satisfy the preliminary requirement mentioned of clarifying its metaphysical commitments about the relation between structures and objects.

Unfortunately, a problem arises. McKenzie (2018) argues that grounding is inadequate to capture the relations between putative fundamental structures such as symmetry groups and objects such as fermions. Therefore, as her argument goes, **Grounding Thesis** is unfit for securing the priority of structures over objects. While McKenzie revolves around a specific case (that of symmetric groups and fermions), her objection generalizes.

According to McKenzie, grounding comes with an *entailment principle* which states that if X grounds Y, then X entails Y. Accordingly, if it is not the case that X entails Y, then it is also not the case that X grounds Y. McKenzie argues that the entailment principles fails to capture the relation between symmetry groups and fermions: symmetries do not entail what kind of fermions we can expect to find in nature (McKenzie 2018, 18). There is an infinite range of possibilities of the determinate values that fermions can have which are not entailed by symmetry considerations. Therefore, given the entailment principle, it is not the case that symmetries ground fermions.

This is unwelcome result threatens the prospects of articulating a grounding version of ontic structuralism. If the relation between putative fundamental structures according to our best science such as symmetry groups and particles cannot be adequately understood in terms of grounding, then we may expect similar problems for other versions of ontic structuralism. Note that McKenzie's objection generalises: any failure of entailment between structures and objects implies a failure of grounding between them and, consequently, we have no reason for regarding structures as more fundamental than objects. McKenzie's objection has some teeth but does not bite. To resist it, we need to consider a canonical distinction between *partial* and *full* grounding (Fine 2012). If X partially grounds Y, then X on its own or with some other entities fully grounds Y.

By introducing the full/partial grounding distinction, it is possible to submit a *revised entailment principle*: if X *fully grounds* Y, then X entails Y. Equipped with this machinery, we can resist McKenzie's objection. While McKenzie claims that symmetries do not fully ground fermions, she does not deny that the former partially ground the latter. Fermions are in factpartially constrained by symmetry groups. By invoking **MFT**, we can maintain the idea that symmetries are more fundamental, and therefore prior to, than fermions. Crucially, if symmetries only partially ground fermions, then we should not think that the revised entailment principle fails in this case. Therefore, we are not forced to reject the claim that symmetries ground fermions. Like McKenzie's objection, this strategy generalizes.

Overall, it allows the ontic structuralist to maintain the priority of structures over objects even if the former only partially grounds the latter. Therefore, this approach leaves grounding as a promising notion for elucidating the Fundamentality Thesis and Priority Thesis.

To corroborate the applicability of the proposed approach, I conclude by showing how to reformulate three standard ontic structuralist views (French 2010) from the viewpoint of the proposed grounding apparatus.

Works Cited

Fine, K. 2012. Guide to Ground. In F. Correia & B. Schnieder (eds.), *Metaphysical Grounding*. Cambridge University Press. pp. 37–80. French, S. 2010. "The Interdependence of Structure, Objects and Dependence". *Synthese* 175 (S1):89–109.
McKenzie, K. 2018. "Structuralism in the Idiom of Determination". *British Journal for the Philosophy of Science* axx061

The uses of truth: Is there room for reconciliation of factivist and non-factivist accounts of scientific understanding?

Lilia Gurova

Section: a) General Philosophy of Science

Title: The uses of truth: Is there room for reconciliation of factivist and non-factivist accounts of scientific understanding?

Keywords: scientific understanding, factivist vs. non-factivist accounts of understanding, effectiveness as a substitute for truth, neutral account of understanding, ideal gas model

Short abstract

The positions of factivism and non-factivism about scientific understanding do not look incompatible anymore if one realizes that both use the same substitute for truth – the notion of effectiveness. I will argue for a neutral account of understanding, based on the notion of effectiveness, which both factivists and non-factivists would be able to accept. Embracing the neutral account will help to quit certain fruitless discussions and to focus on more important questions such as 'What makes an understanding provider effective?' The ideal gas model will be used as an illustration of the advantages of the proposed neutral account.

Extended abstract

Among the most lively debates over scientific understanding in the recent years is the one about the relation between understanding and truth. The factivists and the quasi-factivists who argue that understanding is a kind of knowledge, and hence it implies truth, have insisted accordingly that the understanding providers, whatever they are (theories, models or any other kind of representation) should contain truth. The non-factivists for whom understanding is not reducible to knowledge, have provided arguments that being true (true enough, or partially true) is neither sufficient nor even necessary for a representation to yield understanding. The arguments of the non-factivists usually build on examples showing that non-linguistic representations (e.g. material models, which are not truth-apt), false theories, and models containing idealizations and fictions, are able to provide understanding. For many, the positions of factivism and non-factivism are obviously incompatible. As Khalifa put it, quasi-factivism is "simply denial of non-factivism" (Khalifa, 2017, p. 156). A closer look at these positions, however, reveals a different picture. On the one hand, the prominent non-factivist accounts of understanding (see e.g. Elgin, 2009; de Regt, 2017; de Regt & Gijsbers, 2017) seem not to entirely break with the notion of truth insofar as the concepts which they introduce as substitutes for truth, e.g. "responsiveness to evidence" (Elgin, 2009) and "effectiveness" (de Regt and Gijsbers, 2017) could not be properly construed without any reference to truth. As de Regt and Baumberger confessed, understanding and truth are connected on the non-factivist account "but what needs to be true are predictions

rather than the theories or models", which are used for drawing these predictions and which because of that are told to yield understanding (de Regt & Baumberger, 2019, p. 73). On the other hand, in other to defend factivism against the arguments of the non-factivists, some quasi-factivists like Khalifa have embraced an "expanded" concept of knowledge in the definition of which "acceptance" is taken as a substitute for "belief", and as a substitute for "truth" is chosen ... "effectiveness". Given that the two allegedly rival positions about scientific understanding, factivism and non-factivism, make use of the same substitute for the notion of truth (in this case this is the concept of effectiveness), it becomes difficult to support anymore the view that these positions are obviously incompatible. Rather it becomes clear that the claim for the incompatibility of the factivist and the non-factivist accounts of understanding needs to be revisited. That could be done in various ways. One can argue, for example, that the established conceptual similarity shows that non-factivism eventually degrades to factivism; or (s)he can argue for the opposite, that factivism eventualy degrades to non-factivism. Both of these lines of reasoning, however, are counterproductive as they lead to unresolvable fruitless debates. I suggest instead to get rid of the factivism vs. non-factivism opposition and embrace a neutral account of understanding. The neutral account could be based on the notion of effectiveness, which both the factivists and the non-factivists would be able to accept. The escape from the factivism vs. non-factivism debates will allow the former proponents of these positions to focus and join efforts on more fruitful questions such as 'What makes an understanding provider effective?' Indeed, if effectiveness is defined in terms of successful predictions, and if it is not necessary for a theory (or a model) to be true (true enough, or partially true) in order to be effective, it become pressing to answer the question what makes theories and models effective and thus understanding providers. Does, for example, the effectiveness of a given theory (or a model) depend on specific inherent properties of that theory (model) or is it determined by the context? The advantages of the proposed neutral, effectiveness-based account of scientific understanding will be demonstrated on the example of the ideal gas model, which has been used extensively so far in the arguments of both the factivists and the non-factivists.

References

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

De Regt, H. W., Baumberger, C. (2019). What is scientific understanding and how can it be achieved? In: McCain, K., & Kampourakis, K. (Eds.). *What is Scientific Knowledge?: An Introduction to Contemporary Epistemology of Science*. London: Routledge.

De Regt, H. W., Gijsbers, V. (2017). How false theories can yield genuine understanding. In: Grimm, S. R., Baumberger, G., Ammon, S. (Eds.) *Explaining Understanding. New Perspectives from Epistemology and Philosophy of Science.* New York: Routledge, 50–75.

Elgin, C. Z. (2009). Is understanding factive? In: Pritchard, D., Miller, A., Haddock, A. (Eds.) *Epistemic Value*. Oxford: Oxford University Press, 322–30.

Khalifa, K. (2017). *Understanding, explanation, and scientific knowledge*. Cambridge: Cambridge University Press.

The self-fulfilling prophecy machine. Beyond the science/technology distinction

Germán Hevia Martínez

EENPS 2020 d) Philosophy of Social Sciences

EXTENDED ABSTRACT

The self-fulfilling prophecy machine. Beyond the science/technology distinction

"If men define situations as real, they are real in their consequences". This is known in sociology as the Thomas's Theorem, popularize by the sociologist Robert K. Merton, who used it as the foundation of his notion of self-fulfilling prophecy. For him, this was a key problem that showed the paradoxical nature of human action. The main property of these types of prophecies is that with a *false* definition of the situation they evoke in human agents the behavior that makes the originally false prophecy *true*.

This idea is nowadays no surprising at all -even for the layperson-: is well known in sociology and in the field of economics. In fact, in philosophy most of the postmodern literature had assumed this statement through the concept of "performative" action and so on. But there is a blank space around this notion, a further development that hadn't been address and that relates to one of the biggest challenges of philosophy of social sciences today: the status of social sciences and the "nature" of their objects. Or to use other words: the problem of the epistemological and ontological foundations of the scientific enterprise that deals with social phenomena.

As has been noted by Anna Alexandrova, nowadays exceptionalism conceptions of social sciences are still been used to acknowledge the differential way in which social sciences try to study and understand its main object. That's the case of the philosophical framework developed by the Spanish philosopher Gustavo Bueno (called philosophical materialism) and of other well-known positions (the case of Max Weber or Dilthey). But they fall in the same error, due to the lack of a broader conception regarding the different types of knowledge used in social sciences. In the philosophical framework of Mario Bunge an original proposal can be found regarding this issue. The idea is simple: in social disciplines different types of knowledge are at play, been possible to distinguish in them between socio-scientific knowledge (which aim is to describe and explain through mechanism the way social systems works) and socio-technological knowledge (which aim is to transform social systems using social sciences).

Although it's indeed possible to maintain this distinction between scientific and technological knowledge in social disciplines, there is still a gap that need to be fulfilled. Exceptionalism conceptions aren't wrong in their notions: there is indeed something different in the ontological characteristics of social phenomena. Bunge acknowledge this in his statement that social systems, although they can be studied in a scientific way and possess internal norms to guaranteed its function like any other type of system, are social artifacts: they are what they are because in most of them it was decided that they should perform in that way. And, because of that, they can be, to some extent, modify though purposeful human action.

Nevertheless, Bungean philosophy cannot stand this constructivist statement without collapsing itself: as an integrated philosophical system, the development of this notion would imply a strong demarcation at the ontological level between natural and social sciences. One that Bunge himself rejects. That's because his philosophy of social sciences is still underdeveloped and doesn't give a satisfactory answer to the differential ontological characteristics of social phenomena. So, how is possible to acknowledge the possibility of maintaining a non-exceptionalism conception of the scientific enterprise and to acknowledge the artefactual nature of social systems at the same time?

An answer to that can be found in the notion of nomological machine, from Nancy Cartwright, a fixed arrangement of components that with repetitive operation give rise to the regular behavior that can be found out in the systems through scientific theories. If social techniques and social technologies are been used to shape the performance of social systems (economical systems, political systems, educational systems, public policy design, corporation management techniques, law systems, etc.), it's possible to say that self-fulling prophecies (and its counterpart, suicide prophecies) play the role of a mechanism in the creation of social phenomena from a micro level point of analysis. Furthermore, it's possible to stablished that in the macro level sociotechnological designs play the role of self-fulfilling prophecy machines, helping to stablish the performance and behavior of social systems through the repetitive and stable operation of actions by human agents according to a designed management plan from which arise the characteristic of social systems that we can find in social theories.

In this communication I'm going to address this issue and try to develop a theoretical framework capable of dealing with the complexity of this matter. I will show that using the notion of self-fulfilling prophecy machine it's possible to develop an approach that acknowledge the differential ontological characteristics of social phenomena while maintaining a non-exceptionalism epistemology in social sciences.

SHORT ABSTRACT

The self-fulfilling prophecy machine. Beyond the science/technology distinction

In this communication I'm going to address the problem of the ontological differences between social and natural phenomena through the demarcation problem, and I'll try to develop a theoretical framework capable of dealing with the complexity of this matter. I will show that using the notion of self-fulfilling prophecy machine, based on the work of Robert K. Merton on this issue and the notion of nomological machine from Nancy Cartwright, it's possible to develop an approach that acknowledge the differential ontological characteristics of social phenomena while maintaining a non-exceptionalism epistemology in social sciences.

Neuroscience without Brains: Perspectives on In Silico methods

Michal Hladky

Neuroscience without Brains: Perspectives on *In Silico methods*

EENPS 2020, Belgrade, June 10-12, 2020

Keywords

Models, Simulations, In Silico experiments, Neuroscience, Perspectivalism

Short abstract

In addition to the In Vivo and In Vitro studies, well established in Biology, new *In Silico experiments* are deployed in neuroscience (Blue Brain Project, Human Brain Project). Their philosophical analysis should provide a basis for an appropriate assessment of research projects, their results and the evaluation of clinical applications (Mulugeta et al. 2018). *In Silico experiments* differ from computer assisted research (Markram 2008, 2012; Markram et al. 2011) as they often provide information about the underlying mechanism, beyond the correlations established by data mining techniques. But should we understand *In Silico methods* as experiments or rather as simulations? If experimental methods provide a better epistemic access than simulations, can their classification as one or the other be only perspectival?

Extended abstract

In addition to the In Vivo and In Vitro studies, well established in Biology, new *In Silico experiments* are deployed in neuroscience (Blue Brain Project, Human Brain Project). Their philosophical analysis should provide a basis for an appropriate assessment of research projects, their results and the evaluation of clinical applications (Mulugeta et al. 2018). *In Silico experiments* differ from computer assisted research (Markram 2008; 2012; Markram et al. 2011) as they often provide information about the underlying mechanism, beyond the correlations established by data mining techniques. Furthermore, these highly realistic computer based reconstructions of target systems allow for interventions and manipulations, prompting scientists to talk and to think about these methods as experimental.

But should we understand *In Silico methods* as experiments or rather as simulations? If experimental methods provide a better epistemic access than simulations, can their classification as one or the other be only perspectival?

In this talk, I will first provide a mapping account of simulations based on model-theoretic notions and compare it with accounts based on representation (Frigg and Nguyen 2017; 2016; Weisberg 2013; Giere 2010; Suárez 2010; Goodman 1976). The failings and tensions between the ontic and the epistemic representational analyses of models have been stressed in the past (de Oliveira 2018). But representationalists have provided several arguments against the alternative model theoretic (mapping/morphism) based analyses (Suárez 2003). I will demonstrate how these objections can be overcome and argue that model theory provides the right tools for understanding and evaluating simulations. In addition, this approach is not only compatible with, but provides a framework for evaluating the correction criteria for scientific representation. Finally, I will illustrate my points with a case study from the Blue Brain Project (Markram et al. 2015).

The second part will be dedicated to the discussion on the epistemic power of simulations compared with the gold standard of empirical sciences - the traditional experimental methodology (Guala 2002; Morgan 2003; Parker 2009; Roush 2017). By comparing the typical inferential structures associated with experimental and model based reasoning, I will show why, *ceteris paribus*, experimental techniques are stronger than simulations, similarly to Parke's (2014)

conclusions. At the same time, this comparison leaves enough space for the seemingly opposing view based on cases where models and simulations provide better epistemic access than direct experiments, as target systems might be inaccessible (Winsberg 2009).

In the last part, I will show that it is possible to adopt a form of perspectivalism and describe experiments as simulations and vice-versa. Such an approach can explain the conflicting intuitions about simulations and in silico experiments, but generates a new puzzle. Can we really increase the strength of our conclusions by simply describing the method used to obtain them as an experiment rather than as a simulation? The answer is: we can't. Not all perspectives are created equal. The redescription of *In Silico methods* as experiments comes at the cost of introducing less plausible premisses into the associated inferential patterns and these undermine the strength of the conclusions about target phenomena.

Bibliography

Frigg, Roman, and James Nguyen. 2016. 'Scientific Representation'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/scientific-representation/.

———. 2017. 'Models and Representation'. In *Springer Handbook of Model-Based Science*, edited by Lorenzo Magnani and Tommaso Bertolotti, 25–48. Dordrecht Heidelberg London New York: Springer.

Giere, Ronald. 2010. 'An Agent-Based Conception of Models and Scientific Representation'. *Synthese* 172 (2): 269–281.

Goodman, Nelson. 1976. *Languages of Art: An Approach to a Theory of Symbols*. 2. ed., [Nachdr.]. Indianapolis, Ind.: Hackett.

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

Markram, Henry. 2008. 'The Blue Brain Project'. In . Keynote lecture at Neuroinformatics 2008 in Stockholm, Sweden. https://www.youtube.com/watch?v=8iDR8Z-e_GU.

———. 2012. 'The Human Brain Project'. *Scientific American* 306 (6): 50–55. https://doi.org/10.1038/scientificamerican0612-50.

Markram, Henry, Karlheinz Meier, Thomas Lippert, Sten Grillner, Richard Frackowiak, Stanislas Dehaene, Alois Knoll, et al. 2011. 'Introducing the Human Brain Project'. *Procedia Computer Science* 7: 39–42. https://doi.org/10.1016/j.procs.2011.12.015.

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

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

Mulugeta, Lealem, Andrew Drach, Ahmet Erdemir, C. A. Hunt, Marc Horner, Joy P. Ku, Jerry G. Myers Jr., Rajanikanth Vadigepalli, and William W. Lytton. 2018. 'Credibility, Replicability, and Reproducibility in Simulation for Biomedicine and Clinical Applications in Neuroscience'. *Frontiers in Neuroinformatics* 12 (April). https://doi.org/10.3389/fninf.2018.00018.

Oliveira, Guilherme Sanches de. 2018. 'Representationalism Is a Dead End'. *Synthese*, November. https://doi.org/10.1007/s11229-018-01995-9.

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

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

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

Suárez, Mauricio. 2003. 'Scientific Representation: Against Similarity and Isomorphism'. *International Studies in the Philosophy of Science* 17 (3): 225–244.

———. 2010. 'Scientific Representation'. *Philosophy Compass* 5 (1): 91–101.

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World (Oxford Studies in Philosophy of Science)*. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199933662.001.0001/acprof-9780199933662.
Winsberg, Eric. 2009. 'A Tale of Two Methods'. *Synthese* 169 (3): 575–92. https://doi.org/10.1007/s11229-008-9437-0.

Quantum Fragmentalism

Samuele Iaquinto and Claudio Calosi

Quantum Fragmentalism

Fragmentalism was originally introduced as a new A-theory of time.[1] It was further refined and discussed,[2] and different developments of the original insight have been proposed.[3] Recently it has been advocated —or at least considered— as a possible interpretation of physical theories such as Special Relativity.[4] In a celebrated paper,[5] Simon suggests that fragmentalism offers a new insight into Quantum Mechanics as well. In particular, Simon contends that fragmentalism delivers a new realist account of the quantum state —which he calls *conservative realism*— according to which: (i) the quantum state is a complete description of a physical system; (ii) the quantum (superposition) state is grounded in its terms,[6] and (iii) the superposition terms are themselves grounded in local goings-ons about the system in question.

The key insight in Simon (2018) is to *identify* different *terms* in a superposition state with Fine's *fragments*. In this paper we offer an argument against this core insight. The argument is simple enough:

$\mathbf{P}_1$. Given the basic tenets of fragmentalism, states of affairs that are parts of different fragments *cannot* interact.

$\mathbf{P}_2$. Different terms in a superposition state *can* —and *do*— interact.

C. Superposition terms are not fragments.

Clearly, the burden of the argument lies in the defense of premises $\mathbf{P}_1$ and $\mathbf{P}_2$.

Here is an argument for $\mathbf{P}_1$. Although there are different ways to pin down the notion of "fragmentation", the minimal idea is that states of affairs that do not belong to the same fragment *cannot* obtain together. Simon himself seems to concede this explicitly. Now consider an interaction between the states of affairs s_1 and s_2. It seems clear that *obtaining together* is a necessary condition for (the possibility of) interaction.

[1]See Fine (2005).
[2]See e.g. Correia and Rosenkranz (2012), Lipman (2015, 2016, 2018), and Loss (2017).
[3]See e.g. Pooley (2013) and Iaquinto (2018).
[4]See Hofweber and Lange (2017), and Lipman (Forthcoming).
[5]See Simon (2018).
[6]We are abusing terminology here, for arguably the terms in a superposition state, and the superposition itself are mathematical objects, whereas the quantum state is supposed to be the *referent* of those mathematical objects. This conflation is widespread in the literature and mostly harmless.

In effect, a few words of clarification are in order. Here and in what follows we use the term "interaction" in a specific sense. We don't mean to give a *definition* of *interaction*. Rather we want to provide an informal os the specific sense at issue here.[7] In this specific sense, we contend, x interacts with y iff x acts on y and y acts on x to produce effect z, or, equivalently, x acts *together* with y to produce effect z. This terminology is particularly useful in this context for it highlights that x and y have to obtain *together*, in order to produce z. Once again, we do not mean the previous bi-conditional to be read as definitions. Yet we can provide examples. There are certain dances where the dancers have to *act together* in order to pruduce certain figures. In a chemical reaction the reactants interact in this strict sense in order to produce a different substance or compound. They too, like the dancers, act together. In this specific sense, we claim, s_1 and s_2 *interact* only if they *obtain together*. But s_1 and s_2 can obtain together only if they belong to the same fragment. Thus, according to fragmentalism, state of affairs that belongs to different fragments *cannot* interact, as per premise $\mathbf{P}_1$.[8]

We will argue for premise $\mathbf{P}_2$ by way of an example. That is, we will invoke interaction of different superposition terms to explain some basic quantum phenomena such as the existence of an interference pattern in the double-slit experiment.

The conclusion is that fragmentalism, at least along the lines proposed by Simon, *does not* offer a new, satisfactory realistic account of the quantum state. This raises the question about whether there are some other viable forms of quantum fragmentalism.

References

Correia, F. and Rosenkranz, S. 2012. Eternal facts in an ageing universe. *Australasian Journal of Philosophy* 90: 307-320

Fine, K. 2005. Tense and reality. In his *Modality and Tense*. Oxford: OUP, pp. 261-320

Hofweber, T. and Lange, M. 2017. Fine's fragmentalist interpretation of special relativity. *Noûs* 51: 871-883

Iaquinto, S. 2018. Fragmentalist Presentist Perdurantism. *Philosophia*. Online first

[7]This is because we don't want to deny that there are other uses of the term "interaction" that do not conform to the characterization we provide in the main text.

[8]Clearly, if the states of affairs *cannot* interact, they *do not*.

Lipman, M. 2015. On Fine's fragmentalism. *Philosophical Studies* 172: 3119-3133

Lipman, M. 2016. Perspectival variance and worldly fragmentation. *Australasian Journal of Philosophy* 94: 42-57

Lipman, M. 2018. A passage theory of time. In K. Bennett and D. Zimmerman. Eds. *Oxford Studies in Metaphysics: Volume 11*, pp. 95-122

Lipman, M. Forthcoming. In defense of a fragmentalist interpretation of special relativity. *Philosophical Studies*

Loss, R. 2017. Fine's McTaggart: Reloaded. *Manuscrito* 40: 209-239

Pooley, O. 2013. Relativity, the open future, and the passage of time. *Proceedings of the Aristotelian Society* 113 (Part 3): 321-363

Simon, J. 2018. Fragmenting the wave function. In K. Bennett and D. Zimmerman. Eds. *Oxford Studies in Metaphysics: Volume 11*, pp. 123-145

Drug Agencies as Science Diplomats

Maria Laura Ilardo

Section: General Philosophy of Science

Title: Drug Agencies as Science Diplomats

Keywords: Science Diplomacy, Drug Agency, Policy-Making, Decision-Making Process, Scientific Institution

The relationship of science and policy-making has been increasingly scrutinized in view of its deep impact on both society and science itself (McGarity, & Wagner, 2010). In particular, four paradoxes affect such relationship: 1. "the simultaneous scientification of politics and the politicisation of science", increases the use of scientific expertise by policy-makers but not the degree of certainty; and it becomes de-legitimating; 2. however, given the putative legitimating function of authoritative scientific knowledge in politics, the general accessibility of such knowledge has led to a competition for expertise which intensifies controversies in policy-making, rather than alleviating them; 3. yet, any possible loss of authority of scientific expertise, does not automatically translate into policy-makers' loss of confidence on entrusted advisory bodies; 4) within this intricate framework, science is deeply needed, and at the same time scientific trust risks to be eroded (see also Weingart, 1999).

As a response to such issues, Science Diplomacy (henceforth SD), may be advanced a pivotal paradigm to public policy (Epping, 2020), where the scientific interaction is considered a diplomatic act. SD is defined as "the use of science, its methods and its philosophies in diplomacy" and "involves science in diplomacy, science for diplomacy and diplomacy for science" (Royal Society/AAAS, 2010; Gluckman, 2017).

By assuming the role of science diplomats, Regulatory Agencies (RAs) may coordinate the work of the different actors involved, while dealing with epistemic asymmetries characterizing the scientific ecosystem (Osimani, 2020), and moderating preference misalignments.

Indeed, it has been noticed that policy formulations and decision-making in 'risk' sectors such as medicines, food safety, disease prevention, and so on are more and more influenced by Regulatory Agencies (RAs) (Versluis, 2011; Kim, 2013). Therefore, national and, even more so, international RAs are progressively able to directly or indirectly shape policies, thereby consequently determining global economic, social and geopolitical equilibria.

This talk analyses the role of Drug Agencies (DAs) as science diplomats and presents key epistemic challenges, tentative solutions and future perspectives.

First and foremost, DAs ought to be neutral in the complex process where science serves regulatory, politic and also economic decisions. In this framework a DA can be described as a transparent and filtering balancer during a measurement process or as a referee during a game, who should be *super partes* by definition and equidistant from pure scientific and mere political concerns. Counterexamples in this respect can be observed at the EU level in the current management of the vaccination campaign, characterized by continuous clashes among pharmaceutical companies, DAs and EU Commission/National governments, where DAs recommendations tended to transpass their institutional boundaries, *de facto* dominating also the political decision making. DAs can end up having to choose "whether to engage in a manner that is consistent with science but that is sometimes at odds with the norms of the policy process or vice versa" (Higgings, 2006). Another related issue concerns the vagueness characterizing DAs performance indicators (see the Astra Zeneca controversy).

Which inferential procedures prevailed on DAs decision-making? Did political-economic influences outweigh scientific evidence? And, as Jasanoff (2009), "how can decisionmakers charged with protecting the public's health and safety steer clear of false and misleading scientific research?".

The paper defines what SD amounts to for DAs by offering an answer to the above questions.

References:

Epping, E. (2020). Lifting the smokescreen of science diplomacy: comparing the political instrumentation of science and innovation centres. *Humanities and Social Sciences Communications*, *7*(1), 1-13.

Gluckman, P. D., Turekian, V. C., Grimes, R. W., & Kishi, T. (2017). Science diplomacy: A pragmatic perspective from the inside. *Science & Diplomacy*, *6*(4), 1-13.

Higgins, P. A. T., K. M. A. Chan, and S. Porder, 2006: Bridge over a philosophical divide. Evidence & Policy, 2, 251–257.

Jasanoff, S. (2009). The fifth branch: science advisers as policymakers. *Harvard University Press*. ISBN 9780674300620

Kim, J., Klika, C., & Versluis, E. (2013). Agencies as risk managers? Exploring the role of EU agencies in authorisation procedures. *Balancing between trade and risk. Integrating legal and social science perspectives*, 175-196.

McGarity, T. O., & Wagner, W. E. (2010). Bending science: How special interests corrupt public health research. *Harvard University Press*. ISBN 0674047141, 9780674047143

Osimani, B. (2020). Epistemic gains and epistemic games: Reliability and higher order evidence in medicine and pharmacology. In *Uncertainty in Pharmacology* (pp. 345-372). Springer, Cham.

Royal Society/AAAS, New Frontiers in Science Diplomacy (London: Royal Society, 2010), https://royalsociety.org/topicspolicy/publications/2010/new-frontiers-science-diplomacy .

Versluis, E., Van Keulen, M., & Stephenson, P. (2011). *Analyzing the European Union policy process*. Macmillan International Higher Education.

Weingart, P. (1999). Scientific expertise and political accountability: paradoxes of science in politics. *Science and public policy*, *26*(3), 151-161.

Reconsidering multi-level mechanistic explanation

Stavros Ioannidis

Reconsidering Multi-level Mechanistic Explanation

According to a widespread ontological view associated with new mechanism, levels of nature typically invoked in explanations in life sciences are levels of mechanisms. On Craver's (2007) popular account, the relation between mechanistic levels is the relation between the mechanism as a whole and its components; in turn, this relation is considered to be a non-causal dependency relation, and is to be viewed in terms of mutual manipulability. Craver analyses mutual manipulability in terms of Woodward's concept of ideal intervention. So, Craver's mutual manipulability account serves both to give an account of the non-causal relations between the components and the whole mechanism, and to ground a hierarchy of mechanistic levels.

However, while this account seems to capture the practice of interlevel experiments in life sciences, it has recently come under criticism, as it has been argued that the concept of ideal intervention cannot be applied to constitutive relations (Harinen 2018, Kästner & Andersen 2018). While philosophers have responded to this problem by trying to reconcile interventionism with mechanistic constitution, or by finding an alternative way to characterise mechanistic constitution, there exists a third, and more radical, option: namely, to reject the view that constitution is needed to understand the notion of a mechanism. This option gives rise to the question, how shall we understand mechanistic levels and the relations between them.

In this paper I will first briefly review the main reasons for introducing constitution in the analysis of mechanism and the problems associated with accounting for constitution in terms of interventionism. I will then present a new argument that undermines the motivations behind Craver's account of constitutive mechanisms and mechanistic levels. This argument is based on the claim that typical and paradigmatic biological mechanisms are causal pathways. Lastly, I will sketch an alternative account of multi-level mechanistic explanation, by discussing various biological examples.

Craver describes the components of the mechanism as x1-φing, x2-φing etc, where the xs refer to the entities that comprise the mechanism and the term 'x-φing' refers to an entity engaging in an activity. He describes the phenomenon as S-ψing, where the S is a structure that ψs, and where S-ψing is the phenomenon that is taken to be constituted by the mechanism. Craver requires that all xs (i.e. all components of the mechanisms) be parts of S. For example, when a neuron generates an action potential, S is the neuron, S-ψing the neuron generating an action potential, and the x that φ are the various components of the mechanism for the generation of the action potential.

The main criticism against Craver's account will be that there is no biological motivation for supposing that in cases of biological mechanisms, the organised entities and activities are always parts of a larger entity, whose behaviour the mechanism underlies. Firstly, we cannot always find natural boundaries around mechanisms. Secondly, we cannot expect this: typical and paradigmatic biological mechanisms are causal pathways, which are not confined within biological objects/structures. In other words, in many cases there is no biological S, where every X (component of mechanism/pathway) is part of S. Thirdly, take a mechanism that occurs within a biological object, for example protein synthesis which occurs within cells. What is the S that ψ-s in this case? It cannot be the cell—the reason is that the mechanism can exist outside the cell. but then the parts of the mechanism (the Xs) can exist without the S. But whenever we have the mechanism, we necessarily have S-ψing. So, the S cannot be the cell. Can the S be the mereological sum of all the components of the mechanism? I will claim that this sum is not a 'natural' biological object, and it is better seen as an occurrent and not a continuant. All this shows that it is better to view typical and paradigmatic biological mechanisms as etiological (rather than 'constitutive').

I will then present an alternative view of multi-level mechanistic explanation. The main ingredients of this alternative view are three basic claims: (i) biological mechanisms are causal pathways, (ii) levels and mechanisms are distinct notions and (iii) levels of nature and of multi-level explanations are levels of composition. A key claim of this account is that whatever contributes to the phenomenon is part of the same pathway; but_causal pathways can contain entities at just one level of composition, or they may contain entities from multiple levels of composition. We therefore need to distinguish between ontological levels (i.e. levels of composition) and explanatory levels. While multi-level explanations contain entities from various ontological levels, ontological levels as such do not matter for explanation; what matters is the existence of a causal pathway.

So, according to this account, multi-level mechanistic explanations are causal explanations that identify particular causal pathways; but the components of the pathway are at different levels of composition: for example, a causal pathway may involve cells, hormones, and behavioural outcomes. This view of multi-level explanation is in stark contrast to the widespread view of multi-level mechanistic explanation as presented in Craver (2007), according to which 'levels' in multiple-level explanations are not levels of composition but levels of mechanisms and mechanistic multi-level explanations are instances of constitutive explanations.

In order to motivate and illustrate the view, I will use particular biological examples: the pathway of visual perception, mechanisms of cell death, developmental mechanisms (axis formation and determination mechanisms, mechanisms for the generation of the tetrapod limb) and mechanisms underlying behavioural responses. The main aim of this paper will be to show that this account of multi-level mechanistic explanation applies to many different biological examples.

References

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

Harinen, T. (2018) 'Mutual Manipulability and Causal Inbetweenness', *Synthese* 195: 35-54.

Kästner, L. & Andersen, L. M. (2018) 'Intervening into mechanisms: Prospects and challenges', *Philosophy Compass* 13: e12546.

Perspectivism and the Veridicality Problem in Non-causal Explanations

Daniel Kostic

Long abstract for the EENPS 2020

Sections: a) General Philosophy of Science
c) Philosophy of Cognitive and Behavioral Sciences

Title: Perspectivism and the Veridicality Problem in Non-causal Explanations

(Word count without the title and bibliography: 1272 words)

The ever-growing interest in non-causal explanations in sciences during the last decade has yielded several very sophisticated philosophical accounts (Batterman 2010; Lange 2018; Jansson and Saatsi 2019; Reutlinger and Saatsi 2018). In non-causal explanations, most broadly speaking, some non-causal facts (such as mathematical or metaphysical properties or relations) are used to explain some empirical facts. For example, some topological explanations in neuroscience (as one variety of non-causal explanations) describe how the brain dynamics counterfactually depend on mathematical properties of connectivity patterns in complex brain networks (Kostić 2020). Even though these and other non-causal explanations do not hinge on contingent empirical facts, but instead describe mathematical or metaphysical dependencies, they should not be seen as being completely divorced from physical reality, or in other words the question is how they can be successful explanations. The success of an explanation can be evaluated in several important ways. The first one is how to distinguish explanations from mere predications and descriptions? A broadly accepted strategy is to ensure that the explanation supports relevant counterfactuals, i.e. had A been different in certain ways the B would have been different in certain ways, thus B counterfactually depends on A. However, despite being successful in terms of supporting counterfactuals the non-causal explanations raise a further question, namely, how they can be true of physical phenomena if they abstract away from their causal or microphysical details (Pincock 2021)?

I call this the veridicality problem (VP hereafter).

There are cases in network neuroscience in which the veridicality problem emerges in a most direct way. Those are the cases which consider the relationship between the "structure" and the "function". The "structure" refers to networks of anatomical connections in the brain, also known as the structural connectivity (SC hereafter). In SC models, the connections between nodes are based on physical (rather than merely statistical) connections between brain areas. On the other hand, the "function" refers to various ways in which the information is transmitted in computations in the brain. Functional connectivity (FC hereafter) define edges based on statistical relations between area activity time series, such as a correlation coefficient, coherence, or synchronization index. Both of these types of connectivity are physically embedded into the 3D space of the skull. Such physical embedding should be guided by some natural constraints on development and evolution of brain networks.

The most salient feature of brain networks is unexpectedly short structural edges, also known as "wires". This feature is indicative of wiring minimization in the evolution and dynamics of brain networks (Stiso and Bassett 2018, 256). Presumably, wiring minimization allows for a very efficient information processing in the system. In terms of topological properties, wiring minimization is characterized by fewer long-range wires, which in turn facilitates redundancy and dynamical complexity. The wiring minimization in the human brain specifically, enables very complex topological features despite of significant constraints on wiring (Stiso and Bassett 2018, 257). To understand how wiring minimization differs across individuals as well across species, in healthy brains and in neurodevelopmental disorders such as schizophrenia, Stiso and Bassett suggest to look into the volumetric constraints on the wiring minimization (Stiso and Bassett 2018, 257). A way to do it is by examining the Rentian scaling properties of the 3D volume of the human skull. Such properties are assessed by calculating the Rent's exponent (which quantifies the fractal scaling of the number of connections to or from a region of the brain). In the context of brain networks, the Rent's exponent is computed by placing randomly sized boxes (which capture the volume of the human brain in three geometric dimensions), and then by counting the number of edges, crossing the boundary of a given box, as well as the number of nodes contained the box. Their explanation of how topological structure affects and determines cognitive function describes counterfactual dependencies between wiring minimization and physical Rentian scaling (Stiso and Bassett 2018, 259). These counterfactual dependencies do not capture the causal facts, thus, the spatially embedded networks are idealizations that provide non-causal explanation.

In this case, the explanation-seeking question is:

Why are characteristic path lengths short in spatially embedded brain networks in healthy subjects?

The answer is that the topological volumetric constraints determine the wiring costs in the evolution and development of brain networks, and wiring costs are inversely proportional to efficiency in both signal processing and establishing new connections. This also bears on understanding the differences in topological properties in health and in neurodevelopmental disease. For example, path lengths in healthy brains are short, which enables very efficient signal processing across brain areas. In contrast, path lengths are longer in epilepsy, Alzheimer disease, or schizophrenia, which given the same volumetric constraints of the human skull as in healthy brains, explains why in such disorders signal processing is inefficient or even disrupted.

At this point we can provide a more precise analysis of explanatory power in topological explanation:

(T1) The brain functional connectivity network (*a*) has a Rent's exponent of a certain value (*F*);
(T2) The brain functional connectivity network (*a*) displays wiring minimization of certain value (*G*);
(T3) Had the Rent's exponent been different (had it assumed a different value) the wiring minimization would have been different.
(T4) A is an answer to the explanation-seeking question Q about B, such that the Q determines the explanatory relevance of F to G.

The first condition, T1, distinguishes what kinds of properties figure in an *explanans*, and in that way determines whether an explanation is distinctively topological or some other kind. T2 ensures that *G* is a proper scientific *explanandum* (i.e. it is a description of an empirical phenomenon). The third condition T3, secures explanatoriness, i.e. the T3 captures the change-relating counterfactuals. Finally, the fourth condition provides contextual criteria for using the counterfactual. This condition provides a context which makes it intelligible why some empirical property G counterfactually depends on a network connectivity pattern, which is expressed as its topological property *F*. If such a context was not available, then it wouldn't have been intelligible how the relevant counterfactual figures in an explanation.

I argue that an approach to the VP based on the perspectival and pragmatic criteria embodied in the T4, adapted from Kostic (2020), Lange (2017) and de Regt (2017) provide a fruitful epistemological framework, that does not raise metaphysical problems relating to the concept of truth or causation.

This gives us a very plausible and straightforward way for responding to the veridicality problem. Instead of asking what are the truth conditions of a topological explanation, we could be asking a less metaphysically charged question such as: what are the perspectival constraints that determine explanatory relevance in order for an explanans to successfully apply to an explanandum? In the case study discussed in this paper it amounts to asking what are the volumetric constraints on wiring minimization in brain networks? As we have seen the perspectival constraints in this case can be assessed very precisely in terms of Rentian scaling, i.e. by computing a Rent's exponent for a 3D volume of a human skull. This approach then does not require any assumptions about notoriously difficult metaphysical notions such as truth or causation. All that is needed are assumptions about the typical 3D volume of the human skull, what is a range of values for that volume in a human population and the assumptions about the wiring costs in network models, to name just a few. In this way metaphysical commitments in such explanation are lessened while at the same time we also have a more precise answer to the question such as how can topological explanation be successful of its explanandum if the explanans involves non-causal facts.

References

- Batterman, R. W. (2010). On the explanatory role of mathematics in empirical science. *The British journal for the philosophy of science*, *61*(1), 1-25.
- De Regt, H. W. (2017). *Understanding scientific understanding*. Oxford University Press.
- Jansson, L., & Saatsi, J. (2019). Explanatory abstractions. *The British Journal for the Philosophy of Science*, *70*(3), 817-844.
- Kostić, D. (2020). General theory of topological explanations and explanatory asymmetry. *Philosophical Transactions of the Royal Society B*, *375*(1796), 20190321.
- Lange, M. (2018). A reply to Craver and Povich on the directionality of distinctively mathematical explanations. *Studies in history and philosophy of science*, *67*, 85.
- Pincock, C. (2021). A Defense of Truth as a Necessary Condition on Scientific Explanation. *Erkenntnis,* https://doi.org/10.1007/s10670-020-00371-9
- Reutlinger, A., & Saatsi, J. (Eds.). (2018). Explanation beyond causation: philosophical perspectives on non-causal explanations. Oxford University Press.
- Stiso, J., & Bassett, D. S. (2018). Spatial Embedding Imposes Constraints on Neuronal Network Architectures. *Trends in cognitive sciences*.

Was Jerzy Neyman a Perspectival Realist?

Adam Kubiak

Section: General Philosophy of Science

Title: Was Jerzy Neyman a Perspectival Realist?

Keywords: statistics, frequentism, realism, perspectivism, pragmatism

By referring to the example of the stopping rule problem (Lindley, Phillips 1976) I argue that perspectival realism (PR) (Crețu 2020; Massimi 2018) conforms naturally to frequentist statistical methodology. Nonetheless, PR and frequentism are incongruent as far as Neyman's (1950) methodological-philosophical frequentist conception is concerned.

What makes Neyman's stance close to perspectival realism is that scientific statements are designed to be expressed through observational and conceptual perspectives and at the same time they refer to the perspective-independent, states of affairs in the real world (Neyman 1950). Nonetheless, in line with his ideas, epistemic realism can only be accepted concerning a body of scientific outcomes. Additionally, Neyman is ambiguous in him distinguishing the real world from empirical observations. The particularly troublesome upshot of this is that statistical hypothesis can be "true", he believes, but at the same time, it is always "fictitious" (Neyman 1923).

A solution would perhaps be to tell that only a part of it can be strictly true about the real world—the value of the model's parameter, which accurately represents the mechanism or characteristic of some part of the real world. Moreover, to tell that the whole hypothesis, which is a statement not only about the value of the parameter but also about empirical predictions relative to empirical setup given this parameter value, is approximately empirically adequate. This approximate and not complete empirical adequacy of idealized description could be then understood as what Neyman meant by the fictitiousness of a scientific concept.

The bigger problem with reconciling PR with Neyman's frequentism is that Neyman avoids pluralism in adopting perspectives, whereas the assumption of this pluralism is a precondition for perspectivism. An interesting fact is that Neyman is internally inconsistent by appreciating the element of decisiveness in individual choices that influence the outcome when he speaks of frequentism, but condemns making choices that affect the outcome when he criticizes Bayesianism (Neyman 1957). Surprisingly, the only acceptable by Neyman aspect of appreciating perspective pluralism is the aspect of adopting error risks based on practical considerations that turn out to be opposing PR. Nevertheless, there are some aspects where pluralism of perspectives seems to be unavoidable. A way to reconcile Neyman's views with PR could be a case (or aspect)-dependent relativization of PR.

As Neyman's frequentism is not the only paradigm in philosophy of statistics PR require further reflection from the viewpoint of alternative to Neyman's methodological-philosophical approaches.

References.

Creţu, Ana-Maria. 2020. "Perspectival Realism." In Encyclopedia of Educational *Philosophy and Theory*, ed. M. A. Peters. Singapore: Springer.

Lindley, Dennis V. and Phillips, Lawrence D. 1976. "Inference for a Bernoulli Process (a Bayesian View)." *The American Statistician* 30 (3): 112–119.

Massimi, Michela 2018. "Perspectivism." In *The Routledge Handbook of Scientific Realism*, ed. J. Saatsi, 164-175. Oxford: Routledge.

Neyman, Jerzy 1957. "'Inductive Behavior' as a Basic Concept of Philosophy of Science." *Revue De L'Institut International De Statistique* 25 (1/3): 7–22.

Neyman, Jerzy 1950. *First Course in Probability and Statistics*. New York: Henry Holt and Co.

Neyman, Jerzy. Spława-Neyman, Jerzy. 1923. “Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych.” Roczniki Nauk Rolniczych i, Leśnych 5: 1–51.

Practices as Social Knowledge

Juho Lindholm

a) General Philosophy of Science

Practices as Social Knowledge

Keywords
John Dewey, epistemology, philosophy of science, social knowledge, practice

Short Abstract

In this presentation, I will argue that John Dewey's notion of scientific knowledge is inherently social in four ways. He can be understood as conceiving knowledge as practice. First, knowledge as practice is an adaptation of the behavior of an organism to an environment, of which other organisms are a part. Second, knowledge as practice is public and hence sharable. Third, Dewey conceived knowledge as a vehicle of social and moral reform. Fourth, the value of knowledge is in improvement of the human condition.

Abstract

In this presentation, I will argue that John Dewey's notion of scientific knowledge is inherently social in four ways and explain the theoretical background of that notion.

The notion of *habit* is pivotal in classical pragmatism, especially in John Dewey's large output. His most detailed account on the nature of habits is available in his *Human Nature and Conduct* (1922). He argues that habits are the vehicle of experience and cognition. Far from being mindless repetition, Deweyan habits require mental acuity: acting habitually, the agent is still learning and adapting existing knowledge to novel conditions (cf. Ryle [1949] 1951: ch. 2). Thus he argues that *habits are arts*. He shows how this notion of habit has tremendous significance in almost all of philosophy: epistemology, philosophy of science, philosophy of mind, ethics, social and political philosophy, and especially philosophy of education.

He extended his theory in *Experience and Nature* ([1925] 1929a), *The Quest for Certainty* (1929b) and *Logic* (1938), in which he has generalized his treatment from *habit* to *experience*. What he had said about habits in *Human Nature and Conduct* remains largely implicit in these later works. In his late philosophy, experience does not mean mere ”appearance” as opposed to ”real” (rather, “The world as we experience it is a real world” (Dewey 1929b: 295)) or private mental entities interfering between real objects and concepts. Rather, his notion of experience is modeled as organism—environment interaction and can be equated with *experiment*. Experiment involves the mind and the body, and also the environment and “external” instruments. Hence his notion of experience is mainly objective.

Dewey's ideas have been forgotten for decades in mainstream analytic philosophy, but they might be coming back in sociology of scientific knowledge (SSK) and science and technology studies (STS). Joseph Rouse (1987; 1996; 2002) has argued that scientific knowledge is scientific practice. It is possible that his notion of practice captures better what Dewey intended to say by “habit,” “kind of action,” or “experience” without significant distortion. Hence, I will use these terms as synonymous.

In *Experience and Nature* ([1925] 1929a: ch. IX), Dewey argues that science is an art, or that science is the intelligent factor of any art. The purpose of all art, whether aesthetic or technological, is the improvement of the environment; and if that involves a critically examined method, he calls it “science.” But, on the other hand, Dewey (1929b) can be understood as denying the distinction between everyday problem-solving and science: we all use the scientific method when dealing with mundane tasks intelligently. Thus he posits a continuum of arts between ordinary problem-solving and science.

In *The Quest for Certainty* (1929b), Dewey argues that “knowledge is a kind of action.” He means that particular actions have a general kind and that kind equals knowledge. He does not explain what this kind is like in detail, but it can be surmised that he means *habit* or *practice*. This is a refutation of the notion that scientific knowledge be a disembodied, abstract system of representations (propositions, theories) which is independent of individuals and society and to which concrete action is merely an accidental addendum.

In *Logic* (1938: chs. II—IV), Dewey argues that inquiry (scientific or everyday) takes place in a

biological and a social matrix. Habits are adaptations of the behavior of an organism to the environment; and other organisms and their reactions to the agent's behavior are ineliminably part of the environment. Hence the establishment of habits cannot but take society into account. This is the first sense in which habit- or practice-knowledge is inherently social.

Another social feature of habit- or practice-knowledge is that as a public entity it is sharable. Dewey does not, however, deny the effect of individuals and their idiosyncrasies on the formation of habits. But habits, which are already social, largely (but not necessarily entirely) constitute individuality, rather than already self-sufficient individuals constitute society. It can be said that the relation between society and individuals is an open-ended, ongoing negotiation between already established habits and innovations which might emerge also from individual initiative but which can become new habits. It is at least arguable that, in this way, Dewey's notion of habit overcomes the dichotomy of the individual and society.

The third social facet of habits is that Dewey modeled his program of social and moral reform on them. He conceived education as the vehicle of such reform through improving customs (Dewey 1916; 1920; 1922). In his large literary output, there is no meaningful distinction between his philosophy in general and his philosophy of education. The slogan “learning by doing” derives from Dewey (1916).

Like his general epistemology, his philosophy of science is geared for social progress. He considers science as an agent in social reform. He argued that the value of knowledge is in the intelligent regulation of the environment and society facilitated by science. In the state of nature, goods can be insecure, evanescent, and available only for the upper class. But through technological application of knowledge they can be rendered more secure, more regular, and available more equitably. This is one more sense in which knowledge is social: it can be used to improve social conditions. In this sense, Dewey was very Baconian and optimistic with regard to technology – an optimism which has been questioned by many, such as Max Horkheimer (1947), Max Horkheimer and Theodor W. Adorno (1947), Martin Heidegger ([1953] 2000), Jacques Ellul ([1954] 1964), Herbert Marcuse (1964), and Lewis Mumford (1967; 1970).

References

Dewey, John (1916): *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan.

Dewey, John (1920): *Reconstruction in Philosophy*. New York: Henry Holt and Company.

Dewey, John (1922): *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Henry Holt and Company.

Dewey, John ([1925] 1929a): *Experience and Nature*. (2nd Edition.) London: George Allen & Unwin, Ltd.

Dewey, John (1929b): *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York: Minton, Balch & Company.

Dewey, John (1938): *Logic: The Theory of Inquiry*. New York: Henry Holt and Company.

Ellul, Jacques ([1954] 1964): *The Technological Society*. New York: Alfred A. Knopf.

Heidegger, Martin ([1953] 2000): “Die Frage nach der Technik,” in *Gesamtausgabe, Band 7* (F.-W. von Herrmann, ed.), Frankfurt am Main: Vittorio Klostermann, pp. 5—36.

Horkheimer, Max (1947): *Eclipse of Reason*. Oxford: Oxford University Press.

Horkheimer, Max and Adorno, Theodor W. (1947): *Dialektik der Aufklärung: Philosophische Fragmente*. Amsterdam: Querido Verlag.

Marcuse, Herbert (1964): *One-dimensional Man: Studies in Ideology of Advanced Industrial Society*. London: Routledge.

Mumford, Lewis (1967): *The Myth of the Machine, Vol. 1: Technics and Human Development*. New York: Harcourt Brace Jovanovich, Inc.

Mumford, Lewis (1970): *The Myth of the Machine, Vol. 2: The Pentagon of Power*. New York: Harcourt Brace Jovanovich, Inc.

Rouse, Joseph (1987): *Knowledge and Power: Toward a Political Philosophy of Science*. Ithaca and London: Cornell University Press.

Rouse, Joseph (1996): *Engaging Science: How to Understand Its Practices Philosophically*. Ithaca and London: Cornell University Press.

Rouse, Joseph (2002): *How Scientific Practices Matter: Reclaiming Philosophical Naturalism*. Chicago and London: The University of Chicago Press.

Ryle, Gilbert ([1949] 1951): *The Concept of Mind*. Watford, Herts: William Brendon and Son, Ltd.

Can all natural properties and relations be powers? The case of manifestation-relations

Vassilis Livanios

Section: General Philosophy of Science

Title: Can all natural properties and relations be powers? The case of manifestation-relations

Keywords: powers; manifestation; dispositional monism; manifestation-relations

Short abstract

The main goal of this paper is to show that the widely accepted view that powers can exist unmanifested is inconsistent with the view (known as dispositional monism) that all fundamental natural properties and relations are powers. To this end, two kinds of manifestation-relation (token-level and type-level, respectively) are introduced and it is then argued that dispositional monists need the former in order to offer a metaphysically clear view of the transition from non-manifestation to manifestation of a power. Finally, an argument is given that the token-level-manifestation-relations are not powers.

Extended abstract

Most metaphysicians agree that an essential feature of powers is that they can exist without being manifested. The main goal of this paper is to show that this thesis is inconsistent with the view (known as dispositional monism) that all fundamental natural properties are powers. To this end, first, I introduce two kinds of manifestation-relation, token-level and type-level, respectively. It is essential to token-level-manifestation-relations that one of their relata is the manifestation of a power, while at least one of the other relata is a particular instantiating that power. Analogously, it is essential to type-level-manifestation-relations that one of their relata is a universal 'included' in the manifestation of the relevant power, while at least one of the other relata is the relevant power-type.

Second, I suggest that an actually instantiated power is not manifested iff the associated with that power token-level-manifestation-relation is not instantiated. The main objection to this suggestion is that power realists often think of an unmanifested power as one whose manifestation is not *actualised* and some of them maintain that a power is related to its manifestation(s) *even if the latter is (are) unactualized.* To address this objection, I discuss two ways of understanding the relation between an unmanifested power and its possible manifestations. According to the first understanding, the relation in question is a kind of *directedness* (for some philosophers, a kind of physical intentionality) of powers towards their

possible manifestations. According to the second understanding, the relation under consideration holds between a power and its possible manifestations only provided that the latter are conceived as *actually* existing, contingently abstract, entities (unrealised possibilia). I offer some reasons why those two conceptions of the relation between a power and its possible manifestations cannot provide an adequate metaphysical account of what is going on in the course of the transition from non-manifestation to manifestation. In contrast to that, I argue that my suggested view offers a metaphysically clear account in terms of the instantiation of token-level-manifestation-relations and so power realists have good reasons to embrace it.

After that, I argue that in order for token-level-manifestation-relations to play their explanatory role (that is, to explain what is metaphysically going on when a power becomes manifested) they should be *necessarily* manifested. Given, however, that powers can exist unmanifested, this conclusion implies that token-level-manifestation-relations cannot be powers. Hence, dispositional monists need non-powers in their ontology because when a power 'passes' from non-manifestation to manifestation, a token-level-manifestation-relation (that is, a non-power) is instantiated. Dispositional monists who admit that powers *can* exist unmanifested cannot refute that such transitions occur on pain of committing themselves to the implausible view that all actual powers *always* exist unmanifested. Therefore, dispositional monists cannot hold that powers can exist unmanifested unless they simultaneously admit the existence of relations that are not powers.

Finally, I address two objections to the above conclusion. The first objection is that token-level-manifestation-relations, *qua internal*, constitute no genuine addition of being and so dispositional monists should not worry whether they are powers or not. The second objection is that my conclusion seems redundant as far as the viability of dispositional monism is concerned because Stephen Barker has already argued that the existence of *type*-level-manifestation-relations is inconsistent with dispositional monism since there are cogent reasons to believe that they are not powers.

Time's Direction at the Plank Scale

Cristian López

Time's Direction at the Plank Scale

b) Philosophy of Natural Science

Keywords: philosophy of physics – quantum mechanics – time's direction – time-reversal symmetry

Short abstract (100 words)

Though nature seems to exhibit a time's direction, physics does not. One of the main reasons is that fundamental dynamical laws are blind to any distinction between the past-to-future and the future-to-past direction since they, to a great extent, have the property of being invariant under the direction of time. This has risen several concerns as to the relation between the macro-scale and the micro-scale. In this presentation, I will claim that we have actually good reasons to take time to be directed at the quantum scale, despite contrary opinions on the matter.

Extended abstract (500-1000 words)

Time seems to be directed. We not only think that both directions of time are different, but also that there is a huge temporal bias in nature: past-to-future processes enormously outnumber future-to-past ones. However, when we look at physics, things are not so straightforward: it has been argued that, at the fundamental level, physics is blind to the direction of time since most dynamical laws are invariant under time reversal (Reichenbach 1956, Horwich 1987, Price 1996, Callender 1997, Maudlin 2002), that is, they remain unaltered under a change of the direction of time. The *locus classicus* of the discussion has largely been the relation between thermodynamics and classical statistical mechanics (Earman 1974, Callender 1997): even though we commonly find temporally asymmetric process at the thermodynamical level (mainly given by entropy-increasing processes), the classical statistical descriptions underlying is rather temporally symmetric. This has risen several concerns in philosophy of physics and philosophy of time that have in general boiled down to provide an answer to the following questions: how does temporally asymmetric processes emerge from temporally symmetric ones?

The situation doesn't change so much as we step into the Plank scale: quantum theories, it's widely claimed, are also to great extent time-reversal invariant. This (seemingly) uncontroversial result just deepens the problem: temporally asymmetric processes at the macroscopic scale finds temporally symmetric processes all the way down. For instance, if we look to non-relativistic quantum mechanics (the first quantum theory we find as we enter the quantum realm), we find that the unitary evolutions of elementary quantum systems, given by the Schrödinger equation, turns out time-reversal invariant (Messiah 1966, Sachs 1987, Penrose 1989, Roberts 2017). Temporally asymmetric behaviors in fact start to come up when interactions or forces come into play, but isolated quantum systems don't exhibit any distinction between both directions of time. Consequently, it is claimed that non-relativistic quantum mechanics doesn't treat the past-to-future and the future-to-past direction differently, at least in the simplest and most relevant cases (XX)

In this presentation I would like to temper this extended idea. My argumentation will be twofold. First, I will claim that whether non-relativistic quantum mechanics turns out time symmetric or not depends on which *interpretation* is endorsed. The idea that non-relativistic quantum mechanics is time symmetric mainly rests upon its unitary part (or the *bare* formalism), but many interpretations introduce new non-unitary dynamical elements that get the theory time asymmetric. This is case for collapse theories, like GRWm or GRWf, but it can be also extended to other interpretations like many worlds or Bohmian mechanics. The upshot is that non-relativistic quantum mechanics, at best, turns out time symmetric in a philosophically uninteresting presentation of the theory, but as soon as we adopt a more robust one, new temporally asymmetric elements are rapidly introduced. Second, I will argue that even the bare formalism (that just involving unitary evolutions given by the Schrödinger equation) comes out time symmetric under certain assumptions. If such assumptions are challenged, the theory may come out temporally asymmetric. This point relates to how the time-reversal transformation is defined within the theory. Contrarily to the conventional wisdom on the matter, I will claim that the formal definition of quantum time reversal is not philosophically neutral, but alternatives can be also given. Some of them render the Schrödinger equation non-time-reversal invariant. This argument goes along Callender's and Albert's views (Callender 2000, Albert 2000), though the approach is different. The home-take message of this presentation will be that, despite what is commonly believed, there are good and sound reasons to take time to be directed at the Plank scale.

References

Albert, D. Z. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Callender, C. (2000). "Is time 'handed' in a quantum world?" *Proceedings of the Aristotelian Society*, 100: 247-269.

Callender, C. (1997). "What is "the problem of the direction of time'?". *Philosophy of Science* (Supplement), 64: S223-S234.

Earman, J. (1974). "An attempt to add a little direction to 'The Problem of the Direction of Time'", *Philosophy of Science*, 41: 15-47.

Horwich, P. (1987). *Asymmetries in Time*. Cambridge, MA: MIT Press.

Maudlin, T. (2002). "Remarks on the passing of time", *Proceedings of the Aristotelian Society*, 102: 237-252.

Messiah, A. (1966). *Quantum Mechanics.* New York: John Wiley and Sons.

Penrose, R. (1989). The *Emperor's New Mind*. New York: Oxford University Press.

Price, H. (1996). *Time´s Arrow and Archimedes' point: New Directions for the Physics of Time*. New York: Oxford University Press.

Reichanbach, H. (1956). *The Direction of Time.* Berkeley: University of California Press.

Roberts, B. (2017). "Three myths about time reversal invariance". *Philosophy of Science,* 84, 2: 315-334.

Sachs, R. (1987). *The Physics of Time Reversal*. London: University Chicago Press.

Mathematical Proofs: Two Kinds of Unification

Michele Lubrano

Submission for EENPS21 Conference

Area: General Philosophy of Science

Mathematical Proofs: Two Kinds of Unification (Extended Abstract)

Explanation in mathematics is a hotly debated issue. The best part of the literature on the topic is devoted to the attempt of answering the question: what makes a mathematical proof explanatory? Indeed, working mathematicians generally agree on the fact that some proofs, besides ensuring the correctness of a result, also let them understand the reason why the proved theorem holds, while others don't. The old Aristotelian distinction between 'demonstrating *that* P is true' and 'demonstrating *why* P is true' (see *Posterior Analytics* I, 13) seems particularly timely here. Today most philosophers of mathematics believe that there is no unique definition of explanatory proof and that there are different properties that can make a proof explanatory (see Colyvan, Cusbert, and McQueen 2018).

One of those properties is known as *Unificatory Power*. There is no standard characterization of it to be found in the literature, so I'll tentatively propose this loose definition:

> (**UP**) A mathematical proof of a theorem θ displays unificatory power if and only if it derives θ in a way that highlight θ's connection with other parts of mathematics, making θ appear as a particular manifestation of a more general and pervasive pattern.

This definition is gathered from what Kitcher (1989) says about scientific explanation as theoretical unification. The advantage of theoretical unification in the empirical sciences is pretty clear: unifying means reducing the number of facts that should be taken as primitive. In Kitcher's opinion his model can be successfully applied also to mathematics, but no systematic attempt in this direction has ever been made. In my contribution I would like to fill this gap and explain how a mathematical proof can display unificatory power and why this power makes it explanatory.

I'll start by showing that a close analysis of proofs that, according to working mathematicians, display unificatory power shows that there are at list two distinct patterns of unifications at work.

The first is what I would call *vertical unification*: it is the kind of unification that is generated by those proofs such that, in order to demonstrate θ, they first demonstrate a more general and abstract statement χ, from which θ is successively deduced. I'll argue that the proofs that exemplify this pattern promote unification by showing that the theorem to be proved is one of the many consequences of a deeper and more abstract mathematical truth. I'll provide, as an example of this phenomenon, a proof of the Isomorphism Theorem for Boolean Algebras that starts from a much more abstract result, namely the Isomorphism Theorem in Universal Algebra, and deduces the former from the latter.

The second pattern is what I would call *horizontal unification*: it is the kind of unification that is generated by those proofs that in order to demonstrate θ they first show that θ is equivalent to another statement φ, belonging to a different subfield of mathematics, and then they prove φ. Statement φ is not necessarily more general or abstract of θ, it simply belongs to another research area. I'll argue that, in this case, unification is provided by the "bridging" of two different and apparently distant mathematical theories. I'll offer, as an example of this pattern of unification, a proof of Fermat's Little Theorem (a theorem of Number Theory) based on Lagrange Theorem (a theorem of Group Theory).

I'll argue that the two patterns of unification cannot be reduced in to one another in any satisfactory way. Hence, any reasonable definition of the property generally called Unificatory Power cannot but be disjunctive in its form.

I'll finally show that a proof that displays unificatory power should be considered explanatory on the basis of human cognitive architecture. As it is convincingly argued in Inglis & Mejía-Ramos (forthcoming), the limited capacity of human working memory is a decisive factor in our understanding of mathematics. I'll argue that if "explanatory proof" is understood as "proof that gives/improves understanding", then those proofs that allow us to cover more truths with less working memory effort are those that mathematicians are inclined to consider more explanatory.

References

Colyvan, Mark, Cusbert, John, and McQueen, Kelvin (2018). 'Two Flavours of Mathematical Explanation'. In Reutlinger, Alexander, and Saatsi, Juha (Eds.), *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanation*, Oxford University Press.

Inglis, Mattew, and Mejía-Ramos, Juan Pablo (forthcoming). 'Functional Explanation in Mathematics', *Synthese*.

Kitcher, Philip (1989). 'Explanatory Unification and the Causal Structure of the World'. In Kitcher, Philip, and Salmon, Wesley (Eds.), *Scientific Explanation*, Minnesota University Press.

The meaning of causal claims in biomedicine and its implications for evidence-based medicine

Mariusz Maziarz

Short Abstract:

The purpose of the study is twofold. First, I argue that biomedicine accepts moderate causal pluralism as an epistemic definition of causality. Second, investigate the implications of the pluralism of notions of causality for clinical practice. I differentiate among 'actions' that do not change the relata of causal claims, (mechanistic) 'interferences,' and 'interventions' sensu stricte that act on causes to change effects. My central claim is that causal claims presupposing diverse concepts of causality can deliver evidence for different types of evidence-based policy.

The fully-fledged abstract

My purpose is to argue that, indeed, moderate causal pluralism is the epistemic view on causality implicitly accepted by biomedical researchers. Furthermore, I argue that the pluralism of epistemic concepts of causality present in biomedicine has serious implications for clinical practice. The structure of the paper is as follow. First, I describe the use of referential semantics to reconstructing the meaning of causal claims. Second, I conduct in-depth case studies of contemporary biomedical research and reconstruct the epistemic concepts of causality implicitly accepted by researchers. I argue that the four studies presuppose concepts of causality in agreement with the regularity, probabilistic, mechanistic, and manipulationist approaches to causality. This allows me to conclude that causal pluralism is the adequate view on causality regarding biomedicine (in its entirety). Third, I investigate the implications of causal pluralism for clinical practice. My main claim is that causal claims presupposing diverse concepts of causality can deliver evidence for different types of actions. Therein, I introduce the notion of 'translating' the meaning of causal claims in reference to misuses of causal evidence for clinical practice. The current philosophical views on what is the meaning of causality presupposed by biomedical research are vastly divided. The differentiation may indicate that the results of reconstructing the view on causality presupposed by biomedical researchers depend on the choice of case studies. This implies that different biomedical studies presuppose different concepts of causality. I use referentialist semantics as a tool for studying the meaning of causal claims in biomedicine and take the types of relations that can be discovered with research methods used in a given study as the reference of causal-family words present in the causal claim put forward by researchers.

The argument proceeds with case studies that are representative of the main approaches to causal inference in contemporary biomedicine. First, I analyze the Centre for Disease Control and Prevention (CDC) search for the cause of EVALI (vaping-related lung injury) (Blount et al. 2019) and relate it to the regularity view on causality. Second, I reconstruct the probabilistic view on causality from Reichenberg et al. (2006) cohort study of paternal age and ASD. Third, I analyze the research of Ratnayake et al. (2018) on the influence of blue light on age-related

macular degeneration in humans and relate it to the mechanistic view on causality. Fourth, I study the RECORD trial (Home et al. 2007) aimed at assessing the effectiveness of rosiglitazone and interpret it as presupposing a version of the manipulationist view on causality. Given this, moderate causal pluralism is a stance adequate to research practices in medicine. However, accepting different epistemic concepts of causality has severe implications for clinical decisions. I argue that the evidence for regularity and probabilistic relations suffices for 'actions' that do not modify the relata of causal claims. The reason is that the possibility of confounding cannot be excluded. This allows for undertaking evidence-based decisions even without the knowledge of full causal structure. What follows, putting observational studies at the bottom of the evidence pyramid is not justified for the actions that do not require knowledge of invariance under intervention. In contrast, mechanistic causal claims do not warrant the success of interventions because the represented mechanism may be screened off by other mechanisms. Finally, clinical trials allow for estimating average treatment effects that suffice for putting forward type-level causal claims that are invariant under intervention and therefore conducting 'interventions' in the strict sense. My work can also serve as an example of using referential semantics to study the meaning of philosophical concepts implicitly used in sciences.

References:

Ratnayake, K., Payton, J. L., Lakmal, O. H., &Karunarathne, A. (2018). Blue light excited retinal intercepts cellular signaling. Scientific Reports, 8(1), 10207.

Vandenbroucke, J. P., Broadbent, A., & Pearce, N. (2016). Causality and causal inference in epidemiology: the need for a pluralistic approach. International Journal of Epidemiology, 45(6), 1776-1786.

Blount, B. et al. (2019). Evaluation of Bronchoalveolar Lavage Fluid from Patients in an Outbreak of E-cigarette, or Vaping, Product Use-Assiciated Lung Injury – 10 States, August-October 2019, Morbidity and Mortality WeeklyReport/ November 15, 2019 / 68(45);1040-1041

Reichenberg, A., Gross, R., Weiser, M., Bresnahan, M., Silverman, J., Harlap, S., ... &Knobler, H. Y. (2006). Advancing paternal age and autism. Archives of general psychiatry, 63(9), 1026-1032.

Home, P. D., Pocock, S. J., Beck-Nielsen, H., Gomis, R., Hanefeld, M., Jones, N. P., ... & McMurray, J. J. (2007). Rosiglitazone evaluated for cardiovascular outcomes—an interim analysis. New England Journal of Medicine, 357(1), 28-38.

Practical Realism as Realism

Peeter Müürsepp

Section a
Practical Realism as Realism

Keywords: constructivism, Kantianism, practical realism, realism, reality

Practical realism is a relatively new version practical approach to the philosophy of science started recently by the late Estonian philosopher of science Rein Vihalemm. Vihalemm claims that his approach goes against anti-realism as well as standard scientific realism. Given the emphasis on practical research activity and the constructive nature of physics-like science, practical realism does not cohere with any kind of realism too well. Still, it is possible to show that practical realism can be taken as a version of realism, although quite different one from naïve, metaphysical or internal ones.

Practical realism is a relatively new version of realism started recently by the late Estonian philosopher of science Rein Vihalemm. Vihalemm presented the core of his approach by the five tenets of practical realism:

1. Science does not represent the world "as it really is" from a god's-eye point of view. Naïve realism and metaphysical realism have assumed the god's-eye point of view, or the possibility of one-to-one representation of reality, as an ideal to be pursued in scientific theories, or even as a true picture in the sciences.
2. The fact that the world is not accessible independently of scientific theories – or, to be more precise, paradigms (practices) – does not mean that Putnam's internal realism or "radical" social constructivism is acceptable.
3. Theoretical activity is only one aspect of science; scientific research is a practical activity and its main form is the scientific experiment that takes place in the real world, being a purposeful and critical theory-guided constructive, as well as manipulative, material interference with nature.
4. Science as practice is also a social-historical activity, which means, amongst other things, that scientific practice includes a normative aspect, too. That means, in turn, that the world, as it is accessible to science, is not free from norms either.
5. Though neither naïve nor metaphysical, it is certainly realism, as it claims that what is "given" in the form of scientific practice is an aspect of the real world. Or perhaps more precisely, science as practice is a way in which we are engaged with the world.

The last tenet expresses almost declaratively that we are dealing with a realism here. However, if we look at the tenets in conjunction then the realist essence is not obvious. Different philosophers of science have raised the issue, whether practical realism is realism at all. Perhaps such kind of approach to science should rather carry the label of empirical constructivism or something of the kind. Rein Vihalemm has added logs to the fire himself by claiming that his new approach goes not just against anti-realism but against standard scientific realism as well. The latter considers truth to be the aim of all scientific research. Vihalemm, however, supported the deflationary attitude towards truth. Vihalemm took exact natural science as constructive-hypothetico-deductive by its nature. Consistently with tenet 1 Vihalemm means that the scientist cannot study nature as it is but has to construct the object of research for herself. Doesn't such view go against any kind of realism, not just naïve and metaphysical ones? It rather looks as pure constructivism. Tenet 2 may provide us with a clarification. However, we need to resolve the issue why internal realism and "radical" social constructivism are not acceptable. According to Vihalemm, both approaches contradict themselves, as one cannot construct just anything she likes. Reality would resist. We have an important turn towards realism here. Exact natural science can be constructive-hypothetico-deductive by its essence but the scientist cannot construct any kind of object for research but just those that reality allows her.

Still, the issue of Putnam's internal realism is somewhat more sophisticated. Vihalemm and Putnam agreed on denying the god's-eye point of view and tried to avoid metaphysical realism. However, in internal realism objects do not exist independently of conceptual schemes in the human mind. If these conceptual schemes are the same thing Vihalemm calls theories then there is the question why the latter cannot accept internal realism. Therefore, they cannot be the same thing. Let us remember that Vihalemm prefers to call theories paradigms or even practices emphasizing the practical essence of his approach as well as the connection with reality. Practice was not at stake for Putnam but this does not necessarily mean that his conceptual schemes are something different from Vihalemm's practices in the sense of paradigms. Still, Vihalemm has claimed that internal realism is not realism as it prescribes something to reality through the prism of the conceptual schemes. However, what about Vihalemm's own approach. He has openly admitted the Kantian roots of practical realism stressing that philosophy of science of today cannot ignore Kant's "Copernican revolution". This obviously means accepting Kantian apriorism. It is not that our knowledge of objects has to conform to objects but objects have to conform to our knowledge. This Kantian claim adheres to Vihalemm's idea of physics-like science being constructive-hypothetico-deductive. The scientist can have a prior knowledge of her research object by means of constructing it. However, there is again the question, is it realism. Actually, it is from the point of view of practical realism. The essence of the explanation is the same as in the case of social constructivism. The scientist can have an *a priori* knowledge of her research objects but this knowledge can develop only in constant practical contact with reality. Nothing can be done that reality does not accept. This type of realism is really certainly not a naïve one and not metaphysical either. It does not assume the god's-eye point of view. As explained with the help of Kantian apriorism, the scientist can interact only with bits and pieces of reality, the ones that fit into her cognitive system. In conclusion, practical realism is realism, although perhaps a somewhat special one.

Structural Realist Theory of the Self

Janko Nešić

Philosophy of Cognitive and Behavioral Sciences

Structural Realist Theory of the Self

Keywords: structural realism, structure, object, individual, self, deflation

Abstract: My goal is to investigate if a structuralist theory of the self can be defended successfully. Some proponents of structural realism have pushed towards developing theories in other domains beyond fundamental physics: chemistry, biology, economics, and cognitive science. The question is posed if a version of (ontic) structural realism could be further developed within the domain of consciousness. In particular, could the subject of experience (the self) be explained by appealing to structures?

There are no objects in ontic structural realism or, if there are any carrier entities, they are grounded in the structure, and individuality is contextual rather than intrinsic. The "no object" problem of ontic structural realism seems particularly pressing if the theory is to be applied to the case of the self (the subject of experiences).

I consider a prime example of a structuralist theory of the self (Beni 2019) that aspires to provide a viable alternative to classical substantivalism. It amends the shortcomings of Metzinger's (2003) "self-model" eliminativism and pluralism of Gallagher's (2013) pattern theory. This theory is motivated by an analogous problem as structural realism - the underdetermination of the metaphysics - in this case of the self by various scientific theories of cognitive neuroscience. The structural realist theory of the self emphasizes the role of cortical midline structures of the brain in the processing of pieces of self-related and self-specific information. It is the information processing in the brain's cortical midline structures that would be the realizer of the basic structure of the self. In my talk, I will pose several problems for an ontic structural realist theory of the self.

This particular theory purports to be a non-eliminativist version of ontic structural realism that retains a thin notion of individual objects, "weakly discernible individual selves" and "weakly discernible nonstructural aspects of the self". Beni claims that such features like the sense of agency, sense of ownership and mineness are identified in virtue of their location in the structure of selfhood. I argue for this both from within the theory itself (to show it's empirical inadequacy) and from a general perspective of ontic structural realism itself. I give arguments that if the structural realist theory of the self is a moderate version of ontic structural realism it is in danger of collapse into eliminativist structural realism since the inflation of ontological priority of structure/relations precludes any notion of intrinsic nature of the relata. In the end, I will explain that a structural realist theory is only viable if one adopts a deflationary view of the self/subject.

References:

Beni, M. D. (2019). Structuring the self. Series New Directions in Philosophy and Cognitive Science, Palgrave Macmillan.
Chakravartty, A. (2012) “Ontological Priority: The Conceptual Basis of Non-Eliminative, Ontic Structural Realism” in E. Landry and D. Rickles (eds), Structure, Object, and Causality. The Western Ontario Series in Philosophy of Science. Dordrecht: Springer.
Esfeld, M. and Lam, V. (2008) “Moderate Structural Realism about Space-time”. Synthese 160: 27–46.
Gallagher, S. (2013). “A pattern theory of self”. Front. Hum. Neurosci. 7:443
Metzinger, T. (2003). Being No One. Cambridge, Mass.: MIT Press.
Ladyman, J. (2007) “Scientific Structuralism: On the Identity and Diversity of Objects in a Structure”. The Proceedings of the Aristotelian Society 81 (1): 23–43.

Structural Realist Theory of the Self

Abstract: My goal is to investigate if a structuralist theory of the self can be defended successfully. I consider a theory of the self (Beni 2019) that aspires to provide a viable alternative to substantivalism. If this theory is a moderate version of ontic structural realism it will collapse into eliminativist version since the inflation of ontological priority of structure precludes any notion of intrinsic nature of the relata. A structural realist theory is only viable if one adopts a deflationary view of the self.

Berkeley's Pragmatist Theory of Causation in De motu

Takaharu Oda

Berkeley's Pragmatist Theory of Causation in *De motu*

[500-word abstract]

I will argue that George Berkeley has a pragmatist theory of mechanical causation, in the scope of his eighteenth-century metaphysics of science in *De motu* (1721/1752; hereafter, *DM* section number). By pragmatism, I will mean that Berkeley takes the formulation from mechanical causes or terms to be something indispensable that should be analysed or defined from a deliberative viewpoint of human temporal needs or practices. There are epistemic limitations bounded in physics (*DM* §§41–42), within which we human agents are not atemporal as finite minds/spirits when we deliberate on mechanical causation. Then, mechanical causation on Berkeley's scientific account is assumed to be what is at work, or in use, from our perspective as *deliberators* that distinguish cause and effect. That is, causal laws are required in our linguistic, pragmatic *deliberation*. Regarding Berkeley's discourse on natural philosophy, I argue for a pragmatist theory of causation, according to which finite minds are able to deliberate on, or infer, a series of empirical evidence in causal terms towards our actual explanation. Secondly, for Berkeley, our actual explanation or definition of mechanical causation can reliably make us believe the truth within scientific discourse. This is because, hypothetically, Inference to the Best Explanation is taken inductively (or abductively) from a limited (not infinite) set of evidence to the approximate truth.

Before defending my pragmatist reading, I critically review three major interpretations of the theory of causation in *DM*. I start examining (1) reductionism, according to which one can eliminatively translate theoretical notions like forces in dynamics into observation notions about motions of bodies in kinematics (Hinrichs 1950; Myhill 1957). Secondly I examine (2) instrumentalism, according to which one can empirically hold the utility of dynamics for calculating bodily motions, even if causal terms are fictitious or the theories from them are potentially false (Popper 1953; Buchdahl 1969; Newton-Smith 1985; Downing 2005, *et al.*). Thirdly I examine (3) structural realism, according to which one can dismiss theoretical entities such as occult qualities, but not the theoretical structure of them for scientific progress (Stoneham and Cei 2009). Especially, pragmatism differs from instrumentalism because, on the instrumentalist reading, causal talk of theoretical terms like forces is not necessarily true, but can be merely fictitious, for their utility in mechanics. Clarifying why we cannot totally favour any of the three, finally I vindicate my philosophical rationale for Berkeley's pragmatism about mechanical causation in *DM*. On my reading, there are three generic ingredients (definitions) in a pragmatist theory of causation, as follows:

1. Causal terms are indispensable in scientific deliberation; they cannot be eliminated (*contra* reductionism)
2. Causal laws (theories in causal terms) are genuinely true, not fictitious (*contra* instrumentalism)
3. What a cause is is defined by our temporal deliberative practices, independent of atemporal structure that theories hold (*contra* structuralism)

Following these definitions, I object to the other three construals in contradistinction to my vindication for Berkeley's pragmatism about mechanical causation. Thereby we will newly stand to understand Berkeley's pragmatic method or his pragmatist theory of causation in scientific discourse.

[100-word abstract]

I will vindicate George Berkeley's pragmatism about mechanical causation in his *De motu* (1721). I take three generic ingredients in a pragmatist theory of causation:

1. Causal terms are indispensable in scientific deliberation; they cannot be eliminated (*contra* reductionism)
2. Causal laws (theories in causal terms) are genuinely true, not fictitious (*contra* instrumentalism)
3. What a cause is is defined by our temporal deliberative practices, independent of atemporal structure that theories hold (*contra* structuralism)

Following these definitions, I object to the other three construals in contradistinction to pragmatism. Thereby we will newly stand to understand Berkeley's pragmatist theory of causation in scientific discourse.

Biological organismality and individuality in the Extended Evolutionary Synthesis

Iñigo Ongay de Felipe

B. Philosophy of Natural Science
Biological organismality and individuality in the Extended Evolutionary Synthesis.

Keywords: evolutionary biology, individuality, extended evolutionary synthesis, multi-level selection theory.

Extended abstract

There is increasing debate in the domain of evolutionary biology with regards to the validity of a number of the core tenets of the Modern Evolutionary Synthesis (MES). While many of the current lines of research point out to an extension of the scientific purview of the MES to include an array of evolutionary factors that the original architects of the Synthesis did not envision, the framework of what has been called the Extended Evolutionary Synthesis (EES) does not constitute so far a clearly designed and thoroughly articulated scientific theory but rather designates an eclectic research program encompassing a variety of only partially overlapping areas of concern alongside a multiplicity of new conceptual tools that conjointly point out to a more complex understanding of evolution than the MES would ever allow for. This paper deals with the role of the notion of individual organismality within such new scheme of things in relation to evolutionary theory. First, the author presents what the role of the individual organism was in the light of some of the central tenets of the MES. In this respect, I argue that the concept of organismal individuality receives a reductive treatment in the MES in so long as it tends to be explained away by other more fundamental factors accounting for evolutionary change. In this regard I will consider both the conception of evolution displayed by Dobzhansky´s population genetics as well as the gene-centered view of evolution advanced by R. Dawkins from the decade of the 1970s onwards. The point will be made that in both accounts biological individuals are irremediably seen as passive *reactors* to evolutionary forces and thus an epiphenomenal notion ready to be reductively screened off by other evolutionary forces. Secondly, I will show that a multiplicity of the most recent developments regarding the EES, from niche construction and ecological inheritance to phenotypic and developmental plasticity or genetic assimilation set the stage for a revision of the role of the individual in which organismal agency is given an active role to play in bringing about evolutionary dynamics. However, such a more prominent role of individual organisms as factors shaping the evolutionary process goes hand in hand with an equally new conception of what a biological individual is. The point will be argued that the EES leads to a revision of the traditional temporary and spatial boundaries for individuality in two contrasting directions: on the one hand, in domains such as those of immunology or physiology traditional organisms start being seen as communities of symbiotic interactions in light of an holobiontic view of life. On the other hand, conversely, groups of organisms are presented as individuals of sorts in new approaches to systematics or within the Multi-Level Selection Theory. All in all, these two moves represent a twofold extension- upwards and downwards- of the way organismal individuality has traditionally been conceived of. I will conclude by showing how such a more pluralistic redefinition of individuality relevantly connects with a set of broader epistemic points concerning scientific pluralism in general as well as with the ontological interpretation of the nature of the living world as one composed of processes rather than substances that John Dupré and Daniel J. Nicholson have recently proposed.

Short abstract

This contributed paper addresses two issues regarding the role of individuality in evolution. The author first considers the rather passive role that individual organisms play in the account of evolution designed by the architects of the Modern Evolutionary Synthesis and the later gene-centered view of evolution. Secondly, the paper takes a look at the way this situation is recently changing within the so called Extended Evolutionary Synthesis toward a more prominently active view of organismal agency as a factor explaining evolutionary change. Finally, the paper revises two ways in which the traditional notion of individuality is becoming blurrier by an extension of its frontiers both upwards and downwards. The author will conclude by connecting this more pluralistic image of individuality in biology with a set of more general epistemic and ontological concerns regarding scientific pluralism and the metaphysics of a world of processes rather than substances.

The (causal) structure of epistemic environments

Barbara Osimani

Section: Formal Philosophy of Science

Title: The (causal) structure of epistemic environments

Keywords: Formal Epistemology, Variety of Evidence Thesis, Coherence, Reliability.

According to the Variety of Evidence Thesis (VET), items of evidence from independent lines of investigation are more confirmatory, ceteris paribus, than e.g. replications of analogous studies. Although intuitively plausible, this thesis is known to fail (Bovens and Hartmann 2003, Claveau 2013, Osimani and Landes 2020). We investigated the epistemic dynamics of VET failure by changing the model parameters regarding source "reliability". The comparison of our results with previous attempts to analyse the VET illustrates how distinctive ways in which (un)reliability may be modelled (and thought of), impact on the inferential import of consistent results (see also Wheeler 2009, Wheeler and Scheines 2013).

The aim of the talk is to explore the role of the modelling assumptions such results rely on. In all models presented so far the confirmatory boost of coherent evidence through (in)dependent evidence and instrument reliability are "incarnated" by a specific topological structure relating the evidence reports to the hypothesis. The structure itself and how it relates to scientific uncertainty has not been justified. Uncertainty regarding the reliability of the data generating process could invest not only the measuring instrument, but precisely the (causal) structure relating hypotheses and evidential reports more generally. Hence, uncertainty would not regard whether the measuring instrument is "reliable" or not, but whether we are dealing with a specific (causal) structure or another one.

The results that hold for the canonical case of a common cause model in which conditional independence is satisfied may not be a good approximation of what one would find if the conditional independence condition was "almost satisfied" — that is, where there is some small epsilon of association among the reports that is left after conditioning on the common cause. More generally, items of evidence may be related to one another and to the hypotheses in a number of different ways, and it is this structure that contributes essentially to whether coherence is confirmatory boosting or not. Wheeler and Scheines 2013 address this question. They study and compare scenarios where items of evidence for criminal settings display diverse conditional dependence relations, among themselves and with respect to the hypothesis; e.g. cases where some of the items of evidence are parent nodes of the hypothesis of interest itself (for instance antecedents for motives). In their framework, VET is violated when items of evidence that are incoherent among themselves confirm the hypothesis more strongly than a coherent "equally positive evidence set" (Wheeler and Scheines 2013, Proposition 6).

Their results suggest that a purely numeric analysis must be aided by explicit structural independence assumptions of the model, along with an accounting of the effect these structural assumptions have on the robustness of the model. This sort of consideration leaks also from the different role that the ρ parameter plays in Bovens and Hartmann's model and in ours: whereas in the former, this plays a prominent role, in our case it all but disappears in the results. This is exactly due to the different causal structure between evidence and hypothesis assumed in their model for the randomising instrument: as soon as the instrument is a randomiser (*REL* = *Rel*), the connection between evidence and hypothesis is severed. This does not happen for our unreliable instruments. As a consequence, in the hypothesis that the evidence is coming from a randomising instrument, the arrow from *CON* to *REP* is de facto deleted.

Hence, uncertainty as to whether the instrument we are dealing with is reliable, biased, or a randomiser may be modelled exactly via various topological structures for the epistemic dynamics representing scientific settings and learning scenarios more generally. The analysis of such structural relations between items of evidence, the investigated hypothesis, and hypotheses regarding the evidence source itself promises to be an essential step forward in understanding (scientific) reasoning and in shaping scientific methodology (see also Sprenger and Hartmann 2018).

Inference and the structure of mental representations

Matias Osta-Vélez

Inference and the structure of mental representations

There is a long-standing tradition in philosophy and cognitive science that sees inference as a syntactically-driven process (e.g. Stich ,1983; Braine and O'Brien, 1998; Fodor, 1998). According to it, thoughts are sentence-like structures with specific syntactic properties (logical forms); and inferences are transitions between thoughts motivated only by the formal relations between their truth-functional structure (e.g. Fodor 1985). Reasoning must be, then, a computational/formal process depending on some finite set of topic-neutral rules of inference like those of proof-theory.

In recent decades this view began to lose strength. On the one hand, cognitive scientists started to doubt the psychological plausibility of it because of some important findings of empirical psychology: (1) reasoning is highly sensitive to content (e.g. Pollard and Evans 1980); and (2) people are quite bad when reasoning following abstract rules (see Johnson-Laird *et al*, 1973). On the other hand, philosophers also started to challenge the aforementioned view. For instance, Gilbert Harman (2002) influentially argued that logic has almost nothing to do with reasoning, and that their association comes from a categorical mistake of identifying *inference* with *implication*.

There is an alternative way of understanding inference that challenges the formalist thesis and propose to understand it as semantically-driven. This view emerged during the last decades, both in philosophy and in cognitive science, from ideas of Sellars (1953), Piaget (Byrnes, 1992) and Sperber and Wilson (1995). According to it, rules of inference are mainly conceptual (or "material") rules: they are based on the content of the concepts involved in predicates and not only on the meaning of the logical constants. In this talk, I will push this line of thought further by proposing a general definition of inference as a cognitive mechanism that works by exploiting structural properties of underlying representational structures.

I will explain this approach by focusing on a new model of concept-based inference that uses conceptual spaces as the underlying representational system (Osta-Vélez & Gärdenfors, forthcoming). I will then show how this general definition can work as a common framework for understanding a wide variety of inferential processes that have been studied in psychology, artificial intelligence and philosophy (notably, perceptual inference, logical inference, and model-based inference).

Finally, I will discuss how this approach relates to the structural view of mental representation (Shagrir 2012; Cummins 2010), and how it provides a solution to the discussion between the formal and the semantic view of inference and reasoning.

References

Braine, M. D. S. & O'Brien, D. P., (eds.) (1998) Mental logic. Erlbaum

Byrnes, J. P. (1992). Meaningful Logic: Developmental Perspectives. In P. B. Pufall & H. Beilin (Eds.), Piaget's Theory: Prospects and Possibilities (pp. 126–140).

Cummins, R. (2010). The World in the Head. Oxford University Press.

Fodor, J. A. (1985). Fodor's Guide to Mental Representation: The Intelligent Auntie"s Vade-Mecum. Mind, 94(373), 76–100.

Fodor, J. A. (1998). Concepts. Oxford University Press. Piaget's Theory: Prospects and Possibilities (pp. 126–140)

Harman, G. (2002). Internal Critique: A Logic is not a Theory of Reasoning and a Theory of Reasoning is not a Logic. In D. M. Gabbay _et al_(Eds.), Handbook of the Logic of Argument and Inference, (Vol. 1, pp. 171–186).

Johnson-Laird, P. N. (2010). Against logical form. Psychologica Belgica, 50(3-4), 193–221.

Johnson-Laird, P. N. *et al* (1972). Reasoning and a Sense of Reality. British Journal of Psychology, 63(3), 395–400.

Oaksford, M., & Chater, N. (1991). Against Logicist Cognitive Science. Mind & Language, 6(1), 1–38.

Osta-Vélez, M. & Gärdenfors, P. (forthcoming). Category-based induction in conceptual spaces. Manuscript.

Pollard, P., & Evans, J. S. B. T. (1980). The Influence of Logic on Conditional Reasoning Performance. Quarterly Journal of Experimental Psychology, 32(4), 605–624.

Sellars, W. (1953). Inference and Meaning. Mind, 62, 313–338.

Shagrir, O. (2012). Structural Representations and the Brain. The British Journal for the Philosophy of Science 63(3), 519-545.

Sperber, D. & Wilson, D. (1995). Relevance. Basil Blackwell.

Stich, S. (1983). From folk psychology to cognitive science. Cambridge, MA. : MIT Press.

Use and abuse of Weber's methodology by economics

Aleksander Ostapiuk

Philosophy of Social Sciences

Title: Use and abuse of Weber's methodology by economics

Key words: philosophy of economics, homo economicus , Max Weber, ideal type, value-freeness

1. Extended abstract

For many years mainstream economics has been criticized from psychology (behavioral economics), sociology (institutional economics) and philosophy (philosophy of economics). However, no better paradigm has been proposed yet. The goal of this research is not to provide one but to analyze the problems with economics by using a Weberian lens. It is a framework that can shed new light on the limits of mainstream economics because two of Weber's methodological claims are its foundations: 1. Ideal type, 2. Value-free science. The goal is to check whether economists misunderstood Weber and if so how economics can be improved by incorporating Weber's insights.

The best example how economists perceive ideal type and value-freeness is Becker's economic approach where neoclassical economics' assumptions are pushed to the extreme (homo economicus, revealed preference theory). Firstly, he argues that people maximize utility both by egoistic and altruistic behaviors. Secondly, he perceives rationality in instrumental sense where the process of achieving goals is important, not goals themselves (e.g. rational drug addict). Thanks to axiomatic assumptions of rationality and agnosticism towards human motivations some economists still perceive themselves as value-free engineers.

Weber's methodology

Weber is analyzed because many economists perceived him as a father of value-free economics. Moreover, ideal type suits economists who use unrealistic models that are designed to predict and are perceived as value-free tools.

Firstly, Weber did not think that scientists can be completely value-free. He makes a distinction between methodological and instrumental value judgments. In science, we should rely on "instrumental value judgments" (e.g., empirical data or logic). However, in the end we cannot escape from using "methodological value judgments" (e.g. revealed preference theory).

Secondly, Weber perceives ideal type as methodological construction which is based on some fictional assumptions about people or reality. He underlines that ideal type is an instrument, not a description of reality. Moreover, he cautioned against perceiving ideal type as a paragon (e.g. ideal type of Nazism). Ideal type should be used as a benchmark (comparing reality with ideal type). In general, Weber perceives ideal type as a heuristic device that should predict and explain.

Ideal types in economics

Economists believe in their ideal types. In economic approach economists should not decide whether people are egoistic or altruistic. However, economists are only people and they have opinions on human nature and the world. Kuhn argues that theories are lenses by which scientists look at reality. The conception of performativity shows that ideal types used by economists are not value-free tools. Models not only describe but they also shape the world. Moreover, "The limits of my language mean the limits of my world" - when we perceive children as 'consumer good' we think about costs and benefits of having them not about love. Economists also "mistaken beauty for truth" and they believe in elegant mathematical models.

Immunity from criticism. By perceiving rationality in instrumental sense and utility *ad libitum* economic models look like Platonist models. The ideal type of homo economicus cannot be used as a

benchmark because rationality and utility are tautologies. By definition people always maximize utility and are rational. The previous conception of homo economicus with perfect rationality was different (Knight). We could compare whether people are rational or not. It is no longer a case because in economic approach even myopic addict is rational. Neoclassical economics absorbs every descriptive criticism. Economists can argue that people give back a founded wallet to the police because their preference is to be moral and giving wallet back maximize their utility. The immunity from descriptive criticism leads economists to hubris and reluctance towards methodological pluralism.

Instrumentalism. Economics focuses on prediction. However, it is not enough for theory only to predict (e.g. Ptolemaic theory). It should also explain. In science different tools are used for different goals (hammer vs barometer).

Homo economicus - prescriptive model. In the last years economists have incorporated knowledge from psychology, sociology, etc. (reverse imperialism) but homo economicus is still in the center. However, now it is perceived as prescriptive, not descriptive model. Behavioral economists look for purified preferences and they argue that this how people should behave.

Problems

The ideal type of homo economicus seems as value-free framework but it has an influence on reality.

Firstly, understanding utility *ad libitum,* where is no distinction between values, does not mean that economists are value-free but it leads to a particular perception of human nature - psychological egoism. Becker gives the example of a wife who behaves 'altruistically' and lets her husband read before sleep because it maximizes HER utility. Economics' moral relativism is often used to justify selfish behavior ("greed is good").

Secondly, economists treat welfare with revealed preferences as descriptive theory. It leads to many problems (e.g., Pareto Optimum and inequality, fetishization of GDP, consequentialist cost-benefit analysis).

Thirdly, economics' theory supports free market (more options=better) and negative freedom (people have autonomous preferences). Therefore, state shouldn't interfere. Moreover, economists perceive capitalism as ideal where invisible hand always works.

Weber's advice

If economists cannot escape from 'methodological value judgments' they should put them on the table. Economists should realize the normativity of their assumptions which are treated as positive (e.g., welfare, instrumental rationality). It can be done thanks to genealogy of economics. It is essential to understand the role of social factors because values and interests of economists are hidden in disguise of objectivity.

Economists should perceive ideal types only as instruments, not coverings laws that claim universal applicability. Therefore, economists should build models for specific purposes. Moreover, prediction is important but understanding people's motivations and values is also important (*verstehen*). Secondly, models are not value-free tools. They are like stories and sometimes we can forget that it is only a story. The ideal type of homo economicus started to live its own life. Although it was designed as a value-free concept, most social scientists (also economists) see it as a fully egoistic calculator which chooses the best option. Economists to protect themselves from confusing metaphor with reality need to scrutinize their ideal types.

2. Short abstract

For many years mainstream economics has been criticized from psychology (behavioral economics), sociology and philosophy. However, no better paradigm has been proposed yet. The goal of this research is not to provide one but to analyze the problems with economics by using a Weberian lens. It is a framework that can shed new light on the limits of mainstream economics because two of Weber's methodological claims are its foundations: 1. Ideal type, 2. Value-free science. The goal is to check whether economists misunderstood Weber and if so how economics can be improved by incorporating Weber's insights.

Is Confirmation Inductive?

Alik Pelman

EENPS 2020

Section: general philosophy of science; formal philosophy of science

Keywords: confirmation; induction.

Title: Is Confirmation Inductive?

Abstract (100 words)

The common way of putting a hypothesis H to an empirical test is by drawing empirical predictions E from H and observing whether E is true. If it is indeed true, H is *inductively* confirmed by the evidence. So presumably H is a conclusion of a strong inductive argument involving *E*. However, the argument we use for the above confirmation is, 1. If H then E; 2. E; 3. Therefore, H. But this is *not* an inductively strong argument; in fact, it is the *fallacy of affirming the consequent*. I propose a solution, by showing that *within the context of scientific inquiry*, this argument is indeed inductively strong.

Extended abstract (1000 words)

Overview

I wish to share a surprising challenge to the common assumptions that hypotheses are inductively confirmed by their empirical predictions. I shall then propose a solution to this challenge.

Introduction

The widely accepted view is that although empirical evidence cannot *verify* universal hypotheses, it can still *confirm* such hypotheses, and that such confirmation is given by induction. In short, it is believed that universal hypotheses can be inductively confirmed by empirical evidence. More specifically,

CI **Evidence *E* confirms hypothesis *H* iff *H* is a conclusion of a strong inductive argument in which *E* is a premise.**[1]

The problem

Let us have a look at the most common scientific method of assessing a hypothesis *H*. First one needs to draw from *H* empirical predictions *E*. Then, if *E* turn out true, then *H (& the background assumptions K*) is confirmed. If *E* turn out false, then *H (&K)* is falsified. When we put this in terms of arguments, we get:

Falsification Argument:

1. If H(&K) then E
2. Not E
3. Therefore, not H(&K) (1, 2, Modus Tolens)

[1] Two notions of strong: a. absolute (P(*H*)>0.5); b. incremental (P(*H*|*E*)>P(*H*))

<u>Confirmation Argument:</u>

1. If H(&K) then E
2. E
3. Therefore H(&K)

But, at lease as it stands, this Confirmation Argument is *not* an inductive argument at all, let alone a strong one. In fact, it is simply the *fallacy of confirming the consequent*.

Hempel calls this most intuitive confirmation principle, the Converse Entailment Condition:

CEC if *H* entails *E* then *E* confirms *H*.

There are very few who deny this principle. In fact, Bayesianism takes pride in its ability to derive this principle from the Bayesian confirmation condition which is:

BC If P(*H*|*E*) > P(*H*), then *E* confirms *H*.[2]

The problem is that if the Confirmation Argument is not an inductively strong argument – as seems to be the case - then it is not clear whether the premises indeed lend support to the conclusion, and specifically, it is not clear whether *E* indeed confirms *H*. Is there a solution to this problem?

A proposed solution – the case of universal generalisation

Let us begin with a simple case. Let,

H All Fs are Gs

E All *observed* Fs are Gs

Now, let us plug these into the Confirmation Argument (which normally, as we have seen, is the fallacy of affirming the consequent):

1. If all Fs are Gs then all observed Fs are Gs
2. All observed Fs are Gs
3. Therefore, all Fs are Gs.

Now (1) is a tautology. (2)+(3) simply form an inductive generalization, which is an inductively strong argument. Hence, the whole argument turns out an inductively strong argument.[3] We can conclude then that when the hypothesis *H* is a generalisation of the form 'all Fs are Gs' and the observation *E* is an instance of this generalisation, i.e., samples that are both F and G, then the Confirmation Argument turns into an inductively strong argument. In other words, a universal hypothesis is indeed *inductively confirmed* by its instances.

The proposed solution – the general case

What about confirmations that are not of universal hypotheses by their instances, e.g., what about the confirmation of the General Theory of Relativity (GTR) by the perihelion of Mercury? In order for this observation to *inductively* confirm GTR, we need an inductive argument with the perihelion of Mercury among its premises and GTR as its conclusion. But if we simply plug *these* into the Confirmation Argument, we get the normal fallacy:

---

[2] When we combine BC with Bayes theorem – namely P(*H*|*E*)=P(*E*|*H*)*P(*H*)/P(*E*) – we get the principle:
BC* If P(*E*|*H*) > P(*E*), then *E* confirms *H*.
Now when H entails E, P(E|H) = 1, and hence clearly P(E|H)>P(E). So CEC follows from BC*. In fact, since BC* applies to any case in which *H* raises the probability P(*E*), BC* clearly applies to a case in which *H* raises this probability to P(*E*)=1, and hence CEC is a *special case* of BC* (and hence of BC combined with Bayes theorem).
[3] Note that the strength of this argument is incremental and not absolute. (See footnote 1.)

1. If GTR then perihelion of Mercury
2. Perihelion of Mercury
3. Therefore, GTR.

So this won't work. We need something more sophisticated. Here is my proposal.

Let H be a *meta*-hypothesis about GTR,

H All empirical predictions of GTR are true,

and let E be the *meta*-observation about the empirical predictions of GTR (including the perihelion of Mercury),

E All *observed* empirical predictions of GTR are true.

Now plugging those into the Confirmation Argument, we get:

1. If all empirical predictions of GTR are true then all observed empirical predictions of GTR are true
2. All observed empirical predictions of GTR are true
3. Therefore, all empirical predictions of GTR are true

For reasons similar to the ones stated in the previous example, this also is a strong inductive argument. However, its conclusion is not GTR, but rather the meta-hypothesis *about* GTR. Yet (3) is a statement about the *empirical adequacy* of GTR. Which is a very strong (strongest?) support for a theory. Hence, the next move from (3) seems inductively strong:

4. Therefore, GTR is true.

This can be generalized into the following form of argument about any theory T:

1. If all empirical predictions of GTR are true then all observed empirical predictions of GTR are true
2. All observed empirical predictions of GTR are true
3. Therefore, all empirical predictions of GTR are true
4. Therefore, GTR is true.

Since (1) is a tautology, and (2)+(3) is an inductively strong argument, and so is (3)+(4), it follows that the whole argument (1)–(4) is inductively strong.

We have seen that the most common way of confirming a hypothesis by its empirical predictions has the logical form of the fallacy of affirming the consequent, rather than of a strong inductive argument. However, as I hope to have shown, *within the context of scientific inquiry*, this form of argument is indeed inductively strong. Consequently, such an argument can serve to justify the confirmation of a theory by its empirical predictions.

Integrating first-person data with neuroscience. A case of migraine aura research

Marek Pokropski

Title: Integrating first-person data with neuroscience. A case of migraine aura research

Section: Philosophy of Cognitive and Behavioral Sciences

Keywords: mechanistic explanation, dynamical explanation, migraine, first-person methods

Short abstract:

According to the idea of mechanistic integration of cognitive sciences, different fields of research are integrated when they provide constraints on the space of possible mechanisms and thus contribute to a complete multilevel explanation. The mechanistic model of integration is typically applied to various research fields in cognitive (neuro)science, such as network neuroscience and dynamical modelling. It is still unclear whether the mechanistic framework could be integrated with first-person methodologies. In this paper, I argue that first-person data may provide constraints on mechanistic explanations. To illustrate that it is possible, I discuss the case of migraine visual aura explanation.

Extended abstract:

It was argued recently that the mechanistic model of multilevel explanation could deliver the framework for integration of cognitive (neuro) science (e.g. Craver 2007, Miłkowski 2016). According to this idea, different fields of research are integrated when they provide constraints on the space of possible mechanisms, and thus contribute to a complete multilevel explanation. Such mechanistic explanations oscillate between different levels, from the behavior of an organism to processes at a molecular level, and link together all the processes relevant to the constitution of the target phenomenon. Mechanistic integration was typically considered in relation to various research fields in cognitive science, such as network neuroscience (Zednik 2018) or dynamical modelling (Bechtel and Abrahamsen 2010). It is still unclear whether mechanistic framework could be integrated with research fields that apply first-person methodologies. First-person methodologies are notoriously problematic in the studies of mental phenomena. Some argue that they are unreliable and should be replaced by some sort of mental states ascription (e.g. Dennett 1991), others develop introspective methodology in order to make it more reliable and complementary with empirical research (e.g. Piccinini 2010, Hurlburt and Schwitzgebel 2007). In this paper, I defend the claim that first-person data may provide constraints on mechanistic explanations and thus become integrated with other research fields. To illustrate this claim, I discuss the case of research on migraine visual aura.

Visual aura is a common symptom of a migraine attack, which manifests in visual hallucinatory experience. A significant role in understanding what kind of experience is such visual aura, as well as what is the responsible mechanism, played illustrations made by migraineurs

(Shott 2007). These graphical representations revealed that migraine auras often have a uniform structure, which typically consists of a characteristic zig-zag scintillating pattern (sometimes called a fortification figure due to resemblance to castle fortifications) followed by a scotoma. Interestingly, the migraine aura is a dynamic phenomenon, i.e. the visual pattern propagates and moves through the visual field (Lashley 1941). For example, the scintillating pattern may appear in the center of the visual field, and then it is growing and moving towards the periphery, where it disappears. Establishing the dynamic pattern of the aura was key to discover the underlying mechanism. In 1958 Milner noticed that there is a similarity between the speed of propagation of scintillating visual patterns and the velocity of the neural phenomenon called cortical spreading depression (CSD). More recent studies (e.g. Silberstein 2004) show that there is a growing body of evidence that CSD constitutes the underlying mechanism of the migraine visual aura.

In this paper, I argue that we should understand the explanation of migraine aura as an example of first-person constraints on mechanistic explanatory models. In particular, first-person data provided two types of constraints relevant to the mechanistic model – first, dynamical constraints related to the dynamical and temporal pattern of the target phenomenon. Second, constraints on the localization of a mechanism responsible for particular visual patterns (e.g. according to Hansen et al. (2013), visual pattern corresponds to the region of the occipital cortex that is involved).

References

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science*, Part A, 41(3), 321-333.

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

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown & Co.

Hansen JM, Baca SM, Vanvalkenburgh P, Charles A. (2013) Distinctive anatomical and physiological features of migraine aura revealed by 18 years of recording. *Brain* 136 (Pt 12):3589-95.

Hurlburt, R. T., and Schwitzgebel E. (2007). *Describing Inner Experience? Proponent Meets Skeptic*. Cambridge, MA: MIT Press

Lashley, K. S. (1941) Patterns of cerebral integration indicated by the scotomas of migraine. *Arch NeurPsych*. 46 (2): 331–339.

Milner, P.M. (1958) Note on a possible correspondence between the scotomas of migraine and spreading depression of Leão. *Electroencephalogr Clin Neurophysiol Suppl*; 10: 705.

Miłkowski, M. (2016). Integrating cognitive (neuro)science using mechanisms. *AVANT* VII(2), 45–67. doi: 10.26913/70202016.0112.0003

Piccinini G. (2010) “How to Improve on Heterophenomenology: The Self-Measurement Methodology of First-Person Data.” In the Journal of Consciousness Studies, 17.3-4 (2010), pp. 84-106.

Schott, G. D. (2007) Exploring the visual hallucinations of migraine aura: the tacit contribution of illustration, *Brain*, Vol. 130, Issue 6, pp. 1690–1703.

Silberstein, S.D. (2004) Migraine. *Lancet* 363(9406):381-91.

Zednik, C. (2019) Models and mechanisms in network neuroscience, *Philosophical Psychology*, 32:1, 23-51

Definable Conditionals

Eric Raidl

Section (f) or (c). [This is a slightly revised version of the accepted abstract to EENPS2020]
Keywords: Conditionals, Sufficient/Necessary Reason, Difference Making, Counterfactual Dependency, Evidential Conditional.

Definable Conditionals

Conditionals, that is natural language sentences of the form 'if A, [then] C', are notoriously difficult to analyse. A standard account has however emerged in the 70'ies, the so-called possible worlds account (Lewis, 1973b). This account even spread into the fields of linguistics and formal semantics in the work of Kratzer (1979), and into the psychology of reasoning (Over, 2009). According to this account, a conditional $A > C$ is true in the actual world (roughly) if and only if the closest A-worlds are C-worlds. However, recent reflections suggest to strengthen the defining clause by additional conditions. What these conditions are is not settled. Different approaches argue for different conditions (Crupi & Iacona, 2020; Krzyżanowska, Wenmackers, & Douven, 2013; Lewis, 1973a; Raidl, 2019; Rott, 2020; Spohn, 2015; van Rooij & Schulz, 2019). Some of these logics are not worked out yet, or they are only worked out for specific models. To get a grasp to compare them, we need to know what kind of logics they generate depending on the variation of the semantics. This article proposes a general method which generates completeness results for such strengthened conditionals.

The problem in a nutshell, is this: Imagine that you have a strengthened conditional of the form

- $\varphi \rhd \psi$ in world w iff closest φ-worlds to w are ψ-worlds and X.

Suppose additionally that X is also formulated in terms of the closeness apparatus. It thus seems that one can rephrase the conditional $\varphi \rhd \psi$ in the language for $>$, namely as $(\varphi > \psi) \wedge \chi$, where χ is the expression corresponding to the semantic condition X. The main question is this:

- Can we use known completeness results for $>$ to obtain completeness results for $\rhd$?

The answer is yes and the paper provides a general method. The idea goes as follows. First redefine $>$ in terms of $\rhd$. This backtranslation of $\varphi > \psi$ will yield a sentence α in the language with $\rhd$. If everything is well behaved, and bracketing some details, one can just use this backtranslation to translate axioms for $>$ into axioms for $\rhd$. In other words, the back-translation is a looking glass which provides a distorted picture of the logic for $>$, in terms of $\rhd$. This distorted picture is in fact a logic for $\rhd$.

The method applies to Lewis' (1973a) counterfactual dependency, to Spohn's (2015) sufficient and necessary reason, to Rott's (2020) difference making and dependency conditional, to Raidl's (2019) neutral, doxastic and metaphysical conditional (Raidl, in press), to Crupi and Iacona's (2020) evidential conditional (Raidl, Iacona, & Crupi, 2021) and even to counterpossible conditionals (Berto, French, Priest, & Ripley, 2018). And the list probably continues. The hope is indeed, that more complex conditional constructions could equally be treated, such as Spohn's (2015) supererogatory reason, or the probabilistic-causal conditional of van Rooij and Schulz (2019).

Let me provide the ingredients of the method here, concentrating on the completeness transfer.

Definition 1. Let $\mathcal{L}_{\rhd}$ and $\mathcal{L}_{>}$ be conditional languages. A *translation* is a total function $\circ\colon \mathcal{L}_{\rhd} \longrightarrow \mathcal{L}_{>}$ such that $p^{\circ} = p, (\neg\varphi)^{\circ} = \neg\varphi^{\circ}, (\varphi \circ \psi)^{\circ} = (\varphi^{\circ}\sharp\psi^{\circ})$ for $\sharp \in \{\wedge, \vee, \rightarrow\}$, and there is a formula $\alpha[p,q] \in \mathcal{L}_{>}$ with propositional variables among $\{p,q\}$ such that $(\varphi \rhd \psi)^{\circ} = \alpha[\varphi^{\circ}/p, \psi^{\circ}/q]$.

Definition 2. Let $\circ\colon \mathcal{L}_{\rhd} \longrightarrow \mathcal{L}_{>}$ be a translation, N a model class in $\mathcal{L}_{\rhd}$ with truth relation $\vDash_{\rhd}$ and M a model class in $\mathcal{L}_{>}$ with truth relation $\vDash_{>}$. $g : M \overset{\circ}{\hookleftarrow} N$ is *a co-embedding of M into N modulo* $\circ$ iff

1. $g : M \longrightarrow N$ and $g : W(\mathfrak{M}) \longrightarrow W(g(\mathfrak{M}))$ are total functions,
2. for all $\varphi \in \mathcal{L}_{\rhd}$ and all $w \in W(\mathfrak{M})$: $w \vDash^{\mathfrak{M}}_{>} \varphi^{\circ}$ iff $g(w) \vDash^{g(\mathfrak{M})}_{\rhd} \varphi$.

M co-embeds into N modulo $\circ$, $M \overset{\circ}{\hookleftarrow} N$, iff there is $g\colon M \overset{\circ}{\hookleftarrow} N$.

Definition 3. Let $\circ\colon \mathcal{L}_{\rhd} \longrightarrow \mathcal{L}_{>}$ and $\Gamma_{>}$, $\Gamma_{\rhd}$ axiom systems in $\mathcal{L}_{>}$ and $\mathcal{L}_{\rhd}$ respectively. $\Gamma_{>}$ *simulates* $\Gamma_{\rhd}$ *modulo* $\circ$, $\Gamma_{\rhd} \overset{\circ}{\propto} \Gamma_{>}$, iff for every $\varphi \in \mathcal{L}_{\rhd}$, $\Gamma_{\rhd} \vdash \varphi$ implies $\Gamma_{>} \vdash \varphi^{\circ}$.

We can then transfer a known completeness result for the defining conditional $>$ to the defined conditional $\rhd$:

Theorem 1. *Let N and M be model classes in $\mathcal{L}_{\rhd}$ and $\mathcal{L}_{>}$ respectively, and $\Gamma_{\rhd}, \Gamma_{>}$ systems in $\mathcal{L}_{\rhd}, \mathcal{L}_{>}$ respectively. Assume*

1. *$\Gamma_{>}$ is complete for M,*
2. *$\circ : \mathcal{L}_{\rhd} \longrightarrow \mathcal{L}_{>}$ and $\bullet : \mathcal{L}_{>} \longrightarrow \mathcal{L}_{\rhd}$ are translations,*
3. *$M \overset{\circ}{\hookleftarrow} N$,*
4. *$\Gamma_{>} \overset{\bullet}{\propto} \Gamma_{\rhd}$,*
5. *$\bullet$ inverts $\circ$ in $\Gamma_{\rhd}$, i.e., $\Gamma_{\rhd} \vdash \varphi^{\circ\bullet} \leftrightarrow \varphi$.*

Then $\Gamma_{\rhd}$ is complete for N.

Along similar lines, one can transfer soundness results and the correspondence theory for $>$ to obtain soundness results for $\rhd$ and the correspondence theory (Raidl, 2020).

The whole method can be depicted as a knowledge transfer. It allows to generate a logic for a defined conditional $\rhd$, based on the logic for the defining conditional $>$. The resulting logic for $\rhd$ is of course not the same as the one for $>$. It is a distorted image of the logic for $>$. The distortion comes from the converse definability of $>$ in terms of $\rhd$ – the *backtranslation* $\bullet$ of $>$ into $\rhd$. This backtranslation is the looking-glass which creates the distorted image. I apply the results to state the complete logic for some of the new conditionals mentioned at the beginning. This will be done in a minimal conditional logic for $>$ based on neighborhood selection frames. The above mentioned correspondence theory then allows to lift the results to stronger semantics. The method is semantically flexible, and the new defined conditionals can equally be developed in probabilistic threshold semantics, in ranking semantics, belief revision semantics or in standard Lewisean possible worlds semantics.

References

Berto, F., French, R., Priest, G., & Ripley, D. (2018). Williamson on counterpossibles. *JPL*, *47*(4), 693–713.

Crupi, V., & Iacona, A. (2020). The evidential conditional. *Erkenntnis*. doi: 10.1007/s10670-020-00332-2

Kratzer, A. (1979). Conditional necessity and possibility. In R. Bäuerle, U. Egli, & A. von Stechow (Eds.), *Semantics from different points of view* (pp. 117–147). Heidelberg: Springer.

Krzyżanowska, K., Wenmackers, S., & Douven, I. (2013). Inferential conditionals and evidentiality. *JoLLI*, *22*(3), 315–334.

Lewis, D. (1973a). Causation. *JoP*, *70*(17), 556–567.

Lewis, D. (1973b). *Counterfactuals.* Oxford: Blackwell.

Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*(4), 431-438.

Raidl, E. (2019). Completeness for counter-doxa conditionals – using ranking semantics. *RSL*, *12*(4), 861–891.

Raidl, E. (2020). Definable Conditionals. *Topoi*, *40*, 87–105.

Raidl, E. (in press). Strengthened conditionals. In B. Liao & Y. N. Wáng (Eds.), *Context, conflict and reasoning. logic in Asia Series.* Springer.

Raidl, E., Iacona, A., & Crupi, V. (2021). The Logic of the Evidential Conditional. *Review of Symbolic Logic*. doi: 10.1017/S1755020321000071

Rott, H. (2020). Difference-making conditionals and the Relevant Ramsey Test. *The Review of Symbolic Logic*, 1–32. doi: 10.1017/S1755020319000674

Spohn, W. (2015). Conditionals: A unifying ranking-theoretic perspective. *Philosophers' Imprint*, *15*(1), 1–30.

van Rooij, R., & Schulz, K. (2019). Conditionals, causality and conditional probability. *JoLLI*, *28*(1), 55–71.

Where does causal knowledge in macroeconomics come from?

Stevan Rakonjac

d) Philosophy of social sciences

Where does causal knowledge in macroeconomics comes from?

Keywords: Macroeconomics, causal structure, structural econometrics, VAR models, microfoundations

Extended abstract

The aim of this paper is to present different methodological approaches in empirical macroeconomics and to point out that these different approaches can be regarded as giving different answers to the question from the title, and consequently offering different methods for obtaining knowledge about causal structure of macroeconomic phenomena. We will also point out to some of the difficulties each of these approaches faces.

In 1947 Tjaling Koopmans criticized Arthur Burns and Wesley Mitchell for the attempt of "measurement without theory" in their study of business cycles (Koopmans 1947). Empirical regularities that Burns and Mitchell observed with their "atheoretical" methods were mere aggregate descriptions of macroeconomic data, but every economy is made up from individuals, so aggregate macroeconomic relathionships are just the product of behaviour of individuals. Therefore, it was obvious for Koopmans that observed aggregate regularities are the product of the simultaneous validity of a number of "structural" equations, the latter supposedly describing behaviour of individuals. An implicit intepretation seems to be that structural equations represent causal relathionships, while observed aggregate relathionships represent mere correlations arising out of simlatenous working of many causes.

The problem with the "atheoretical" approach is that without the help of the theory there is no way to get from the observed regularities to the structural equations that produce them. Theory should supply us with the form of structural equations and and with restrictions on the coefficients of these equations. If there are enough restrictions supplied, we can use econometric techniques to estimate coefficients of structural equations from empirical data (this is the "identification" problem). Measurement is in this picture just the proces of filling in the quntitative details of the theoretically given (causal) structure. This is the picture of econometric research develoeped by the Cowles Comission, whose member Koopmans was.

Cowles Comission approach came under attack from different positions. Robert Lucas (Lucas 1976) argued that existing structural macroeconometric models were inadequate for economic policy analysis because their parameters were not invariant to policy changes. One of the main reasons for this non-invariance was that these models did not account for the rational way in which agents form expectations about policy variables. We could also say that this non-invariance of "structural" macroeconometric models meant that they are missing important parts of (causal) structure of economic phenomena, and were therefore actually not structural. One way to answer Lucas' critique was to try to incorporate the missing part of the (causal) structure into macroeconomic models by building macroeconomic models in which rational agents are in general equilibrium. These macroeconomic models would then have proper "microfoundations", which is where the true structure of all economic phenomena presumably lies, in microeconomics. But microfoundations project has its own difficulties. Building macroeconomic models with proper microfoundations is usually feasible only in highly idealized models. Secondly, microfoundations are in practice usually achieved by a methodologically problematic shortcut of "representative agent" (Kirman 1992). Thirdly, practical application of models with microfoundations requires estimating parameters of "taste and technology", a task unlikely to be performed successfully (Sims 1986).

Another critique of structural macroeconometric models was that they could not be identified in practice, and that identification of models was usually achieved at the cost of their structurality (Liu 1960; Sims 1980). Sims' own approach was to give up the attempt of specifying the true "structure" of macroeconomic models *a priori* and try to learn as much as possible about it from the data alone. He proposed to build nonstructural macroeconometric models with as few *a priori* restrictions as possible. These models are called VAR models.

VAR models were soon deemed inadequate for policy analysis, one of the most important goals of econometric modeling (Cooley & LeRoy 1985; Leamer 1985). Atheoretical, unstructural VAR models are purely descriptive, but policy analysis requires modeling causal relathionships. Granger causality test are performed in VAR models, but "Granger causality" is not causality at all, it is pointed out, but a name for a particular kind of correlation between variables. Furthermore, using impulse response functions for analysing policy impacts with VAR models requries making errors from different equations uncorelated, which in turn requires introducing some causal structure to the model. Imposing *a priori* restrictions and structure to the VAR model gives us structural VAR models, or SVAR models.

Although restrictions required by SVAR models may be weaker than restrictions required by models in Cowless Comission approach, we are in a sense back at were we started, requiring untestable *a priori* assumptions to ground policy analysis in VAR models. (Hoover 2012). The question is where does this *a priori* knowledgde of causal structure of economic phenomena come from? The answer that Koopmans and Lucas seem to be giving is that it comes from microeconomic theory. But how did this knowledge came to be incorporated in this theory? Microeconomic theory is based on the principle that individuals optimizing under constraints.

This principle by itself has no empiricaly observable consequences, and requires additional hypotheses about preferences and constraints. Causal knowledge supplied by microeconomic theory is therefore likely to be substantialy dependent on these additional assumptions, which suggest which should shift our attention on the way additional assumptions are acquired.

To summarize, the problem of macroeconomics is that theory is often not substantive enough to structure empirical research successfully, and empirical research has hard time getting from observations to underlying causes without the help of *a priori* imposed structure. Lawrence Summers thinks we should give up attempts of developing econometric methodology for systematically gaining knowledge about causal structure of macroeconomic phenomena, because such attempts represent a "scientific illusion" (Summers 1991). Illusion or not, such attempts were defining methodological debates in empirical macroeconomics, and at the hart of these debates stand the question "Where does causal knowledge in macroeconomics come from?".

Short abstract

Different methodological approaches to empirical macroeconomics will be described and it will be explained that they represent different answers to the question from the title. Structural approaches require that macroeconometrical research should be explicitly founded on the (micro)economic theory in order to be able to measure the causal structure of the macroeconomic phenomena. Unstructural VAR approach suggest using econometric models to try to find out as much as possible about causal structure from the data, without prior restrictions from the theory. Problems with both are described.

References:

- Cooley, T. F., & LeRoy, S. F. (1985). Atheoretical macroeconometrics: a critique. *Journal of Monetary Economics*, 16(3), 283-308.
- Hoover, K. D. (2012). First Principles, fallibilism, and economics. *CHOPE Working paper, no. 2017-01*. Duke University, Center for the History of Political Economy (CHOPE).
- Hoover, K. D. (2012). Economic theory and causal inference. *Philosophy of economics, 13*, 89-113.

- Kirman, Alan (1992). Whom or What Does the Representative Agent Represent? *Journal of Economic Perspectives*, 6 (1):126–39.
- Koopmans, T. C. (1947). Measurement without theory. *The Review of Economics and Statistics*, *29*(3), 161-172.
- Koopmans, T. C. (1949). Koopmans on the Choice of Variables to be Studies and the Methods of Measurement: A Reply. *The Review of Economics and Statistics*, 86-91.
- Leamer, E. E. (1985). Vector autoregressions for causal inference?. In *Carnegie-rochester conference series on Public Policy* (Vol. 22, pp. 255-304). North-Holland.
- Liu, T. C. (1960). Underidentification, structural estimation, and forecasting. *Econometrica: Journal of the Econometric Society*, 855-865.
- Lucas, Robert E., Jr. (1976). Econometric Policy Evaluation: A Critique. In *Carnegie-Rochester conference series on public policy* (Vol. 1, No. 1, pp. 19-46).
- Simon, H. A. (1977 [1953]). Causal ordering and identifiability. In *Models of Discovery* (pp. 53-80). Springer, Dordrecht.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 48(1), 1-48.
- Sims, C. A. (1986). Are forecasting models usable for policy analysis?. *Quarterly Review*, (Win), 2-16.
- Sims, C. A. (1996). Macroeconomics and methodology. *Journal of Economic Perspectives*, 10(1), 105-120.
- Summers, L. H. (1991). The scientific illusion in empirical macroeconomics. *The Scandinavian Journal of Economics*, 129-148.
- Stock, J. H., & Watson, M. W. (2001). Vector autoregressions. *Journal of Economic perspectives*, 15(4), 101-115.
- Vining, R. (1949a). Koopmans on the Choice of Variables to be Studies and the Methods of Measurement. *The Review of Economics and Statistics*, 77-86.
- Vining, R. (1949b). Koopmans on the Choice of Variables to be Studies and the Methods of Measurement: A Rejoinder. *The Review of Economics and Statistics*, 91-94.

Disagreement and Progress in Philosophy and in Empirical Sciences

Işık Sarıhan

Re-submission for EENPS 2021

Note to organizers: This abstract is the same as last year's, with a few stylistic changes that don't affect the substantive content.

Section: General Philosophy of Science

Author: Işık Sarıhan

Title: Disagreement and Progress in Philosophy and in Empirical Sciences

Keywords: Philosophy of Science, Metaphilosophy, Disagreement, Philosophical Progress, Scientific Progress, Philosophical Method

Short Abstract

Philosophers who claim that there has not been enough progress in philosophy in terms of resolving big philosophical debates often make their point by contrasting philosophy with natural sciences. In response, some have argued that the state of progress in philosophy and in sciences is not much different: In both fields, we see a partial progress on finding answers to big questions. Against this view, I will argue that there has indeed been more substantive progress in sciences. The claim that there is no agreement on many central questions in sciences results from a misleading comparison, and it obscures the fact that the extremely slow pace of philosophical progress results from deep (but surmountable) methodological problems of the profession.

Long Abstract

The long-neglected issues of philosophical progress and peer disagreement in philosophy have recently started to attract more attention from philosophers (Chalmers 2015, Dietrich 2011, Stoljar 2017). The issue of philosophical progress, defined as philosophers' convergence on true answers to central philosophical questions, encompasses questions such as whether this type of progress is possible and how much has it been achieved to this day. The discussion of the latter question often involves contrasting the progress in natural sciences with progress in philosophy, and the bleak state of progress in philosophy is pointed out with the help of this contrast. In response to this, some philosophers have argued that the state of progress in philosophy and in natural sciences is not much different. For instance, Balcerak Jackson (2013) suggests that philosophy nevertheless arrives at partial and approximate solutions for problems, an aim that is shared with the sciences. Frances (2017) claims that there has been a lot of agreement on smaller questions in philosophy if not the central ones, and the situation in the sciences is similar, where there is a lot of progress but also disagreement on bigger questions. Olson (2019) similarly argues that there is less convergence than we think on the answers to scientific questions.

In response to this view, I will argue that there has indeed been much more progress in natural sciences compared to philosophy. First of all, I will point out that there is more substantial agreement among the practitioners of natural sciences compared to the practitioners of philosophy, and the claim that there is no agreement on many central questions in sciences results from a misleading comparison between the two fields in the context of the "centrality" of the questions asked. I will also try to demonstrate that while scientific disagreements result from the lack of clear evidence (and that agreement is reached relatively quickly when evidence is obtained), philosophical disagreements result from a deeper problem regarding a lack of explicit agreement about what counts as clear philosophical evidence and what are the steps to resolve a disagreement based on philosophical data. This methodological defect, coupled with the fact that the falsity of a philosophical view cannot be straightforwardly demonstrated, enables philosophers to easily resist various theories no matter how detailed and appealing are the arguments in defense of a rival theory. The fertile ground for philosophical disagreement, created by the above factors that relate to the method and the subject matter of philosophy, is further reinforced by psychological and professional factors such as a philosopher's not giving up on a pet theory, or the constant search for avenues of disagreement in response to the pressure to come up with novel material for publication – which is an easy task given that in many areas of philosophy there is little or no immediate practical consequence of a theory's being wrong, unlike the theories in empirical sciences.

The overall aim of the above argument is not to arrive at a pessimistic picture of philosophy where there has not been any progress at all or there won't likely be much more progress. On the contrary, I will claim that a proper understanding of the predicament of philosophical progress in contrast to natural sciences can help us fix the shortcomings of philosophical methodology that prevents philosophy from progressing in a more robust manner. I will conclude by speculating on how academic philosophical practices can be reformed to bolster philosophical progress, and what, if any, can be learned from the example of progress in empirical sciences.

References

Balcerak Jackson, M. (2013) Conceptual analysis and epistemic progress. *Synthese* 190:3053–3074.

Chalmers, D.J. (2015) Why isn't there more progress in philosophy? *Philosophy* 90 (1):3-31.

Dietrich, E. (2011) There is no progress in philosophy. *Essays in Philosophy* 12 (2):9.

Frances, B. (2017) Extensive philosophical agreement and progress. *Metaphilosophy* 48 (1-2):47-57.

Olson, D. (2019) Epistemic progress despite systematic disagreement. *Epistemology & Philosophy of Science* 56 (2):77–94.

Stoljar, D. (2017) *Philosophical Progress: In Defence of a Reasonable Optimism*. Oxford University Press.

Why Does Symbolic Logic Emerge During the Industrial Revolution?

Michael Shenefelt and Heidi White

976 words

Why Does Symbolic Logic Emerge During the Industrial Revolution?

Extended Abstract:

Before the nineteenth century, farsighted thinkers had long toyed with the idea of a fully symbolic logic, but they had never turned any such project into reality. Only with the advent of large-scale manufacturing did symbolic logic finally take shape. The first fully symbolic systems were laid out by the English logicians George Boole and Augustus De Morgan, both of whom published major books in 1847—just as the Industrial Revolution in England was in full swing. Is the correlation between symbolic logic and the Industrial Revolution merely a coincidence—or is it cause and effect? We argue that a key factor behind symbolic logic's growth was the Industrial Revolution itself. Specifically, the Industrial Revolution convinced large numbers of logicians and mathematicians of the immense power of mechanical operations—and thereby supplied an audience for a new approach.

In fact, the correlation between industrialization on the one hand and the development of abstract algebras and symbolic logic on the other is close. Boole, De Morgan, and George Peacock (author of the influential *Treatise on Algebra*) all came from England during a period of intense industrialization, and, later in the century, the eminent figures of Gottlob Frege, Georg Cantor, and Richard Dedekind appeared in Germany—just as Germany, too, industrialized. Giuseppe Peano perfected his axioms at the University of Turin at about the same time that the Automobile Factory of Turin (whose acronym in Italian is "FIAT") built its first automobiles.

Is there a causal connection between industrialization and the development of symbolic logic—or is there perhaps some *other* explanation for why the rise of symbolic logic and advent of industrialization would appear to go hand in hand?

In this paper, we'd like to suggest that a key factor behind symbolic logic's growth was the Industrial Revolution itself. Specifically, the Industrial Revolution convinced large numbers of people, in all walks of life, of the immense power of mechanical operations. Whole generations witnessed this power, and out of these generations the logicians of the age emerged. Of course, logical discoveries depend on individual insight, and speaking for ourselves, we have no particular theory of how such insights occur. But logic as a *discipline* requires something more—insight with an audience. Logicians need *other* people who are willing to listen, and audiences are a consequence of social forces, forces that affect large numbers of people quite apart from individual will. As a result, one might expect logic to have a social history no less than an abstract one. Logic considers unchanging, abstract truths (or at least they *seem* to be unchanging, abstract truths; this question is of course philosophically complicated); nevertheless, the extent to which large numbers of people will ever really explore such truths still depends, in part, on their social setting. And their social setting turns on various factors—political, economic, technological, even geographical. On this view of things, one might expect the history of logic to be a mix of the abstract *and* the mundane.

We'd like to argue that the development of symbolic logic was encouraged by the nineteenth-century success of large-scale manufacturing.

The seeds of symbolic logic certainly go back farther than the nineteenth century; they are already plainly visible in the seventeenth century in the work of the philosopher Leibniz—who aimed at a logical calculus that he said would be "mechanical." In particular, Leibniz was interested in geared machinery that could be used to solve mathematical problems, and he was especially drawn to the idea of an artificial language that would allow people to express all observable facts unambiguously and to make from them all valid deductions. All the same, only in the nineteenth century did this sort of project finally become real, and the nineteenth century invested far more energy in the effort than did any earlier period.

But why should developments in manufacturing have had any effect on the specialized thinking of nineteenth-century logicians?

An old jibe against logicians of the ancient world was that they acted like men eating crabs—dismantling the shell with a great deal of labor just so they could eat a tiny morsel of meat (the morsel being an argument whose logical validity was probably already obvious). This was a complaint against classical Greek logic (from Ariston of Chios, by the way), but with symbolic logic the logician is plainly dismantling a great deal more. And this is part of what sets nineteenth-century logicians apart from their predecessors. They were willing to submit to the apparent tedium of a far more elaborate approach. If one approaches argumentation in an admittedly tedious but mechanical way, making explicit everything that we usually grasp intuitively, without analysis, one can construct a system that captures much larger areas of valid argumentation, and one can bring out its apparent form.

In some ways, the process might be likened to the building of a machine—for example, an immense, steam-driven loom. Building such a loom would be laborious, and putting it together would require forging some odd-looking components. If, having finally built the loom, one were then to use it to manufacture only one piece of cloth, the whole exercise would be pointless. Yet if one brought the loom up to full speed and set it running for prolonged periods, one would then have a machine of great power.

Just so, building an abstract system of symbolic logic requires tedious labor and odd-looking components—all the components, for example, that one learns about in mastering first-order predicate logic. But if designed correctly, the system can then express a great array of theorems, and the proposed proof of a theorem can be checked mechanically.

Much depends, of course, on just what one means by the words "mechanical" and "mechanically," and these are words on which nineteenth-century logicians often relied.

Tactile field and the dual nature of touch

Błażej Skrzypulec

c) Philosophy of Cognitive and Behavioral Sciences

Tactile field and the dual nature of touch

Keywords: tactile perception; interoception; exteroception; spatial perception

It seems that there is an important difference concerning the way in which space is presented in visual and tactile modalities. In case of vision, visual objects are experienced as located in a visual field. On the other hand, it is controversial whether similar field-like characteristics can be attributed to space in which tactile objects and tactile sensations are experienced to be located. In the presentation, I investigate whether postulating the presence of a tactile field is justified. I argue that the answer is both 'yes' and 'no'. This is so due to the dual nature of touch: touch is both an interoceptive modality, which presents states of one's body, and an exteroceptive modality, which presents external entities. I claim that the interoceptive tactile space has the character of a spatial field. On the other hand, the exteroceptive tactile space does not have characteristics necessary for ascribing field status.

The main intuition concerning the presence of a visual field is that visual objects are not presented merely as standing in spatial relations, but are presented as positioned within a topologically connected space. This space seems to constitute a 'container' or 'form' in which perceived objects are located, and which remains structurally unchanged despite the appearances and disappearances of objects (Richardson 2010). Relying on the above intuition, I propose that a perceptually presented space is a spatial field if and only if (a) it is topologically connected, (b) its structure is independent from the pattern of spatial relations between presented entities, and (c) it can contain empty locations.

Interoceptive tactile space

In arguing that interoceptive tactile space is a field, I focus on two bodily representations: (a) skin-space which represents body relying on layout of cutaneous receptors (Cheng and Haggard 2018) and (b) stable, off-line body schema which represents body as a structure made of parts connected by joints (de Vignemont 2010). The feature of topological connectedness is explicitly present in case of skin-space as it represents body relying on the pattern of receptors covering the skin (Longo and Golubova 2017). Furthermore, also the stable body models body as a structure of topologically connected parts (de Vignemont et al. 2009). Also, the second characteristic of spatial fields is supported by both of the considered representations. The topological structure of the skin-space is determined by the positions of skin receptors and so stays the same no matter which receptors receive stimulation at a given moment. Similarly, the structure of stable bodily schema is determined by the way in which fragments of body are connected by joints and is not influenced by the current patter of tactile stimulation. The situation is no different when the final characteristic of spatial field is considered: the possibility of empty locations. In particular, the structure of skin-space is determined by the array of receptors no matter their current activities and so skin-space can model the body as a space containing places in which no tactile stimulation is present. Analogously, in case of stable body schema the structure of body parts is represented no matter whether there is some tactile stimulation affecting these parts.

Exteroceptive tactile space

While that there are strong reasons for characterizing interoceptive tactile space as a spatial field, the same is not true about exteroceptive tactile space. Exteroceptive space is not the bodily space in which tactile sensations are experienced but is a space in which external tactile items are presented to be located.

The external tactile objects are represented in virtue of exploratory procedures which, by using cutaneous and kinesthetic information, allow representing external objects as instantiating a variety of tactile properties (Fulkerson 2011). Nevertheless, it is not justified to claim that such objects are presented as positioned in a tactile spatial field. Due to obtaining cutaneous data, and by utilizing the fact that properties of tactile bodily sensations can reflect the properties of external objects (Richardson 2011), tactile perception may represent the item touching the skin as a common subject of various spatial properties such as shape and size. However, while data gathered by skin receptors allow recognizing the arrangement of surfaces influencing the body, they are salient in regard of how the space extending beyond the bodily boundaries is organized. In case of interoceptive tactile space, its topological structure was in an important respect determined by the array of skin-receptors. Nevertheless, when applied to exteroceptive space, such information is not sufficient for determining the topological structure of space as this structure is not bodily structure but the structure of space outside the body.

Of course, in usual situations the exteroceptive tactile perception utilizes not only static cutaneous data but also relies on kinesthetic data and engages in dynamic exploratory procedures. Because of that one may believe that in virtue of these complex, dynamic operations exteroceptive tactile space is presented as a spatial field. For instance, by processing both cutaneous and kinesthetic information, one may be presented not only with some surfaces inflicting pressure on the skin but with an object extending in external space from one body part to another. In consequence, it seems that such experiences present a topologically connected, external spatial path between bodily parts. However, in such tactile experiences the topological structure of space is not independent from the experienced patters of spatial relations between presented entities. It is so because in such exteroceptive experiences presentation of a topological connectedness of a fragment of space depends on presenting particular, synchronic and diachronic, relations between tactile objects and body parts. For instance, it is not the case that connecting and disconnecting fingers are experienced as positioned in topologically connected space that is presented independently from their movements. Quite contrary, a fragment of external space is presented as topologically connected in virtue of dynamic pattern of relations between fingers and would not be presented as such if relations between fingers were different.

Cheng, T., & Haggard, P. (2018). The recurrent model of bodily spatial phenomenology. Journal of Consciousness Studies, 25(3–4), 55–70.

Fulkerson, M. (2011). The unity of haptic touch. Philosophical Psychology, 24(4), 493–516.

Longo, M. R., & Golubova, O. (2017). Mapping the internal geometry of tactile space. Journal of Experimental Psychology: Human Perception and Performance, 43(10) 1815–1827.

Richardson, L. (2010). Seeing empty space. European Journal of Philosophy, 18(2), 227–243.

Richardson, L. (2011). Bodily sensation and tactile perception. Philosophy and Phenomenological Research, 86(1), 134–154.

de Vignemont, F. (2010). Body schema and body image—Pros and cons. Neuropsychologia, 48, 669–680.

de Vignemont, F., Majid, A., Jola, C. & Haggard, P. (2009). Segmenting the body into parts: Evidence from biases in tactile perception. The Quarterly Journal of Experimental Psychology, 62(3), 500–512.

Causal Attribution and Partial Liability: A Probabilistic Model

Jan Sprenger

Causal Attribution and Partial Liability: A Probabilistic Model

March 26, 2021

The Problem

Probabilistic models have been successful at quantifying the strength of the causal influence of an intervention on a target event (e.g., Pearl, 2001; Fitelson and Hitchcock, 2011; Sprenger, 2018). These models can be embedded seamlessly into our currently best theories of causation expressed by directed acyclical graphs (DAGs). It is unclear, however, whether such measures of causal strength transfer to the problem of **causal attribution**—more specifically, the degree to which a certain event X_1 contributes to the occurrence of another event Y, especially when other causes $X_2, \ldots X_n$ are present.

This is not a purely philosophical problem: it is relevant for the practice of tort law, whenever the legal system allows for **partial liability** and not only for all-or-nothing liability (Hart and Honoré, 1985; Wright, 1988). There is a widespread intuition in tort law, encoded in various rulings and scholarly texts, that a defendant is liable for a claimant's loss to the degree—and only to the degree—to which the defendant's actions caused the claimant's loss (Kaiserman, 2017).

The question is, of course, how this degree should be quantified. Kaiserman (2016) proposes an account of causal contribution where the contribution of each cause is proportional to $p(Y|X_i)$ and defends this choice by some examples. However, his account is devoid of a compelling principled motivation. This contribution develops a more sophisticated probabilistic model using a recent case from Dutch tort law (*De Munnik Schoenen B.V. vs. ABAB Stichting*) as a running example.

The Case

In the running example the defendant's negligent advice A influenced the claimant's decision D that should, finally, lead to heavy financial losses L for him. The defendant was the claimant's financial advisor and was supposed to act with due professional diligence. The negligence itself was not disputed.

The court identified the defendant's liability as proportional to the "lost chance for a better result" *[verloren kans op een beter resultaat]*. How should we calculate this "lost chance"? (The court's reasoning was demonstrably fallacious and shall not be discussed.) A natural approach is that it is a function of $p(L|A)$ and $p(L|\neg A)$, i.e., that it should depend on the chance of the occurrence of the loss in two scenarios: the defendant's advice is negligent; the defendant's advice is *not* negligent.[1]

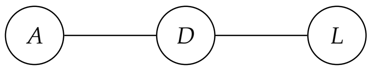

Figure 1: A causal Bayesian network with the relevant variables.

To calculate these probabilities, we need to make some causal modeling assumptions, e.g., those in Figure 1. The causal Bayes net states that the negligent advice influences the final result (loss or no loss) exclusively via the claimant's decision whether or not to follow the advice. The calculation of $p(L|A)$ and $p(L|\neg A)$ then depends on the conditional probabilities $p(L|D)$, $p(L|\neg D)$, $p(D|A)$ and $p(D|\neg A)$. Finally, it seems natural that the relevant "lost chance" should be quantified as the *chance difference*

$$\Delta = p(L|A) - p(L|\neg A).$$

This quantity corresponds to the prospective causal strength measure defended by Sprenger (2018) and the absolute risk reduction (ARR) measures in medicine (e.g., Sprenger and Stegenga, 2017).

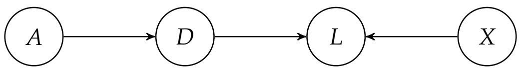

Figure 2: A causal Bayesian network with the relevant variables. X = another enabling cause for L.

An advantage of this model over the causal contribution model suggested by Kaiserman (2016) is its independence on the number of actual causes of the loss: while Kaiserman's model normalizes the influence of each cause $X_1, \ldots X_n$ by the sum $\sum_{j=1}^{n} p(Y|X_j)$ and thus depends on how specific the model is, the value of Δ is insensitive to the presence of further enabling conditions for the target events (e.g., such as in Figure 2).

An Alternative Model

The talk will discuss an alternative probabilistic model where partial liability is given by

$$\Delta' = \frac{p(L|A) - p(L|\neg A)}{p(L|A)} \tag{1}$$

[1]For purposes of exposition, I do not draw the distinction between $p(L|A)$ and $p(L|do(A))$ as long as L is causally downstream of A.

The quantity Δ' is called *relative risk reduction* in medicine. The motivation of this model is that the compensation to the claimant must be equal to the *expected damage caused by a negligent advice in the long run* (i.e., in repetitions of the same situation). Alternatively, the measure Δ' can be motivated by dividing the risk of the occurrence of the damage (i.e., $p(L|A)$) between the baseline risk (i.e., $p(L|\neg A)$) and the increase caused by the defendant's negligence (i.e., $p(L|A) - p(L|\neg A)$). The degree of partial liability then corresponds to the share of total risk that is attributable to the defendant.

The choice between both models, between Δ and Δ', is not trivial: it goes beyond the purely technical, mathematical dimension and involves matters of legal doctrine, such as whether we want to determine partial liability as a function of the *actually* lost chance (Δ), or in proportion to the expected damage of the defendant's action (Δ').

Concluding Thoughts

Elementary insights from probabilistic reasoning and causal inference still have to find their way into legal practice. Using precise causal and probabilistic models avoids fallacies in legal reasoning and quantifies partial liability in a rigorous way. However, certain methodological obstacles remain.

Similar problems also arise in the sciences, e.g., in medicine where scientists try to attribute adverse events to particular factors. Developing a successful model for causal attribution and partial liability requires mathematical schooling of legal practitioners as well as a interdisciplinary collaboration between philosophers, legal scholars and scientists.

References

Fitelson, Branden and Christopher Hitchcock (2011). Probabilistic Measures of Causal Strength. In Phyllis McKay Illari, Federica Russo, and Jon Williamson (eds.), *Causality in the Sciences*, pp. 600–627. Oxford: Oxford University Press.

Hart, H. L. A. and Tony Honoré (1985). *Causation in the Law*. Oxford: Oxford University Press.

Kaiserman, Alexander (2016). Causal Contribution. *Proceedings of the Aristotelian Society 116*, 387–394.

Kaiserman, Alexander (2017). Partial Liability. *Legal Theory 23*, 1–26.

Pearl, Judea (2001). Direct and Indirect Effects. In Jack Breese and Daphne Koller (eds.), *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 411–420.

Sprenger, Jan (2018). Foundations for a Probabilistic Theory of Causal Strength. *Philosophical Review 127*, 371–398.

Sprenger, Jan and Jacob Stegenga (2017). Three Arguments for Absolute Outcome Measures. *Philosophy of Science 84*, 840–852.

Wright, R.W. (1988). Allocating Liability among Multiple Responsible Causes: A Principled Defense of Joint and Several Liability for Actual Harm and Risk Exposure. *University of California at Davis Law Review 21*, 1141–1211.

The Origins of Observation

Athamos Stradis

The Origins of Observation

In statistical mechanics, a system E at any given moment is described by a point in its state space ('phase space'). This corresponds to a 'microstate', an exact microscopic configuration of its constituent particles. For example, if E is a glass of water, its microstate would describe the positions and momenta of every water molecule. However, each point in phase space inhabits a region corresponding to a 'macrostate', a coarse-grained characterisation that we use in daily life. In our example, this might be 'glass of cold water with a volume of one litre'.

Macrostates consist of sets of microstates that are indistinguishable. But to whom? This definition implicitly invokes a human observer O to whom some sets of microstates (the 'familiar macrostates, $\{F_i\}$), but not others ('alternative macrostates', $\{A_i\}$), look the same. This raises an obvious question:

> **Macrostate Question**: *Why do we observe* $\{F_i\}$ *rather than* $\{A_i\}$*?*

A good answer to this question should explain why two features of $\{F_i\}$ coincide. The first feature, already discussed, is that its members consist of indistinguishable microstates; hence, they are *homogenous.* The second is that they exhibit *regularities*, e.g. the Second Law.[1] As Hemmo and Shenker (2012, ch.5) point out, homogeneity and regularity are logically distinct properties of $\{F_i\}$.

Hemmo and Shenker (2012, 110) argue that the Macrostate Monitor Question has an evolutionary answer. On their view, organisms which can observe robust regularities will have a considerable survival advantage, since they will be able to make reliable predictions. However, $\{F_i\}$ *does* in fact exhibit reliable regularities, such as the Second Law.[2] Since we humans have survived

[1] See Shenker (2017) for details.

[2] Since $\{A_i\}$ can be arbitrarily gerrymandered in phase space, it's doubtful that they tend to exhibit robust regularities.

1

this far, it's therefore no wonder that $\{F_i\}$ constitutes the set of macrostates we actually observe. Hence, the fact that $\{F_i\}$ exhibits both homogeneity and regularity is explained as follows: beings for whom these features of macrostates coincide will tend to succeed.

Is this a good answer to the Macrostate Question? To assess this, we need to be clearer about what we mean in saying O can 'observe' E. On a very minimal interpretation, this could mean 'monitor':

> **Monitor**: *O can monitor $\{L_i\}$ iff some of O's states correlate with some of $\{L_i\}$.*

where L_i is just some state of E. Humans certainly monitor $\{F_i\}$: a human (O) feeling cold/hot correlates with its environment (E) being cold/hot. Reading 'observe' in this liberal way, the Evolutionary Account might address this interpretation of the Macrostate Question:

> **Macrostate Monitor Question**: *Why do we monitor $\{F_i\}$ rather than $\{A_i\}$?*

But monitoring $\{F_i\}$ is a very generic feature of physical systems shared by many inanimate objects. For example, imagine a bubble (O) in some sea foam washed up on a beach. Its volume being small/large correlates with its environment (E) being cold/hot, so it monitors $\{F_i\}$ just as we do. This is a problem for the evolutionary account: since inanimate objects like bubbles have not been subject to natural selection, there must be a *non*-evolutionary explanation for why they monitor $\{F_i\}$. And since this presumably extends to physical systems generally, it would make an evolutionary explanation for why *we* monitor $\{F_i\}$ redundant.

I propose the following non-evolutionary explanation. According to my definition, $\{L_i\}$ is monitorable iff it enters into correlations with states of other systems (such as O). This just means $\{L_i\}$ has to exhibit regularities. But we know that this is true of $\{F_i\}$, and not $\{A_i\}$ (see footnote). Hence, we monitor $\{F_i\}$ and not $\{A_i\}$ because this is possible, whereas the reverse is impossible.

Hemmo and Shenker certainly think humans *at least* monitor $\{F_i\}$. However, its obvious that we also *enlist* $\{F_i\}$ to guide our actions. For example, a human (O) feeling cold/hot correlates with its environment (E) being cold/hot, *and then* the human switching on her home's thermostat up/down to maintain an optimal body temperature. We observe things not just as passive monitors, but as 'agents':

> **Monitor**: *O can enlist* $\{L_i\}$ *iff some of O's states correlate with some of* $\{L_i\}$*, and O's states then cause actions.*

Reading 'observe' in this richer way, the Evolutionary Account might take it as a given that we merely monitor $\{F_i\}$, but aim to address a different reading of our original question:

> **Macrostate Agent Question**: *Why do we enlist* $\{F_i\}$ *rather than* $\{A_i\}$*?*

However, my earlier explanation undercuts even this. In order for us to enlist $\{F_i\}$, we need to monitor in the first place with cognitive states that then cause our actions. Hence, monitoring $\{F_i\}$ is a prerequisite for enlisting $\{F_i\}$. But as we saw, $\{F_i\}$ is monitorable whereas $\{A_i\}$ is not, so this already answers why we enlist the former rather than the latter.

In conclusion, we monitor $\{F_i\}$ rather than $\{A_i\}$ because regularity is a prerequisite for monitorability, and $\{F_i\}$ exhibits regularity whereas $\{A_i\}$ generally doesn't. Moreover, since monitorability is a prerequisite for enlistability, this also explains why we enlist $\{F_i\}$ rather than $\{A_i\}$. Because all this simply follows from what's observable in statistical mechanics *in principle*, it undercuts evolutionary explanations. I suggest evolution may explain something else instead: out of all the states which *are* monitorable/enlistable, why do we *in fact* monitor/enlist some more frequently than others?

References

Hemmo, Meir, and Orly R. Shenker. 2012. *The Road to Maxwell's Demon: Conceptual Foundations of Statistical Mechanics*. New York: Cambridge University Press

Shenker, Orly. 2017. "Foundations of statistical mechanics: mechanics by itself." *Philosophy Compass*, https://doi.org/10.1111/phc3.12465

Deep Convolutional Neural Networks and How-Possibly Explanations in Cognitive Neuroscience

Vanja Subotić

EENPS 2020, *Philosophy of Cognitive and Behavioral Sciences* Section

Deep Convolutional Neural Networks and How-Possibly Explanations in Cognitive Neuroscience

Short Abstract

Deep convolutional neural networks (DCNNs) have been proved as successful in pattern recognition tasks, especially tasks pertaining to the field of computer vision. Buckner (2018) argues that DCNNs may implement even higher cognitive processes such as transformational abstraction, thereby vindicating classical empiricist epistemology. On the other hand, Marcus (2018a, 2018b) is rather skeptical of DCNNs posing any serious threat to nativism. By drawing on Stinson (2018, 2020) and Craver (2007), I will be arguing that models based on DCNNs can be used as tools for discovering general-level mechanistic explanations of cognitive processes, and as sketches of mechanisms are being filled with details from cognitive neuroscience (as in e.g., Ritchie & Op de Beeck 2019), one can expect to jump from current how-possibly to how-plausibly explanations. The chances are that such explanations will be more akin to empiricist point of views rather than nativist.

Keywords: *cognitive neuroscience, deep convolutional neural networks*, *deep learning*, *empiricism*, *explanation*.

Extended Abstract

At least since 2010, deep convolutional neural networks (DCNNs), a subclass of neural networks employing deep learning algorithms, have been proved as successful in pattern recognition tasks, especially tasks pertaining to the field of computer vision (*cf*. Hassabis *et al*. 2017 for an overview). Nonetheless, it was only recently that the relevance of deep learning and DCNNs was recognized by philosophers of cognitive science and AI (Buckner 2018, 2019, Buckner & Garson 2018, López-Rubio 2018, 2020).

This is all the more surprising given the enthusiasm that struck philosophers when connectionism, based on (shallow) neural networks, entered the scene of cognitive science in the 1980s. The main promise of connectionism and its appeal was in the presumed vindication of empiricism. It seemed that when a neural network is trained on data to perform a certain task, the emergent knowledge of such a network is learnable from experience, instead of being hardwired as in traditional symbolic models. However, higher-cognitive processes represented an insurmountable obstacle for connectionist modeling. Early critics of connectionism, who opted for symbolic models, claimed that *en principe* without combinatorial syntax and semantics, along with either explicit or implicit encoded rules, no neural network can ever account for language processing, reasoning, or abstracting (cf. e.g., Fodor & Pylyshyn 1988, Pinker & Prince 1988).

Several decades later, the crux of the dispute between *pro* connectionist and *pro* symbolic philosophers and cognitive scientists rests pretty much the same, even though there are methodological differences between early connectionist models and state-of-the-art DCNNs. Specifically, DCNNs have three distinctive features – namely, convolution, depth, and pooling – which account for computational efficiency and neural constraining. Buckner (2018) claims that these three features of DCNNs allow them to implement even higher cognitive processes, such as transformational abstraction, which means, in turn, that DCNNs support the traditional empiricist idea of abstraction, like Locke, Berkeley, and Hume endorsed; as well as contemporary empiricist idea in cognitive science that abstract category representations can be deployed using only domain-general mechanisms. On the other hand, Marcus (2018a, 2018b) is rather skeptical about the possibility that DCNNs have a natural way to deal with the hierarchical structure which could be constrained only by encoding rules; or that they could ever pose a serious threat to nativism, which treats language and other higher cognitive processes as domain-specific.

Contra Marcus, I firstly propose to grant models based on DCNNs an exploratory role by drawing on Stinson (2018, forthcoming). Thus, it would be possible to tinker the parts of a neural network and see what they do and how do they behave – which would be a good way to explore both nativist and empiricist claims without committing to them in advance. Moreover, by granting them such a role, I am leaving the possibility of methodological progress wide open. If current models have flaws, that does not mean that they cannot ever be improved – as the history of connectionist models teaches us – DCNNs are methodologically superior to shallow neural networks from the 1980s and therefore, can account for way more nuanced tasks and even outperform humans.

The explanations that one infers from models based on DCNNs can be construed as how-possibly explanations with an aspiration towards how-plausibly explanations. Recall Craver's (2007) view that when models describe how a set of parts and activities might be organized together, they provide us with a how-possibly mechanistic explanation of the phenomenon. One might have no idea if the assumed parts exist or whether they engage in the activities ascribed to them in the model. However, as more and more details are being adduced, how-plausibly models that are consistent with the known constraints on the parts emerge (Craver 2007: 121). In a similar vein, as more and more details from cognitive neuroscience are becoming available to us, models based on DCNNs can start offering how-plausibly explanations of cognitive processes. To better grasp this proposition, consider the following study in cognitive neuroscience. Ritchie & Op de Beeck (2019) use fMRI coupled with formal modelling to evaluate categorization models based on different degrees of abstraction so that they could examine how neural representations for object categories arise. A complete how-possible explanation which moves towards how-plausibly explanation of, say, mechanisms underlying the cognitive process of abstraction would include both insights from Buckner (2018) and Ritchie & Op de Beeck (2019). As Catherine Stinson (2018, 2020) puts it, exploratory connectionist models are promising tools for discovering general-level mechanisms underlying cognitive processes, and the chances are that such explanations will be more akin to empiricist point of views rather than nativist.

Keywords: *cognitive neuroscience, deep convolutional neural networks*, *empiricism*, *explanation*.

3

References

Buckner, C. 2018. Empiricism without Magic: Transformational Abstraction in Deep Convolutional Neural Networks. *Synthese*, 195(12), 5339–5372.

Buckner, C., & Garson, J. 2018. Connectionism and Post-connectionist Models. In M. Sprevak, & M. Columbo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 76–91). Routledge University Press.

Buckner, C. 2019. Deep Learning: A Philosophical Introduction. *Philosophy Compass*: e12625.

Craver, C. F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Clarendon Press.

Fodor, J., & Pylyshyn, Z. 1988. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition* 28, 3-71.

Hassabis, D., Kumaran, D., Summerfield, C., et al. 2017. Neuroscience-inspired Artificial Intelligence. *Neuron*, 95 (2), 245–258.

López-Rubio, E. 2018. Computational Functionalism for Deep Learning Era. *Minds & Machines* 28: 667-688.

López-Rubio, E. 2020. Throwing light on black boxes: emergence of visual categories from deep learning. *Synthese*, https://doi.org/10.1007/s11229-020-02700-5

Marcus, G. 2018a. Deep Learning: A Critical Appraisal. Preprint ArXiv:1801.00631.

Marcus, G. 2018b. Innateness, AlphaZero, and Artificial Intelligence. Preprint ArXiv:1801.05667.

Pinker, S., & Prince, C. 1988. On Language and Connectionism. *Cognition,* 28, 73-193.

Ritchie, J. B., & Op de Beeck H. 2019. A Varying Role for Abstraction in Models of Category Learning Constructed from Neural Representations in Early Visual Cortex. *Journal of Cognitive Neuroscience,* 31 (1), 155-173.

Stinson, C. 2018. Explanation and Connectionist Models. In M. Sprevak, & M. Colombo (Eds.), *The Routledge Handbook of the Computational Mind* (pp. 120–134). Routledge University Press.

Stinson, C. 2020. From Implausible Artificial Neurons to Idealized Cognitive Models: Rebooting Philosophy of Artificial Intelligence. Philosophy of Science 87 (4), https://doi.org/10.1086/709730

Function Acquisition in Genomics

Predrag Šustar and Zdenka Brzović

Section b) Philosophy of Natural Science

Title: Function Acquisition in Genomics

Keywords: function, genomics, etiological theory, selected effects, causal role

Short abstract

How do new functions arise at the genomic level and at which point can we say that a certain genomic trait, more precisely, its activity or diverse activities, have become *functional*? We argue that an articulation of a set of evolutionarily-based constraints for a causal role account of functions can provide a good answer to that question. By taking into consideration leading explanatory models in current genomics, we aim at positioning our etiological account with regard to other competitors in this strand of the function debate, primarily, the "generalized selected effects" account.

Extended abstract

How do new functions arise at the genomic level and at which point can we say that a certain genomic trait, more precisely, its activity or diverse activities have become *functional*? This question is especially salient in the light of advances in medicine, such as those in the area of genomic and gene editing. The fact that we can add or remove certain genomic elements, because they perform (or fail to perform) a determined function in a living system brings us to the above basic question. Furthermore, how do genomic traits of synthetic organisms, whose entire genomes or substantial portions of them are engineered in the lab, acquire functions?

It seems that here evolutionary history is not relevant for the ascription of functions. This reflects the 'received view' in the function debate, according to which the strong

etiological, i.e., the traditional selected effects (SE) account is not applicable within the biomedical sciences (see, e.g., Germain, Ratti, and Boem 2014). In other words, the rival causal role (CR) account of functions should be endorsed instead, because it deals with actually contributing activities to the functioning of the containing biological system. The scientific and philosophical debate on the "ENCODE controversy", however, has shown that CR on its own, that is, without other constraints on function ascriptions than those concerned primarily with the working scientists' specific research interests, cannot account for a stratified programmed character of functional activities of a trait at the genomic level (see Doolittle et al. 2014; Graur et al. 2013; Brzović and Šustar 2020).

Accordingly, we argue that an articulation of a set of evolutionarily-based constraints for CR can give a good answer to our departing question and warrant a unified account of functions, which applies to various areas in genomics, from more historical ones to the areas such as medical genomics. On our account, there are two main sources for evolutionary constraints on CR: *firstly*, the constraints related to the notion of fitness. Namely, on this view, we ascribe a function to a genomic trait of an organism if and only if performing that function persists in causally contributing to the organism's and its ancestors' fitness. *Secondly*, in order to respond to the main issue of the paper, further constraints are needed that specify at which point can we consider an initially accidental contribution to the fitness in question to have a *programmed* character (Cummins 1975) and exhibit a lawlike regularity. Here the relevant constraints are related to the notion of natural selection and selection processes more broadly, but also other, selectively neutral processes that result in regular behavior. As to both points, we position our etiological selectionist account with regard to Buller's account of "weak etiology" (see in particular Buller 1998), Garson's "generalized selected effects" account of functions (see Garson 2019) and Maley and Piccinini's unified mechanistic account of teleological functions (see Maley and Piccinini 2018).

We illustrate our overall account, especially the way in which evolutionary constraints apply to CR in the areas of genome biology and medicine, by considering the following explanatory models: (1) the gene duplication and divergence models; (2) *de novo* gene origin explanatory models, and (3) simpler genome editing models. However, we mostly concentrate on (2), because it is interesting to examine how *de novo* evolved genes transition from a state without a function to a state of having a function (see Keeling et al. 2019). This makes it interesting for examining function acquisition in both evolutionary and more present-related, biomedical aspects of genomics.

References

Brzović, Zdenka, and Predrag Šustar. 2020. "Postgenomics Function Monism." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, https://doi.org/10.1016/j.shpsc.2019.101243.

Cummins, Robert C. 1975. "Functional Analysis." *Journal of Philosophy* 72: 741–64.

Doolittle, W. Ford, Tyler D. P. Brunet, Stefan Linquist, and T. Ryan Gregory. 2014. "Distinguishing between 'Function' and 'Effect' in Genome Biology." *Genome Biology and Evolution* 6 (5): 1234–37. https://doi.org/10.1093/gbe/evu098.

Germain, Pierre-Luc, Emanuele Ratti, and Federico Boem. 2014. "Junk or Functional DNA? ENCODE and the Function Controversy." *Biology & Philosophy* 29 (6): 807–31. https://doi.org/10.1007/s10539-014-9441-3.

Graur, Dan, Yichen Zheng, Nicholas Price, Ricardo B. R. Azevedo, Rebecca A. Zufall, and Eran Elhaik. 2013. "On the Immortality of Television Sets: 'Function' in the Human Genome According to the Evolution-Free Gospel of ENCODE." *Genome Biology and Evolution* 5 (3): 578–90. https://doi.org/10.1093/gbe/evt028.

Keeling, Diane Marie, Patricia Garza, Charisse Michelle Nartey, and Anne-Ruxandra Carvunis. 2019. "The Meanings of 'function' in Biology and the Problematic Case of de Novo Gene Emergence." *ELife* 8. https://doi.org/10.7554/eLife.47014.

Maley, Corey J., and Gualtiero Piccinini. 2018. *A Unified Mechanistic Account of Teleological Functions for Psychology and Neuroscience*. Vol. 1. Oxford University Press. https://doi.org/10.1093/oso/9780199685509.003.0011.

On politics and social science – the subject-object problem in social science and Foucault's engaged epistemology

Milan Urošević

Philosophy of social sciences

On politics and social science – the subject-object problem in social science and Foucault's engaged epistemology

Keywords: social science, epistemology, researcher, genealogy, Foucault

Abstract: The epistemological problem of the relationship between the subject of knowledge and the object being known has it's form in social science as a problem of the relationship between a social scientist as a researcher and society and it's phenomena as an object of this inquiry. As Berger and Kellner note in their book "Sociology Reinterpreted" a social scientist is necessarily a part of the object he studies, being embedded in a position in society from which he studies it. Hence social sciences as scientific endeavors face a problem of the inseperability of their researchers from object they study. Two main solutions two this problem have arisen: positivism and interpretivism. Positivism postulates that rigorous methods for research will insure that objective knowledge will be produced while interpretivism sees society only as an aggregate of individuals whose interactions should be interpreted. A third epistemological framework has arisen in the first half of the twentieth century usually called "critical theory". Critical theory states that researchers should aim their research towards changing the object they are researching, therefore their scientific practice should have extra-scientific effects, namely political effects. This perspective violates Webers postulate of value neutrality which claims that social sciences can only study the state of affairs but can't subscribe desirable ways of action. As we will see the main topic of our paper is the epistemological framework of the work of Michel Foucault and his contribution to the resolution of the problematic relation between a researcher and his research object in social science. We will claim that Foucault broadly falls into the critical theory paradigm but manages to solve it's conflict with the value neutrality postulate. Foucault envisions society as an amalgam of discursive and non-discursive practices that interconnect in a way that gives them regularity and coherence through time. As Gayatri Spivak notices for Foucault discursive practices create meaning and in doing so chart a way for non-discursive practices and therefore for action. This can be seen as an explanation for Foucault's well known postulate of the relationship between power and knowledge, discursive practices create knowledge that makes visible certain paths for action. Both of these types of practices intertwine to create what Foucault calls "dispositifs" that can be seen as mechanisms that bind

discursive and non-discursive practices in a coherent manner and enable their regular repetition through time. Foucault calls his methodology "genealogy" and sees it as a historical research of the emergence of dipositifs. Genealogy is a historical research of the contingent ways in which practices got interconnected in the past to create dispositifs we see today. As Foucault claims genealogy begins with a "question posed in the present" about a certain dispositive and then charts historical events and processes that led to its current form. The main aim of genealogy is to show that there is no transcendental necessity for a certain dispositif to exist in it's current form by exposing the historical contingency that led to it's current state. Foucault claimed that his intent was to show that there is no metaphysical necessity that grounds the existences of dispositifs and hence that their current form is arbitrary. As we can see Foucault follows his postulate on the relationship between knowledge and power and formulates his scientific practice as an opening of possibilities for different forms of action. This is way he calls his books "experiments" and claims that they are to be used for readers to re-examine their own links to the currently existing dispositifs and possibilities of their alternative arrangements. But as Foucault claims the genealogical method doesn't include normative prescriptions and can be seen only as a form of an anti-metaphysical "unmasking" of current dispositifs. This unmasking doesn't prescribe a desirable form to any dispositive but only shows that it can be arranged in different ways. Hence we can say that Foucault sees the relationship between a researcher and his object of study as a form of an intervention of the subject that aims at showing that the object is an arbitrary construction. In that regard Foucault falls into the critical theory paradigm. Where he differs from critical theory is his anti-normative stance that refuses to prescribe any desirable form of action unlike for example Horkheimer who in his essay on critical theory claims that "the task of the theorist is to push society towards justice". Foucault claims that his research results should be used as "instruments" in political struggles but he himself doesn't ever proclaim a desirable political goal. So we can conclude that Foucault solves the problem of the subject-object relation in social science by envisioning the research process as a practice of production of tools for social change. Therefore he connects social science to extra-scientific political goals but doesn't violate the value neutrality postulate because his research doesn't prescribe any concrete political goals but only shows the possibility for social change.

Philosophy of social sciences

On politics and social science – the subject-object problem in social science and Foucault's engaged epistemology

Keywords: social science, epistemology, researcher, genealogy, Foucault

Abstract: This paper will deal with the epistemological problem of the relationship between the subject and the object of knowledge in social science. We will try to show how the epistemological framework of Michel Foucault can be used to tackle this problem. Our point will be that Foucault's methodology solves this problem by envisioning the research process as a practice of an intervention into the object. This intervention is envisioned as a form of criticism of the researched object by showing its contingent and arbitrary nature.

The hypothesis of “hidden variables” as a unifying principle in physics

Louis Vervoort

Section b) Philosophy of Natural Science

The hypothesis of “hidden variables” as a unifying principle in physics.

12.01.2020

Abstract. In the debate whether ‘hidden variables’ could exist underneath quantum probabilities, the ‘no hidden-variables’ position is at present favored. In this article I attempt to provide a more equilibrated verdict, by pointing towards the heuristic and explanatory power of the hidden-variables hypothesis. I argue that this hypothesis can answer three foundational questions, whereas the opposing thesis (‘no hidden variables’) remains entirely silent for them. These questions are: 1) How to interpret probabilistic correlation ? (a question considered by Kolmogorov “one of the most important problems in the philosophy of the natural sciences”, and first analysed by Reichenbach); 2) How to interpret the Central Limit Theorem ?; and 3) Are there degrees of freedom that could unify quantum field theories and general relativity, and if so, can we (at least qualitatively) specify them ? It appears that only the hidden-variables hypothesis can provide coherent answers to these problems; answers which can be mathematically justified in the deterministic case. This suggests that the hidden-variables hypothesis should be considered a legitimate candidate as a guiding, unifying principle in the foundations of physics, against the mainstream position.

Keywords: determinism, indeterminism, hidden variables, quantum mechanics, probability theory, Bell’s theorem, Reichenbach’s principle

1. Introduction.

The question whether the universe is ultimately deterministic or indeterministic (probabilistic) concerns one of the oldest debates in the philosophy of physics. Modern quantum mechanics, a probabilistic theory, has convinced many that indeterminism wins; but a more careful analysis, based notably on the interpretation of the theorems of Gleason, Conway-Kochen and Bell, shows that the debate is actually undecided (cf. e.g. Wuethrich 2011). This may be a shared belief in the philosophy of physics community; but it surely is unpopular outside that community. A broader question can be condensed in following slogan: Can hidden variables exist underneath quantum probabilities? Can quantum probabilities be reduced to, or ‘explained by’, deeper-lying variables? (A precise mathematical formulation of this question will be given in Section 2.)

Assuming the existence of such hidden variables is a more general hypothesis than determinism: determinism corresponds to the extremal case where the variables are deterministic, a mathematically well-defined special case of more general probabilistic variables (cf. Section 2). Now, the generally received wisdom is again that such hidden variables can*not* exist, at least if one only considers *local* hidden variables – which is what I will do throughout this talk, since nonlocal variables involve, by definition, superluminal interactions, so interactions that are not Lorentz-invariant and in overt contradiction with relativity theory. Thus, I will only inquire here about the possibility of local degrees of freedom (variables).

The question of the possibility of hidden variables (HV) underlying quantum probabilities is likely most clearly investigated through Bell's theorem (Bell 1964). Indeed, Bell's theorem is the only standard chapter in physics that explicitly mentions the concepts of cause and determinism and investigates them in the light of quantum mechanics. Bell's seminal article of 1964 starts as follows: "The paradox of Einstein, Podolsky and Rosen was advanced as an argument that quantum mechanics could not be a complete theory but should be supplemented by additional variables. These additional variables were to restore to the theory causality and locality" (1964, p. 195). Now, according to Bell's theorem local HV theories contradict quantum mechanics in certain experiments. Since these Bell experiments have vindicated the quantum prediction countless times, most scholars believe now that the prospects for local hidden-variable theories, and thus for the HV-hypothesis, are dim. In other words, according to the standard view quantum probabilities are irreducible in general; they cannot be understood as emerging from a deeper-lying HV level.

My goal here is to argue that not only is the HV-hypothesis not refuted by Bell's theorem and the Bell experiments (in line, notably, with Wuethrich 2011), but this hypothesis, compared to its competitor ('no HV' or irreducibility), has the greater explicatory power. Specifically, I will show in Section 3 that the HV-hypothesis can provide coherent answers to three questions of the foundations of probability and of physics, whereas irreducibility remains entirely silent for them. These questions are: 1) How to interpret probabilistic correlation ? (a problem raised by Kolmogorov and first analyzed by Reichenbach); 2) How to interpret the Central Limit Theorem?; and 3) Are there variables that could unify quantum field theories and general relativity, and if so, can we say at least something about them ? It appears that in particular *deterministic* HV allow to mathematically prove answers to questions 1) and 3); and to conjecture a proof of an answer to question 2). To further illustrate the heuristic power of determinism[1], I will recall that it corresponds, at least in certain physical situations, to the 'best informed' epistemic position (Section 2).

This talk is organized as follows. In Section 2 I will extract from the literature a straightforward definition of 'deterministic' and 'probabilistic / stochastic' variables, needed in the remainder of the article. This will allow to define 'reducibility' of (quantum) variables via (yet unknown) additional variables – the so-called hidden variables. Next, it appears that some confusion exists regarding the concept of 'objective (and subjective) probability': this concept can have two quite different meanings. To illustrate my argument, it will be helpful to look in detail at a realistic case, namely the automated tossing of a large die: an experiment that can equally well be described as a deterministic or a probabilistic system, depending on the epistemic status of the experimenter (in line with e.g. Suppes 1993). Hence it will appear useful to introduce the notion of "relativity or subjectivity of (in)deterministic ascription". This simple thought experiment does not claim to provide new results, but will allow to illustrate and clarify the notions of reducibility and determinism, and thus to throw light on the three foundational questions mentioned above (related to probabilistic dependence, the Central Limit Theorem and the unification of QFT and GR). These questions are the subject of Section 3, containing the main results of this work. Section 4 will conclude.

Selected References.

Bell, J. (1964). "On the Einstein-Podolsky-Rosen Paradox", Physics, 1, 195-200
Earman, J. (1986). A Primer on Determinism, Reidel, Dordrecht
Earman, J. (2008). "How determinism can fail in classical physics and how quantum physics can (sometimes) provide a cure", Philosophy of Science 75, 817-829.
Earman, J. (2009). "Essential self-adjointness: implications for determinism and the classical-quantum correspondence", Synthese 169, 27-50.
Hoefer, C. (2016). "Causal Determinism", The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2016/entries/determinism-causal/>.
Reichenbach, H. (1956). The Direction of Time, Univ. of California Press, Berkeley - Los Angeles
Suppes, P. (1993). "The Transcendental Character of Determinism," Midwest Studies in Philosophy, 18: 242–257.
Wuethrich, C. (2011). "Can the world be shown to be indeterministic after all?" In Probabilities in Physics, ed. C. Beisbart and S. Hartmann, pp. 365-389, Oxford University Press, Oxf

[1] For a detailed treatment of physical determinism, see notably Earman (1986), (2008), (2009); and Hoefer (2016).

Modules as an adaptationist discovery heuristic

David Villena

Modules as an adaptationist discovery heuristic
David Villena

Cognitive modules are internal structures. Some theorists invoke them in order to explain human cognitive capacities that are hypothesized to perform specific tasks. In particular, cognitive modules could be understood either as information-processing mechanisms or as bodies of mental representations. Traditional approaches to the different fields of psychology set a methodology centered in the study of proximate causes of behavior. Arguably, the novelty and uniqueness, if any, of the approach championed by evolutionary psychology is to suggest a computational quest for ultimate causes in the realms of psychology and cognitive science. Thanks to their methodological adaptationism, evolutionary psychologists predict either behavior caused by modules whose existence is unknown to us so far or behavior that indicates unknown structural properties of already known modules. If the corresponding empirical tests are successful, we are said to gain new knowledge about the internal structure of the human mind. I discuss the heuristic value of this research program as well as the cogency of its underlying adaptationist approach to evolution and computationalist view about the mind.

Keywords: philosophy of cognitive science, adaptationism, computationalism, massive modularity of mind, evolutionary psychology

Quantum contextuality and ontic indefiniteness

Marek Woszczek

Contextuality is a fundamental, irreducible physical property of quantum systems, which has been experimentally demonstrated on many diverse physical systems. Mathematically, it is expressed by the (Bell–)Kochen–Specker Theorem, which says that it is impossible to consistently assign the pre-existing 0,1-values to all possible physical observables on a quantum system for dim>2. However, there is a growing controversy concerning its interpretation, both physical and philosophical. In the first part of the talk, I shall define it in the sheaf-theoretic framework as a topological property, and argue that it should be construed fully ontologically as a very primitive physical phenomenon behind the purely quantum information processing in nature, which has little to do with epistemic constraints of observers. In the second part, I shall discuss in detail the ontic definiteness as a metaphysical assumption responsible for preferring the hidden variable models of quantum contextuality. It will be indicated why there are good reasons to reject it altogether, and why such a rejection is sometimes conceived as philosophically controversial. I shall also discuss the connection between contextuality, ontic indefiniteness and the quantum probability theory.

Vitality: A Multi-Layered View

Mustafa Yavuz

b) Philosophy of Natural Science (Better if there is a symposium on Philosopyhy of Biology)

Vitality: A Multi-Layered View

Abstract
In this study, referring to the historical definition of biology that can be considered as the scientific study of life, I will try to put forth some assumptions and explanations in order to give a basic definition of life or vitality. After investigating some relevant terms in the contemporary biology to improve them, I will mention the necessity of revision and distinction in accordance with the given explanations. The first is the need for an update of the term known as "homeostasis" into "homeokinesis". For this reason, I emphasize some propositions in order to clarify the distinction of meaning. The following three questions were repeated in the motivation and structure of this study: 1) What is vitality? 2) Does it have any degrees or grades? 3) Then what is the fundamental state of life?

In the quest of an answer to these questions, different topics in biology will be mentioned as references. At the same time, by claiming that vitality is an emergent process, I follow the emergentist view, rather than the reductionist view on the statements of biological facts and phenomena. As a result, considering the cell level as a base, in the hierarchy of the biological organization, I put forth a threefold-vitality-view as emergent in the levels of the cell, the organ and the organism. I also count consciousness and memory immanent in every level of the proposed multi-layered model. In this multi-layered view of vitality, I can define vitality as the regulation of the flow process between the internal and external loads of a cell, organ and organism. In addition to this definition, I emphasize that vitality is mechanized by a metabolism to balance the homeokinetic range between the minimum and maximum degrees at the cell, organ and organism levels. Vitality is not only an emergent coordination, but also immanent a type of consciousness and memory in each level. I support the idea that a whole is more than the sum of its parts. Therefore, I give an example from chemistry which provides and effective understanding of some phenomena. Vitality is an emergent phenomenon that occurs within the cell, the organ and the organism consequently. For these three levels, I make an allusion to the anoetic, noetic, and autonoetic consciousness types posited by Tulving (1985) in his APA Award winning study.

Key Words: Life, Cell, Homeokinesis, Consciousness

Modeling in Biomedicine: Extending the notion of models

Martin Zach

According to a widely held view – herein called description-driven modeling account – scientific modeling of real-world phenomena consists of constructing a simplified surrogate system that serves as the target of investigation and is only later compared to the real-world system. This strategy of theorizing has been construed as a way of representing target systems indirectly, and distinguished from yet another strategy of theorizing, the abstract direct representation (Godfrey-Smith [2006]; Weisberg [2007]). Though traces of doubt, sometimes more sometimes less explicit, can be found in the literature (Knuuttila and Loettgers [2017]; Mitchell and Gronenborn [2017]; Plutynski and Bertolaso [2018]), more can be said with respect to the general applicability of this account of modeling. Rather than raising worries due to mostly metaphysical reasons as some have done (Toon [2012]; Levy [2015]), the focus of this paper will be on scientific practice.

Drawing on methods of participatory observation in laboratory settings and an extensive analysis of scientific literature, this paper aims to describe the process of building mechanistic models in the field of cancer immunology. More specifically, the various experiments involving, among others, cell cultures, animal models and imaging techniques in studying the role of myeloid-derived suppressor cells in facilitating the phenomenon of cancer dissemination will be discussed. As a result, it will be argued that building mechanistic models in much basic biological research proceeds in a different way than the widely held account would have us believe. To capture these practices, the notion of 'experimentation-driven modeling' will be introduced and conceptually distinguished from the notions of abstract direct representation and data models (Woodward [2011]; Leonelli [2019]). Mechanistic models that are derived (no connection to the logical notion of derivation) from experiments are being built by integrating piecemeal experimental results into a comprehensive account of a mechanism that is responsible for the phenomenon at hand.

Importantly, this account of modeling should be construed as complementing (rather than replacing) the existing and widely held account in that it captures practices that the received view does not account for except for in a highly artificial way.

Topological explanation in system-theoretical linguistics

Lukas Zamecnik

c) Philosophy of Cognitive and Behavioral Sciences

Topological explanation in system-theoretical linguistics

Key words:

Non-causal explanation, topological explanation, counterfactual dependency, explanatory asymmetry, system-theoretical linguistics

In the current debates on the nature of scientific explanation, we note the controversy over the nature of non-causal explanations. The spectrum of views on the nature of this is varied, ranging from those who reject its exclusivity (e. g. Skow 2016), to those who claim a plurality of explanatory strategies (e. g. Bokulich 2018) to those who claim the exclusivity of non-causal explanations (e. g. Kostic 2019a, 2019b, Lange 2017). Our position advocates the exclusivity of non-causal explanations in the context of physics, life sciences and in cognitive and behavioral sciences, including linguistics.

In recent discussions, it has been often recalled that the essential conditions to be met by a valid model of explanation include support of counterfactual (i.e. the existence of counterfactual dependency between explanans and explanandum) and explanatory asymmetry (i.e. If *A* explains *B* then *B* cannot explain *A*.). Some critics have refuted that the non-causal explanation is unable to incorporate these conditions. Following Kostic (2019a) we hold that these conditions can be fulfilled for at least some types of non-causal explanations. The non-causal model of explanation that Kostic explicitly deals with is a topological model (Kostic 2019a, 2019b, 2018). This explanatory model has so far been applied primarily in the context of life sciences, such as Huneman 2010 and Kostic 2019a.

The first aim of this paper is to extend the spectrum of applications of the topological model of explanation beyond the examples given so far, to the area of quantitative linguistics (Altmann 1978), namely to system-theoretical linguistics (Köhler 2012, 1986). The second aim is to show, through the new example cited, the feasibility of the stated conditions (counterfactual dependency, explanatory asymmetry) for the topological type of non-causal explanation.

Quantitative linguistics seek to describe and explain the properties of texts and speech through corpus analysis and by means of mathematical statistics, using some statistical distribution, in quantitative linguistics typically named as laws, for example the Zipf law and Menzerath-Altmann law. A specific feature of system-theoretical linguistics is the ability to represent individual linguistic subsystems (e.g. lexical, morphological, syntactic) as self-regulating circuits of linguistic variables, for example for lexicon: length, frequency, number of meanings of lexical unit, and provide functional explanation of the state of the system, which is determined by outside-system requirements such as coding and decoding efforts, and memory limitation, etc.

The functional explanation brings some obligations, for example the problem with functional equivalents, (Hempel 1965) that system-theoretical linguists could not eliminate (Köhler 2012). Therefore, we consider it useful as the third aim of the paper, and in connection with a pluralistic explanatory strategy (Bokulich 2018), to explain the behavior of linguistic subsystems using a topological model. Of course, the pluralistic explanatory strategy assumes that the topological model of the explanation also brings some undesirable obligations. Therefore, the fourth goal of the paper is to ascertain the new commitments arising from the application of this model of explanation.

In summary, the contribution seeks:

1. To apply a topological model of explanation in a new context, namely in system-theoretical linguistics.

2. To demonstrate the feasibility of counterfactual dependency and explanatory asymmetry conditions for a topological model of explanation by a new example from system-theoretical linguistics.

3. To demonstrate the benefits of a pluralistic approach to explanation, i.e. that the difficulties of functional explanation are not shared by topological explanation.

4. To ascertain any new undesirable obligations that will result from the application of the topological explanation model for system-theoretical linguistics.

References:

Altmann, G. (1978) Toward a Theory of Language. In: Glottometrika, 1, 1–25.

Bokulich, A. (2018) Searching for Non-causal Explanations in the Sea of Causes. In: Reutlinger, A., Saatsi, J. (eds.) Explanation Beyond Causation, Oxford: Oxford University Press.

Hempel, C., G. (1965) The Logic of Functional Analysis. In: Hempel, C., G. Aspects of Scientific Explanation and other Essays, New York: The Free Press, 297–330.

Huneman, P. (2010) Topological explanations and robustness in biological sciences. In: Synthese, 177, 2, 213–245.

Köhler, R. (2012) Quantitative Syntax Analysis. Berlin: De Gruyter.

Köhler, R. (1986) Zur linguistischen Synergetik. Struktur und Dynamik der Lexik. Bochum: Brockmeyer.

Kostic, D. (2019a) General Theory of Topological Explanations and Explanatory Asymmetry. In: Philosophical Transactions of the Royal Society B: Biological Sciences (forthcoming).

Kostic, D. (2019b) Mathematical and non-causal explanations: an introduction. In: Perspectives on Science, 27, 1, 1–6.

Kostic, D. (2018) Mathematical and topological explanations: an introduction. In: Synthese, 195, 1, 1–10.

Lange, M. (2017) Because without Cause. Oxford: Oxford University Press.

Skow, B. (2016) Scientific Explanation. In: Humphreys, P. (ed.) The Oxford Handbook of Philosophy of Science, Oxford: Oxford University Press, pp. 524–543.

Seeing colours unconsciously

Paweł Zięba

Seeing colours unconsciously

According to unconscious perception hypothesis (UP), episodes of the same fundamental kind as ordinary conscious seeing can occur unconsciously. The proponents of UP often support it by invoking empirical evidence for a more specific hypothesis, according to which colours can be seen unconsciously (UPC). That evidence includes studies using transcranial magnetic stimulation (Boyer et al. 2005), metacontrast (Norman et al. 2014), and dichoptic chromatic masking (Moutoussis and Zeki 2002).

The aim of this talk is to investigate the ramifications of UPC for the debate about the phenomenal character of visual perception. The latter is a set of properties that determine what it is like to consciously see. Colour is a paradigmatic example of such a property. For present purposes, it is useful to categorize theories of the phenomenal character with respect to how they respond to the following questions:

(Q1) Is the claim 'a perception P has a phenomenal character' equivalent to the claim 'a perception P is conscious'?

(Q2) Is the phenomenal character of perception constituted by mind-dependent properties, or is it constituted by mind-independent properties?

The resulting classification divides theories of the phenomenal character into four types:

	Q1: equivalence	Q1: no equivalence
Q2: mind-dependent	equivalence internalism	inequivalence internalism
Q2: mind-indepenent	equivalence externalism	inequivalence externalism

While UPC admits of being incorporated into all four accounts, closer inspection reveals that some of them are better suited to accommodate it than others.

Equivalence theories reject the idea of unconscious phenomenal character (see e.g. Prinz 2012). Proponents of such views are also prone to dismiss both UP and UPC (Phillips 2018). Inequivalence theories, by contrast, argue that the phenomenal character is consciousness-independent (see e.g. Marvan and Polák 2017), and thereby predict that UP and UPC are true. Inequivalence theories are thus preferable over the equivalence theories, at least as far as accommodating UP and UPC is concerned.

Internalist theories claim that the phenomenal character is produced in the subject when the visual system is stimulated in the right way by environmental factors. According to externalist theories, by contrast, the phenomenal character is constituted by mind-independent properties of the perceived scene. Hence the internalism vs. externalism distinction in the table above maps onto intentionalism vs. relationalism distinction in metaphysics of

1

perception. Nevertheless, the choice between inequivalence internalism and inequivalence externalism does not boil down to picking a side in the intentionalism vs. relationalism debate. This is because the internalist faces a dilemma that the externalist does not face.

On the first horn, the phenomenal character is produced in visual cortex. This validates phenomenal overflow, i.e. the possibility of phenomenally conscious access unconscious perception, which in turn casts doubt on the possibility of unconscious perception (because it reinforces the objection that the putative instances of unconscious perception are in fact cases of overflow).

On the second horn, the phenomenal character is not produced in visual cortex. In this case, the colour internalism that postulates unconscious production of phenomenal character in visual cortex is false. If the phenomenal character can be unconscious, it cannot be produced in the neural basis of consciousness either (so the prefrontal cortex is not an option). And indicating any other pattern of brain activity as the producer of unconscious phenomenal character will prompt the overflow theorist to argue that this is where phenomenal consciousness is produced, which will undermine the idea of seeing colours unconsciously.

The tension between overflow and unconscious perception hinders any internalist account of UPC. While the dilemma does not show that colour internalism is completely unable to accommodate UPC, it renders externalism preferable because the latter does not face the dilemma. Consequently, UPC constitutes a good reason to think that the phenomenal character of visual perception is comprised of primitive mind-independent properties.

References:

Boyer, J.L., S. Harrison, and T. Ro. 2005. Unconscious processing of orientation and color without primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* 102: 16875–16879.

Marvan, T., and M. Polák. 2017. Unitary and dual models of phenomenal consciousness. *Consciousness and Cognition* 56: 1–12.

Moutoussis, K., and S. Zeki. 2002. The relationship between cortical activation and perception investigated with invisible stimuli. *Proceedings of the National Academy of Sciences of the United States of America* 99: 9527–9532.

Norman, L.J., K. Akins, A. Heywood, and R.W. Kentridge. 2014. Color Constancy for an Unseen Surface. *Current Biology* 24: 2822–2826.

Phillips, I. 2018. Unconscious Perception Reconsidered. *Analytic Philosophy* 59: 471–514.

Prinz, J. 2012. *The Conscious Brain*. New York: Oxford University Press.

Symposia

Symposium on the History of Formal Logic in Eastern Europe

Mate Szabo

Symposium on the History of Formal Logic in Eastern Europe

proposed by Máté Szabó

Even though formal logic has a strong tradition in Eastern Europe, its rich history is under-represented in the international academic literature (except possibly some of the Warsaw-Lvov School's achievements). The aim of this symposium is to change this situation, and to bring together logicians, philosophers and historians to address some of the understudied chapters of logic in Eastern Europe.

The four talks of the symposium discuss chapters in the history of logic in Czechia, Hungary (and partially Poland), Romania and Serbia. They cover many aspects of the history of the field. The diversity of the focus of the talks, ranging from recursion & computability theory to set theory, as well as more foundational issues, exemplify the richness of the field in the region. They also encompass a long period of time, with the talks concerning Hungary and Romania beginning in the 1940s, while those from Czechia and Serbia cover the 1960s and 1970s up to today.

Vopěnka's Alternative Set Theory
within Twentieth Century Mathematics

Zuzana Haniková, e-mail: zuzana.hanikova@gmail.com

From Elementary School to $\mathcal{E}_3$,
the Early History of Elementary Functions

Walter Dean, e-mail: whdean@gmail.com
Máté Szabó, e-mail: www.mate@gmail.com

The Plurality of Logics,
Between Human Thinking and Formal Systems

Constantin C. Brîncuş, e-mail: c.brincus@yahoo.com

From Recursion to Deduction,
Two Strands of Modern Logic in Serbia

Jovana Kostić, e-mail: kostic.jovana345@gmail.com
Katarina Maksimović, e-mail: katamax21@gmail.com
Miloš Adžić, e-mail: milos.adzic@tutanota.com

Vopěnka's Alternative Set Theory within Twentieth Century Mathematics

Zuzana Haniková

Petr Vopěnka presented his Alternative Set Theory (AST) in the monograph, *Mathematics in the Alternative Set Theory*, published by Teubner Verlag in 1979. Prior to this publication, the AST had been developed for a decade in a research seminar headed by himself and attended by a group of graduate students and collaborators, including Antonín Sochor, Josef Mlček, or Karel Čuda, and these explorations yielded dozens of papers available via the Czech Digital Mathematical Library (dml.cz). As a sidenote, during this decade Vopěnka was prevented from publishing his own work, maintaining international scientific contacts, or developing his career, for political reasons. Vopěnka subsequently published another monograph on the AST in 1989, and offered several papers and lectures that detailed the philosophical and historical motivations for the AST. The latter include the works of Leibniz and Bolzano; as a matter of fact, Vopěnka's interests in various areas of the history of mathematics were quite broad, producing works ranging from geometry to analysis and set theory.

However, the AST also has tight links to developments in mathematics in the 20th century, roughly from the 1930's onwards: namely, Skolem's work in nonstandard models of arithmetic, Vopěnka's nonstandard models of set theory, Robinson's nonstandard analysis, Vopěnka and Hájek's work in the theory of semisets, or Parikh's feasible arithmetic. These are only some major mathematical influences; the AST also reflects Vopěnka's growing interest in phenomenology. Moreover, one can find parallel developments that were presumably independent of the AST but share some of its motivations, background, and its timing, such as the axiomatic nonstandard set theories proposed by Nelson and by Hrbáček. This talk will present and discuss the twentieth century mathematical narrative for the AST and demonstrate how the AST can be understood as a natural continuation of the above inspirations and experience.

From Elementary School to $\mathcal{E}_3$,
the Early History of Elementary Functions

Walter Dean and Máté Szabó

Kalmár introduced *elementary functions* in his 1943 paper *Egyszerű példa eldönthetetlen aritmetikai problémára* (A Simple Example for an Undecidable Arithmetical Problem). The paper presents a simplified proof of Gödel's first incompleteness theorem wherein Kalmár described a class of functions sufficient for the arithmetization of the syntactic notions he employs but which is apparently narrower than the primitive recursive functions. Péter (1954) later explained the naming of the class by the use of operations (successor, addition, subtraction, bounded sum and product, and substitution) that one learns in elementary school. Kalmár (1957) would later propose that elementary functions originate from the experience of counting and organizing concrete objects.

Kalmár's paper was originally published in Hungarian and some of the earliest results about the elementary functions were obtained by his colleagues Csillag and Péter, and his student Bereczki. Nonetheless, the elementary functions quickly became more broadly known for at least two reasons: 1) Kleene (1952) prominently advertised Kalmár's result that the Normal Form Theorem for recursive functions can be strengthened such that the T-predicate is elementary (rather than primitive recursive); 2) Grzegorczyk (1953) introduced a hierarchy which properly stratifies the primitive recursive function in which the elementary functions form the fourth level, $\mathcal{E}_3$. Such results would go on to inspire other proof- and complexity-theoretic characterization of the elementary functions (e.g. Rose 1984, Clote 2002). They additionally suggest that the elementary functions may be understood as analyzing a notion more restrictive than that of effective computability.

In our talk we will analyze Kalmár's philosophical motivation and contrast his approach with the route by which Skolem (1923) was led to his "recursive mode of thought" and also the role recursive definition were assigned by Hilbert and his collaborators (e.g. 1926, 1934). In addition, we will examine the relation of the elementary functions relative to subsequent investigations of computational feasibility.

The Plurality of Logics,
Between Human Thinking and Formal Systems

Constantin C. Brîncuş

Logical pluralism is a fact, i.e., there is a plurality of logics. Each system of logic can be presented as a deductive system of theorems obtained from axioms or assumptions on the basis of certain rules. An important question suggested by this fact is the one that relates human thinking with this plurality of logics: is there a unique system of thinking used to construct and characterize all these logics or there are different systems of thinking characterized by these logics? Starting from the observations that the Romanian logician Grigore C. Moisil made in his 1944 article, *Logical Pluralism*, I will analyze in my presentation two problems: 1) is there a unique system of thinking that we use, expressed by a unique system of logical rules, when we talk about these different logics, or in each metalogic we have to use the system of logic that it describes? 2) Classical mathematics has been developed by using classical logic and it is an important instrument in mathematical physics for describing nature. Can we build a mathematical physics by using a non-classical mathematics based on a non-classical logic? I will argue in my presentation, in the spirit of Saul Kripke's line of thought, used in his 1974 lectures on the nature of logic, that some logical rules cannot be "adopted" and in this sense many logical systems are only formal systems that can be used by human thinking for different purposes.

From Recursion to Deduction,
Two Strands of Modern Logic in Serbia

Jovana Kostić, Katarina Maksimović and Miloš Adžić

We will trace some of the keystones of the development of logic in Serbia, during the last seventy years. The first strand that we will follow begins in the early 1960s with Vladeta Vučković, the first modern Serbian logician. At the time, he published a series of papers in computability theory culminating in his investigation [2] of the notion of *retraceable* sets, introduced by Dekker and Myhill. Call a set $A \subseteq \mathbb{N}$ *almost recursive*, if for every $x \in A$ we can effectively find the number of elements of A less than x. Vučković showed that every retraceable set is almost recursive, and that every almost recursive set with recursively enumerable complement is retraceable, thus giving a useful characterization of this interesting generalization of retraceable sets. He also showed that every Turing degree is almost recursive and that each recursively enumerable Turing degree is a degree of recursively enumerable set whose complement is almost recursive.

The second strand began during the late 1970s. At that time, Kosta Došen investigated the characterizations of logical constants in classical, intuitionistic, and other substructural logics by equivalences between sequents in which the constant to be analyzed appears in a certain position inside the sequent and a structural sequent, i.e. one in which no logical constant appears [1]. Behind this type of characterization is the assumption that the structural part is the main part of logic and that the logical constants are secondary - they play the same role inside different structural contexts. Approaching proof theory using the tools of category theory, Došen came to realize that his early analyses of logical constants are but superficial aspects of *adjunction*, one of the central notions of category theory. It is in the field of categorial proof theory that Došen, together with his student Zoran Petrić, made his greatest achievements.

Cited works

[1] Došen, Kosta, LogicalConstants as Punctuation Marks, *Notre Dame Journal of Formal Logic*, Vol. 30, 1989.

[2] Vučković, Vladeta, Almost Recursive Sets, *Proceedings of the American Mathematical Society*, Vol. 23, 1969.

Free will and consciousness as a problem of philosophy of science: What can philosophy and science learn from each other?

Maria Sekatskaya and Louis Vervoort

EENPS-2021 Symposium, Belgrade

Free will and consciousness as a problem of philosophy of science: What can philosophy and science learn from each other?

Louis Vervoort, School of Advanced Studies, University of Tyumen, Russia
Maria Sekatskaya, Heinrich Heine University, Dusseldorf, Germany

EENPS-2021 Symposium, area: ***c)*** *Philosophy of Cognitive and Behavioral Sciences, or* ***b)*** *Philosophy of Natural Science*

General Description

The "problem of free will" (What is free will? Do we really have free will?) is a classical problem of philosophy; a staggering number of great minds have expressed their opinion on this topic, virtually from all disciplines, including Aristotle, Aquinas, Hume, Kant, Spinoza, and Einstein. While philosophy was the mother discipline from which the topic sprang, in recent times several other disciplines have joined the debate, in particular psychology, neuroscience, cognitive science, computer science and physics. Both in professional and broad-public texts the link between free will, consciousness and (in)determinism is often immediately made. For instance, the Internet Encyclopedia of Philosophy starts its entry on free will thus[1]: "Most of us are certain that we have free will, though what exactly this amounts to is much less certain. According to David Hume, the question of the nature of free will is 'the most contentious question of metaphysics.' If this is correct, then figuring out what free will is will be no small task indeed. Minimally, to say that an agent has free will is to say that the agent has the capacity to choose his or her course of action. But animals seem to satisfy this criterion, and we typically think that only persons, and not animals, have free will. […] This article considers why we should care about free will and how freedom of will relates to freedom of action. It canvasses a number of the dominant accounts of what the will is, and then explores the persistent question of the relationship between free will and causal determinism […]." The typical human capacity referred to in this passage, not shared by animals, is usually considered to be consciousness or (other) cognitive capacities.

In recent years, a strong impetus has been given to the theoretical philosophical research by experimental advances in neurobiology, and by an increasing interest in humanoid functions and capacities that could be realized by robots, in general systems steered by artificial intelligence (AI). For instance, in 2008 neuroscientists have reported, based on the measurement of brain activity by fMRI, that 'free' choices of test persons (namely the choice to lift their left or right hand) could be predicted up to 10 seconds (!) before the test person consciously made the decision to pick one or the other hand[2]. To many researchers, especially neurobiologists, scientific results as these put free will in question. As another example among the many, in 2017 cognitive neuroscientists published an article in *Science* entitled "What is consciousness, and could machines have it?"[3] – an example of the exponentially rising interest in machine-based forms of consciousness. In physics too, the question of free will has been discussed in 2018 and linked to one of the key problems of physics, i.e. the unification of quantum mechanics and relativity theory – namely by Nobel laureate Gerard 't Hooft[4].

In this symposium, we start from the assumption that there is a clear case for studying free will and consciousness through an interdisciplinary lens. While free will has traditionally been studied, mostly, within metaphysics and philosophy of mind, we believe the recent upsurge of scientific findings warrants the

[1] Cf. internet website https://www.iep.utm.edu/freewill/, retrieved 12.01.2020.
[2] Chun Siong Soon et al., *Unconscious determinants of free decisions in the human brain*, Nature Neuroscience 11, 543 - 545 (2008)
[3] Stanislas Dehaene et. al, *What is consciousness, and could machines have it?*, Science 27, Vol. 358, Issue 6362, pp. 486-492 (2017).
[4] Gerard 't Hooft, *Free Will in the Theory of Everything*, arXiv:1709.02874 [quant-ph] (2018), cf. https://arxiv.org/abs/1709.02874

participation of methods and insights from philosophy of science to study this problem. Thus, for this symposium, we invite contributions from philosophy, computer science / IT, psychology, natural science etc., with special emphasis on contributions that open a door towards other disciplines than the original one. An indicative list of the topics of the symposium:

- Recent insights in the philosophy of free will and consciousness
- Recent insights in computer science and AI research related to consciousness / free will / cognition
- Free will and consciousness from the perspective of philosophy of science and of cognitive science
- Insights from neuroscience, psychology, natural science etc.

****

The list of the symposium's speakers

Artem Besedin	email: artbesedin@hotmail.com
Damir Čičić	email: damircicic@gmail.com
Anton Kuznetsov	email: anton.smith@philos.msu.ru
Sergei Levin and Mirko Farina	email: serg.m.levin@gmail.com
Andrew Mertsalov	email: a.mertsalov@hardproblem.ru
Maria Sekatskaya and Gerhard Schurz	email: maria.sekatskaya@gmail.com
Louis Vervoort	email: l.vervoort@utmn.ru

****

The abstracts of the symposium's contributions

Artem Besedin

Implicit Attitudes: A Challenge to Agency and Moral responsibility

One challenge to our theories of agency and moral responsibility from the cognitive science and psychology is the problem of implicit attitudes mainly discussed in the form of the problem of implicit bias
Some experiments show that people who are not explicitly sexist or racist, in certain experimental conditions, are demonstrate discriminatory behavior. The features of such implicitly biased behavior are automatism, lack of consciousness and control over the course of action. Uhlmann and Cohen (2005) have shown that implicitly biased agents often are under the "illusion of objectivity": they tailor the criteria of choice according to their preferences and cite them as an explanation of the action. The experiments of Norton, Vandello and Darley (2004) that implicit bias can result in the distortions in memory. Payne (2006) demonstrates that "bias can coexist with a conscious intention to be fair and unbiased": even if subjects were told before the experiment to avoid biased actions, they continued to perform them.
It is not clear, to what class of cognitive phenomena the implicit attitudes must be attributed: whether they are something different from beliefs (Gendler's "aliefs" (Gendler 2008a, 2008b)), a species of beliefs (Schwitzgebel 2010), or sui generis phenomena (Levy's "patchy endorsements" (Levy 2015)). The choice is important for the relation between implicit attitudes and agency.
One possible conclusion of the analysis of these cases is skepticism about free will and moral responsibility. The features of implicitly biased behavior, as Levy suggests (Levy 2017), are incompatible with free will and moral responsibility. Implicit bias is one kind of implicit attitudes. The arguments concerning this type of behavior can be generalized to a broader class of actions. Another option is an optimistic position, defended by Vargas: the implicit bias cases can stimulate us to develop a better theory of agency and moral responsibility.

References

1. Gendler, T. S. (2008a). Alief in action (and reaction). Mind & Language, 23(5), 552–585.
2. Gendler, T. S. (2008b). Alief and Belief. Journal of Philosophy 105(10), 634-663.
3. Levy, N. (2015). Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsement. NOUS, 49(4), 800–823.
4. Levy, N. (2017). Implicit Bias and Moral Responsibility: Probing the Data. Philosophy and Phenomenological Research. 94(1), 3-26.
5. Norton, M. I., Joseph A. V., and John M. D. (2004). Casuistry and social category bias. Journal of Personality and Social Psychology, 87(6), 817–831.
6. Payne, B. K. (2006). Weapon Bias: Split-Second Decisions and Unintended Stereotyping. Current Directions in Psychological Science, 15(6), 287–291.
7. Schwitzgebel, E. (2010). Acting Contrary to Our Professed Beliefs, or The Gulf Between Occurrent Judgment and Dispositional Belief. Pacific Philosophical Quarterly, 91, 531-553.
8. Uhlmann, E. L. & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. Psychological Science, 16(6), 474-480.
9. Vargas M. (2017). Implicit Bias, Responsibility, and Moral Ecology. in: D. Shoemaker, ed., Oxford Studies in Agency and Responsibility. Oxford University Press.

Damir Čičić

Two Accounts of the Problem of Enhanced Control

According event-causal libertarianism, an action is free in the sense relevant to moral responsibility when it is caused indeterministically by an agent' beliefs, desires, intentions, or by their occurrences. This paper is an attempt to understand one of the major objections to this theory: the objection that it cannot explain the relevance of indeterminism to this kind of freedom (known as free will). Christopher Evan Franklin [2011, 2018] has argued that the problem of explaining the relevance of indeterminism to free will (which he calls the 'problem of enhanced control') arises because it is difficult to see how indeterminism might enhance our abilities, and disappears when we realize that beside the relevant abilities free will involves opportunities. In this paper, I argue that the problem occurs not because of the focus on abilities, but because it is hard to see how indeterminism might contribute to the satisfaction of the sourcehood condition of free will (in the framework of event-causal theory of action).

Anton Kuznetsov

Existence of Consciousness and Integrated Information Theory of Consciousness.

There are number of empirical theories of consciousness. They are based on different philosophical assumptions. The global neuronal workspace theory (GNWT) is motivated by an illusionists' account of consciousness. Integrated information theory of consciousness (IITC) uses an axiom of intrinsic existence of consciousness or reality of consciousness. Contrary to GNWT this axiom of IITC is not supported by well-developed arguments. Instead, it is based on dualistic and panpsychist approaches to consciousness, which are themselves problematic.

I will argue for a way to defend the first axiom of IITC. The infallibility of introspection, the transcendental status of consciousness, the specific nature of phenomenal concepts are usual ways of defending the existence of consciousness. But they are problematic. Introspection could be inadequate, the specific nature of phenomenal concepts is faced by the strategy of phenomenal concept or relies on introspection, or incoherent the transcendental argument is a non-starter. I'll try to show it in details. The way out is to use the methodology of common sense. I think the existence of the physical in general relies on common sense. And in the same way common sense works for consciousness. If you reject the existence of the last one then you must reject the existence of the first one. If this approach would work then there is a basis for IITC.

Sergei Levin
Mirko Farina

How I Learned to Stop Worrying and Love Free Will

Abstract: Willusionists believe that science has proven that free will is an illusion. We propose a thought experiment demonstrating that it is possible to spot a difference between merely having an illusion and having some real ability. The experiment involving a fictional character Dr. Strangelove, who suffers from alien hand syndrome in the right hand. He asks another fictional character Mary, who is a world-renewed neuroscientist and a willusionist, to return his control over the movements of his right hand; thus, he asks Mary to return his free will. Mary thinks Strangelove confuses the free will with the illusion of free will and she can only return him the latter. For that, she performs brain surgery leaving intact real casual roots of the alien hand movements, but creating an illusion of control. Even though Strangelove may be artificially happy with the results, a third-party observer would notice that Strangelove has been tricked into believing that he is in control of the movements of his hand. If the difference between the illusion of control of his right hand and the way he controls the left hand is real and not imaginary, then it is misleading to call free will an illusion. We conclude that the model of free will proposed by willusionism is incomplete.

Andrew Mertsalov

The Relevance of the Interpretations of Quantum Mechanics to the Question of Free Will and Determinism

It is widely agreed that our best physical theories don't fit determinism precisely[5]; it is often claimed[6] that most of the interpretations of quantum mechanics (QM; Copenhagen, objective-collapse, etc.) can provide a strong basis to deny determinism. Nevertheless, determinism is still treated as the main threat to free will, and the question about their compatibility remains to be the central in the field. Both compatibilists and incompatibilists[7] provide a number of arguments to show that the refutation of determinism by QM should not be taken into account. I'll analyze some of these arguments to show that they are unsound.

It is frequently claimed[8] that QM doesn't require indeterminism: there are some powerful deterministic interpretations. But even if QM-indeterminism will turn out to be true, another argument states[9], it can be irrelevant to metaphysical causal determinism. One can insist that metaphysical theory differs from the determinism physics is supposed to refute, to the extent that they can be true or false independently. I'll compare them and show that in spite of many of their distinctions physical and metaphysical determinism can be reduced to a common conceptual core with mutual key theses, so the denial of some of these theses by indeterministic QM interpretations means the rejection of both theories. This analysis will help me to show further that deterministic QM interpretations (with non-contextual hidden variable) that fit that core theses are disproved experimentally[10], while deterministic interpretations that are not refuted (Bohm's, Everett) don't fit the core11 and so are irrelevant to metaphysical determinism. That means that the denial of determinism by QM should be treated seriously in the free will debates: it cannot prove the existence of free will but can alter some arguments and question-setting or even change the general mood of the discussion.

References

1. Aspect et al. 1981 – Aspect A. et al. Experimental Tests of Realistic Local Theories via Bell's Theorem. // Phys. Rev. Lett. 1981. № 47 (7). P. 460–463.
2. Bell 2004 – Bell J.S. Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy. 2nd ed. Cambridge: Cambridge University Press, 2004.
3. Berofsky 2012 – Berofsky B. Nature's Challenge to Free Will. Oxford: Oxford University Press, 2012.
4. Bohm, Hiley 1993 – Bohm D., Hiley B.J. The Undivided Universe: An Ontological Interpretation of Quantum Theory. L. and NY: Routledge, 1993.
5. Born 1969 – Born M. Atomic Physics. London: Blackie and Son, 1969.
6. Dennett 1984 – Dennett D.C. Elbow Room: The varieties of Free Will Worth Wanting. Cambridge, Mass.: MIT Press, 1984.
7. Dennett 2003 – Dennett D.C. Freedom Evolves. NY: Viking, 2003.
8. Earman 1986 – Earman J. A Primer on Determinism. Dordrecht: D. Reidel Publishing Company, 1986.

9. Everett 1973 – Everett H.III. The Theory of The Universal Wave Function / in Dewitt B., Graham N. (eds.). Many-Worlds Interpretation of Quantum Mechanics. Princeton, NJ: Princeton University Press, 1973.
10. Feynman, Hibbs 1965 – Feynman R.P., Hibbs A.R. Quantum Mechanics and Path Integrals. NY: McGraw Hill Book Company, 1965.
11. Feynman, Leighton, Sands 2010 – Feynman R. P., Leighton R.B., Sands M. The Feynman Lectures on Physics. Volume I: Mainly Mechanics, Radiation, and Heat. NY: Basic Books, 2010.

---

[5] See [Earman 1986].
[6] Both by classics of contemporary physics [Heisenberg 1958, 128–146; Born 1969, 94–102; Feynman, Hibbs 1965, 2–21; Feynman, Leighton, Sands 2010, 38-15–38-18; Penrose 1989; Блохинцев 1966, 5–21, 44–45, 133–147; Мякишев 1973, 171–173, 223–231; Hawking 1988, 53–61, 172–174] and by modern scientists [Lloyd 2007, 99; Gisin 2014; Kaku 2005, 153–160].
[7] See, e.g., [Dennett 1984, 2003; Kane 2001; Honderich 2002; Pereboom 2003].
[8] E.g., [Kane 2001, 9; Honderich 2002, 463–466; Pereboom 2003, xviii].
[9] Different variations of this argument can be found in [Dennett 1986, 77, 101–130; 2003, 36–62; Honderich 2002, 463–464; Berofsky 2012, 9–12].
[10] See, e.g., [Bell 2004; Norsen 2006; Aspect et al. 1981; Pan et al. 2000; Hensen et al. 2015].
[11] I'll rely on [Bohm, Hiley 1993; Valentini 1991; Maudlin 2011; Riggs 2009; Everett 1973; Norsen 2007], etc.

Maria Sekatskaya,
Gerhard Schurz

Alternative Possibilities and the Meaning of 'Can'

Abstract: We will provide a new naturalistic account of free will which incorporates the criteria of sourcehood with the criteria of availability of alternative possibilities. Our account combines conditional analysis of abilities with a Frankfurt-style sourcehood psychological approach and has three advantages: 1) it answers the objections against standard versions of classical conditional analysis of 'can' by demanding coherence of what one can freely do with one's personality frame; 2) it is compatible with both determinism and indeterminism as metaphysical background assumptions; 3) it is immune to the Consequence Argument and solves the Luck Problem.

Louis Vervoort

Artificial Consciousness and Superintelligence in Robotics: How to Get There?

Abstract: Can future robots and AI-systems have consciousness and genuinely human intelligence – or even better, superhuman intelligence ? Here I look at these questions from the point of view of philosophy of science and of mind, and argue that these questions are related: their answer hinges on the fulfillment of the same condition. Starting from an analysis of the concepts of consciousness and free will – here philosophy has a much longer tradition than other disciplines – I argue that the key capacity that computers and robots should possess in order to emulate human cognition and consciousness is the capacity to learn and apply theories. I review selected literature in cognitive neuroscience and computer science related to AI to back-up this claim. I will also consider rival theories, thus opening up the debate.

****

Symposium: Philosophy Meets Psychology of Science

Vlasta Sikimic, Kaja Damjanovic, Marko Tesic, Nora Hangel and Tijana Nikitovic

Symposium: Philosophy Meets Psychology of Science

Section: History, Philosophy and Social Studies of Science

Participants (in order of presenting):

Kaja Damnjanović
(Faculty of Philosophy, University of Belgrade, kdamnjan@f.bg.ac.rs)

Marko Tešić
(Birkbeck College, University of London, mtesic02@mail.bbk.ac.uk)

Nora Hangel
(Technical University Munich, nora.hangel@tmu.de)

Tijana Nikitović
(Faculty of Philosophy, University of Belgrade, tijana.nikitovic@f.bg.ac.rs)

Vlasta Sikimić
(Weizsäcker Center, University of Tübingen, vlasta.sikimic@uni-tuebingen.de)

Symposium Description

Philosophy of science profits from empirical insights about psychological dispositions, such as the attitudes of scientists, their reasoning, etc., since they enrich the theoretical understanding of scientific practices. Psychology of science, on the other hand, has explored questions concerning the nature of science or attitudes towards pseudoscientific claims, yet without putting sufficient attention to the philosophical questions in the background. We argue that questions such as how to improve scientific reasoning both on an individual and collective level can be best answered by combining insights from both disciplines. Thus, the purpose of the symposium is to bring together researchers working in both fields, let them present their interdisciplinary research, and engage in a discussion about their methodologies.

There are various fields of philosophy of science that can profit from empirical insights. In our symposium, we will limit ourselves to the empirical approaches in social epistemology of science and to philosophically driven questions in cognitive psychology. Social epistemology of science investigates optimal ways of group knowledge acquisition in science. Understanding the cognitive processes behind the epistemic decisions of scientists, work conditions in science, and values of scientists from the empirical perspective is beneficial for deepening the understanding of how to maximize scientific knowledge on the group level. Thus, the first goal of the symposium is to discuss ways of improving scientific reasoning both on the individual and on the group level.

The second aim of the symposium is to present different methods from psychology and their interaction with the traditional philosophical analysis. The methods that participants used in their research encompass qualitative interviews, quantitative surveys, and psychological experiments. In particular, Kaja Damnjanović will discuss cognitive biases in the scientific context and among scientists, as well as ways of empirically investigate them. Marko Tešić will explore a pattern of causal inference called 'explaining away', prevalent in both scientific and everyday contexts, and present experimental results on how different probability interpretation can drive our explaining away reasoning. Nora Hangel will provide theoretical arguments in favour of empirically-driven social epistemology and present empirical results from a qualitative survey with more than 60 scientists about their collaborative belief formation. Tijana Nikitović will present results from a qualitative survey on existential challenges that early-career researchers face when working abroad, with the conclusion that overcoming these challenges would be epistemically beneficial for the scientific community as a whole. Finally, Vlasta Sikimić will discuss how the socio-political viewpoints of scientists influence their epistemic decisions, based on the results of a quantitative survey on 655 researchers.

Significance

The aim of the symposium is to address the fundamental question of how to optimize scientific reasoning, which lies in the intersection of philosophy and psychology of science. This will be done by analysing some of the cognitive inclinations of scientists, epistemic attitudes, belief formation, and social factors surrounding researchers. Ideally, this interdisciplinary symposium should engage a wide range of audiences. From the perspective of psychology, the theoretical questions from philosophy of science are enriching. Notions such as epistemic tolerance and epistemic authoritarianism, and other epistemic virtues and vices, did not yet get sufficient attention in psychology. The benefit for the more philosophically inclined audience is understanding what actually happens in science on the epistemic level, how researchers form their beliefs, and what influences their epistemic judgment in practice.

Finally, the symposium will give the opportunity both to the speakers and to the attending participants to exchange their knowledge about different methodologies from philosophy and psychology of science, and discuss their potential interactions.

Simplicity bias or Ockham's razor?

Kaja Damnjanović

(Joint work with Ivana Janković, Mateja Manojlović, Tijana Niktiović and Vlasta Sikimić)

It has been commonly argued that different biases negatively affect the epistemic performance of scientists. For instance, confirmation bias is the ungrounded and reinforced belief in positive evidence and it is proposed that it leads to the publication of insufficiently supported results. However, from the perspective of the epistemic community and psychological approaches to rationality, cognitive biases can under certain conditions be advantageous, since the full heuristic power of a cognitive mechanism cannot be assessed in isolation (e.g., Goldstein and Gigerenzer, 2002).

Simplicity bias reflects our preference for more straightforward explanations. In the background of this bias is our need to understand and explain processes in nature and society in a way that we can comprehend easily. However, contemporary science is full of complex explanations. Thus, we raise a legitimate question of what the role of the biased reasoning, specifically phenomenon of "simplicity bias" of scientists is.

Traditionally, simplicity has been understood as one of the virtues of a scientific hypothesis. Johnson et al. (2019) following Lombrozo (2006) empirically showed that simplicity bias works in accordance with Bayesian reasoning: when the prior probabilities change, the strength of one's preference towards simpler explanations decreases. Moreover, scientists, when faced with a dilemma, might opt for a different virtue of a hypothesis, e.g., they will most likely prefer a more comprehensive explanation over a simpler one.

In the talk, we will analyze the role of the simplicity bias in scientists based on the previous empirical results. We argue that it does not represent a real epistemic danger for the scientific community, in terms that the occurrence of the simplicity bias will not necessarily deteriorate scientific findings. In order to provide empirical support for this claim, we propose an operationalization of the trade-off between simplicity and comprehensiveness of hypotheses in the scientific context, which will be experimentally investigated.

Explaining away and the propensity interpretation of probability

Marko Tešić

Explaining away is a pattern of causal inference that is omnipresent in both everyday situations and in scientific contexts. It captures situations where two independent causes (e.g. flu and asthma) can each on their own bring about an effect (e.g. a cough). Knowing that the effect (a cough) is present, subsequently learning that one of the causes obtains (e.g. the patient has an asthma) would reduce the probability of the competing cause (i.e. the patient has a flu), even though the two causes were initially independent. Despite being widely addressed in causal reasoning literature, explaining away remains highly elusive.

Several potential explanations of this insufficiency have been put forward thus far. At present, we explore the novel possibility that the observed insufficiency in explaining may be driven by the way the probabilities are interpreted. More specifically, we test for the possibility that some people interpret probabilities as propensities, that is, as tendencies of a physical system to produce a certain outcome. Studies on explaining away typically assume the subjective interpretation of probabilities. If, however, some people interpret probabilities as propensities, their probabilistic judgments of a cause may stay invariant to the presence/absence of either the effect or the other cause. This then can have consequences for approaches that try to capture causal reasoning in purely subjectivist terms.

The propensity hypothesis was tested by varying (i) the characteristics of scenarios so that the degree by which probabilities could be interpreted as propensities differed and (ii) the prior probabilities of the causes. Results suggested an overall insufficiency of explaining away. However, in line with the propensity hypothesis, a large cluster of participants did not revise their causal judgements. Further results are discussed, as well as their implications in relation to the previous literature.

How do scientists generate scientific claims? Individual, collaborative, and collective accounts from scientific practice

Nora Hangel

Philosophical topics about the nature of scientific knowledge, scientific belief formation, reliability and scope of reconstructed justifications in published articles, epistemic trust and dependence remain of utmost relevance within and outside the field of philosophy of science. For some decades, a naturalized philosophy of science has used science as a resource to inform philosophical questions.

By accessing the microstructure of belief formation processes as described by collaborating experimental scientists, we find a plurality of motives of what drives scientists, collaborating groups and science as a whole. For once, scientific reasoning processes, which accumulate to scientifically justified belief formation are less dependent on individual agents as one might expect. The main motives for individual scientists span between the desire for meaningful contributions to drive scientific process, the requirement of accountability for the reliability of these contributions, and the need for recognition.

I argue to take into account scientists' conceptions about their challenges in collaborative experimental practice: challenges about epistemic dependence and trust that point to an interdependence between the social organization of research groups and the epistemic aim to generate scientifically justified belief collaboratively.

The proposed qualitative analysis adds scientists' conceptions about belief formation processes. The approach is a naturalist social epistemology based on the analysis of conceptions by more than 60 working scientists reflecting on processes of scientific reasoning in collaborative experimental sciences. Those accounts help us to understand how on the one hand the social organization of the research group is utilized for belief forming processes. On the other hand, we learn about actual challenges of working scientists. Even if qualitative analysis does not claim a generalizable picture compared to quantitative research, it includes relevant information about how actual collaborative scientific belief formation develops into reconstructed justifications published in peer reviewed articles.

Challenges of scientific mobility facing early-career life scientists

Tijana Nikitović

(Joint work with Vlasta Sikimić)

From the perspective of social epistemology of science, it has been extensively argued that epistemic diversity is beneficial for the academic community (e.g. Zollman 2010). On the other hand, epistemic performance can be impacted by challenges researchers face in academia and their private lives, especially when doing their research abroad. Given the emphasis on the epistemic benefit of mobility in contemporary science, our study aims at exploring the driving forces behind scientific mobility as well as the main existential challenges that early-career researchers face. We performed semi-structured interviews with seven researchers (five non-EU and two EU citizens) about their experience of academic mobility in the EU. Four participants were women and all were in their early thirties. We used thematic analysis as an analytic method and searched for themes and patterns within and across the interviews. The three main sources of challenges that researchers face are explicit rules (e.g., necessary documentation, limited contracts and positions) and implicit rules of academia (e.g., job insecurity, pressure to publish, overtime, maternity leave), as well as challenges related to adaptation to a new socio-cultural environment (e.g., language barrier, interpersonal conflicts). The two main driving forces were positive motivation to learn and collect new experiences and negative motivation of escaping difficult conditions in the home country. Negative motivation was present solely among non-EU participants who also experienced more difficulty in adaptation to the new environment. Participants notably expressed worry for their professional and personal prospects. Finally, some participants even reported discrimination at the work place. We conclude that in order to achieve epistemic benefit for the group, it is first necessary to provide fair work conditions, i.e. adequate pay and longer-term contracts. Moreover, it is necessary to provide institutional support and to integrate foreign researchers in order to counteract the negative aspects of scientific mobility.

Relationship between political and epistemic values of scientists

Vlasta Sikimić

(Joint work with Tijana Nikitović, Miljan Vasić and Vanja Subotić)

One of the contemporary questions in social epistemology of science is how two scientists believing in opposing theories can still communicate and accept that their opponent might have a good point. The phenomenon explaining the openness for opposing opinions in science is called epistemic tolerance (Straßer et al. 2015). The other phenomenon relevant for the epistemic decisions of scientists is their epistemic authoritarianism. By this we mean the inclination of scientists to epistemically follow dominant paradigms and approaches in their field. Both the notion of epistemic tolerance and authoritarianism are inspired by the political discourse. In our research, we were interested to understand what the actual relationship between socio-political views of researchers and their epistemic attitudes is.

For this purpose, we proposed three scales: for epistemic tolerance, for epistemic authoritarianism, and for detecting skepticism with respect to the scientific method in general. On the other hand, we also measured the level of conservativism of our participants. Our sample (N=655, 352 females) contained both researchers from natural (48.9%) and social (51.1%) sciences. We noticed that scientists in general score high on the epistemic tolerance scale. Consequently, they scored low on epistemic authoritarianism and skepticism about the scientific method. When comparing social to natural sciences, social scientists scored on average slightly higher on the epistemic tolerance scale. We detected that socio-political conservativism is a negative predictor of epistemic tolerance, but the effect was small. Skepticism in the scientific method on the other hand correlated with beliefs in pseudoscience, i.e., astrology. However, skepticism in the scientific method did not correlate with epistemic tolerance. This agrees with the assumption that one can be epistemically tolerant for the opinions of others while still holding own views and the scientific paradigm true.

Symposium: Trends in Formal Philosophy of Science

Vlasta Sikimic, Jan Sprenger, Borut Trpin and Sandro Radovanovic

Symposium: Trends in Formal Philosophy of Science

Section: Formal Philosophy of Science

Talks and Speakers (in order of presenting):

1. **Jan Sprenger**
Center for Logic, Language and Cognition (LLC), University of Turin,
jan.sprenger@unito.it

2. **Borut Trpin**
Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich,
Borut.Trpin@lrz.uni-muenchen.de

3. **Sandro Radovanović**
Faculty of Organizational Sciences, University of Belgrade,
sandro.radovanovic@fon.bg.ac.rs

4. **Vlasta Sikimić**
Carl Friedrich von Weizsäcker Center, University of Tübingen
vlasta.sikimic@uni-tuebingen.de

Description

In formal philosophy of science, different methods are employed in order to optimize scientific knowledge acquisition and to understand how science works. The idea behind formal tools is to provide general explanations for a phenomenon of interest. The choice of such a model is, however, often a non-trivial decision that depends on the questions one wants to answer. For example, highly idealized models are able to provide how-possibly explanations, while empirical calibration is required if one wants to influence scientific practice directly, i.e., to draw conclusions about epistemically optimal practices. Thus, contemporary philosophy of science uses the whole array of different formal approaches, such as agent-based modelling, Bayesian reasoning, data mining techniques, etc. It is less clear, however, how these approaches cohere, how they can be combined, and whether there is something like a (*ceteris paribus*) preferred approach.

In our symposium, all the listed approaches will be represented and discussed, with special attention to their interplay. In particular, we will address the questions of the contemporary role of formal philosophy of science and discuss the mixed-method approach as a prospect for the future development of the subfield. The following three questions provide a guideline through the symposium:

1) What are the "hot" topics and methods in formal philosophy of science?
2) Why, and under which circumstances, are certain formal methods preferred in opposition to others?
3) What is the potential, and what are the limitations of formal methods in philosophy of science? Specifically, does formal philosophy of science use the same methods as science itself, or are there important differences?

Jan Sprenger will open the symposium with a talk that contrasts the role of formal methods in contemporary philosophy of science with the historical origins of formal philosophy of science in logical positivism. Using case studies from recent work, he advocates a mix of formal, conceptual, and empirical methods. Borut Trpin will discuss the famous Variety of Evidence Thesis (VET) in the context of Bayesian epistemology. Moreover, he will provide formal arguments in favour of the epistemic modesty principle, i.e., the principle of withholding the judgment when confronted with strong peer disagreement, which is a novelty in the VET discussion. His work thus combines tools from formal philosophy of science with principles that are discussed in the mainstream epistemology literature. Andrea Berber and Sandro Radovanović will present the results of a study of team structures in science obtained by a data-mining technique. The results of this study were based on citation metrics of the experiments conducted in the high energy physics laboratory, Fermilab, and are an example of carefully conducted data-driven research in formal philosophy of science. Finally, Vlasta Sikimić will discuss the benefits and limitations of data-driven and mixed methods, arguing that a citation-based approach to assessing scientific success can only be applied under a specific set of preconditions. She will also argue that the appropriateness of different formal approaches that are based on data is field-dependent and should be determined on a case-by-case basis.

Significance

The symposium is dedicated to formal philosophy of science and should inform the audience about the tendencies and debates in the subfield. Moreover, the participants will discuss the role of formal approaches in contemporary philosophy of science and the potential of mixed methodologies for providing more comprehensive answers to philosophical questions. For instance, conclusions of empirically informed agent-based modelling are several steps closer to the application in practice than the results of purely abstract modelling. Finally, the attending members of the wider philosophical audience will have the opportunity to expand their knowledge about different approaches in formal philosophy of science, its scope, and its limitations.

Formal Methods and Scientific Philosophy

Jan Sprenger (Center for Logic, Language and Cognition (LLC), University of Turin)

Formal methods in philosophy of science have suffered a setback following the naturalistic turn in the 1960s and 1970s. Yet, they have been resurgent over the last decades: there is an ever increasing number of papers making use of methods and models from Bayesian inference, machine learning, evolutionary game theory and other formal theories. These techniques are used for tackling and solving central problems in philosophy of science. Does this mean that almost 100 years after Carnap's and Reichenbach's programmatic writings on "scientific philosophy", our current philosophy of science qualifies as such?

In answering this question, this talk combines historical and systematic perspectives: First, I present various possible ways of making sense of the label "scientific philosophy", which leads to quite different views on the role of philosophy within the scientific enterprise. Second, I connect these views to historically held positions by proponents of the Vienna Circle and their intellectual surroundings (Popper, Reichenbach). Third, I argue that current philosophy of science should not align with either of these viewpoints, but rather adopt a clever mix of them: (a) to consider philosophical research as a proper part of the scientific enterprise rather than as a meta-science or as necessary prolegomena, (b) to reject a purely "mathematical philosophy" in favor of a mix of formal, conceptual and empirical methods, (c) to adopt the method of explication as being central for classical problems in philosophy of science, such as giving a good theory of explanation, confirmation, or causation. (The explicative method has obvious limits, of course, for example for research at the science-policy interface.)

I will then illustrate how complying with these three requirements allows Bayesian models to achieve substantial progress on philosophical problems. After this case study, I draw the conclusion that good scientific philosophy uses—by and large—the same methods as science, and that there is a continuum between the goals and methods of proper science and those of scientific philosophy.

Methodological triangulation and the value of epistemic modesty

Borut Trpin (LMU Munich) & Mariangela Zoe Cocchiaro (University of Hong Kong)

It intuitively seems that evidence in favour of *h* provides better support for it if it is obtained by multiple different methods rather than if the same evidence is obtained multiple times by the same method. This view has been called the Variety of Evidence Thesis (VET). Surprisingly, when we model these situations in a Bayesian framework, the VET turns out to be false or at least limited in scope (e.g., Bovens and Hartmann, 2003). We focus on a special variant of VET, the so-called *Du Boisian methodological triangulation* (DMT) that has recently been vindicated due to being a better guide to truth than following a single method, given that we are not sure which method is the most reliable (Heesen et al., 2019).

DMT, however, is based on a non-convincing assumption: when it is not possible to triangulate on a single result provided by multiple methods, that is, when evidence is highly discordant, the method prescribes us to randomly endorse one of the provided results. This step is crucial for its justification: without it, the method fails to endorse the results that are more likely true than results obtained by randomly choosing a single method and sticking with it.

We suggest that a triangulator should not endorse any of the results when evidence is unclear and call this approach *epistemically modest triangulation* (EMT). We justify this step on the basis of the insights from the epistemology of disagreements: in case two epistemic peers disagree about *p*, it is rational for them to reduce their confidence in it (e.g., Christensen, 2007). Furthermore, randomly endorsing one of the results, like DMT prescribes, provides a bad ground for epistemic trust in science (Wilholt, 2013).

We then show that if withholding judgment is not considered to be negative, then the expected utility of EMT is greater than that of DMT and that of randomly settling upon a single method. We then discuss the implications of our method for some common types of scientific disagreements (e.g., failed replications and the discussion on the benefits of direct vs. indirect replications), and for the distinction between epistemic and methodological disagreements in science (Šešelja, 2019).

References:

Bovens, L., & Hartmann, S. (2003). Bayesian epistemology. Clarendon Press.
Christensen, D. (2007). Epistemology of disagreement: The good news. Philosophical Review, 116, 187–218.
Heesen, R., Bright, L. K., & Zucker, A. (2019). Vindicating methodological triangulation. Synthese, 196(8), 3067-3081.
Šešelja, D. (2019). Some lessons from simulations of scientific disagreements. Synthese, 1-16.
Wilholt, T. (2013). Epistemic trust in science. The British Journal for the Philosophy of Science, 64(2), 233-253.

Using machine learning algorithms to predict the efficiency of experiments in high energy physics

Sandro Radovanović (University of Belgrade, Faculty of Organizational Sciences)
& Andrea Berber (University of Belgrade, Institute for Philosophy, Faculty of Philosophy)

The social epistemology of science raises the question of how to structure a team in order to maximize its epistemic effectiveness. We will present a study that addresses this question using machine learning algorithms on data from Fermilab, one of the world's largest high energy physics laboratories. Our method has two steps: 1. The analysis of data concerning the structure and outcomes of experiments conducted at Fermilab using Data Envelopment Analysis (DEA). Duration, team number and team size were used as input variables, and the number of citations per paper divided into six weighted categories as output variables. In this step, we obtained the efficiency score for each analyzed experiment. 2. We conducted a predictive analysis, using efficiency scores obtained in the previous step, in order to discern beforehand whether the experiment will be epistemically efficient. We used different machine learning technics such as decision trees, gradient boosted tree algorithms and neural networks. This step enables us to predict efficiency based on relevant numerically expressible data for each experiment. Additionally, predicting project's efficiency with respect to each of the numerically expressed parameters should provide guidelines for the optimization of projects. After presenting our results, we will discuss specific practical instructions on how we should structure a team to maximize epistemic efficiency of a group in accordance with our study, and how these instructions can improve science policy and benefit scientists working in the field.

Benefits and limitations of data-driven approaches in formal philosophy of science

Vlasta Sikimić (University of Tübingen)

Social epistemology analyzes epistemic groups with respect to knowledge acquisition both by experts and laymen. It employs tools from philosophy, psychology and computer science. Computer simulations are used to model group knowledge acquisition. They can provide how-possibly explanations of the phenomena in question, while empirical approaches give a more realistic take on the questions of interest. When the models become empirically calibrated the synergy of these approaches is called mixed-methodology. We will argue in favor of the mixed-methodology, since it brings us closer to the application of philosophical findings in practice. Thus, it increases the role and importance of philosophy for science in general.

Different data mining techniques are an important component of the mixed-methodology. They are powerful tools for addressing optimization questions in science. Still, they should be used with caution, i.e., with a clear understanding of their reach. In particular, data-driven approaches should be field-specific, because data interpretation varies across disciplines. For instance, while the consensus about results in high energy physics (HEP) is relatively quick and reliable, this is not the case in fields such as contemporary experimental biology. The research in HEP has a regular underlying inductive behavior which postulates the conservation principle as the core one. This inductive behavior results in a reliable pursuit (Perović and Sikimić 2019, Schulte 2000). Thus, the research in HEP is suitable for data-drive analyses based on the citation metrics. On the other hand, in experimental biology the consensus about results is generally neither fast nor reliable (Pusztai et al. 2013). The time scale of reaching the consensus is much longer than in HEP. It is often difficult to find a coherent set of inductive rules governing the research in biology. Thus, we will argue that for adequately addressing the epistemic optimization in experimental biology one requires mixed methodology of a different type, e.g., empirically informed agent-based modelling.

The EENPS 2021 Organizers

Program committee

Co-chairs:

- Barbara Osimani (Università Politecnica delle Marche)
- Borut Trpin (MCMP, LMU Munich)

Members:

- Lukas Bielik (Faculty of Arts, Comenius University in Bratislava)
- Lorenzo Casini (University of Geneva)
- Milan Ćirković (Astronomical Observatory, Belgrade)
- Vincenzo Crupi (Department of Philosophy - University of Turin)
- Richard David-Rus (Institute of Anthropology 'Francisc I Rainer', Romanian Academy)
- Andreea Esanu (Alexandru Dragomir - Institute for Philosophy)
- Aldo Filomeno (Pontificia Universidad Católica de Valparaíso)
- Monika Foltyn-Zarychta (University of Economics in Katowice)
- Lilia Gurova (New Bulgarian University)
- Stephan Hartmann (Ludwig Maximilian University of Munich)
- Catherine Herfeld (University of Zurich)
- Gábor Hofer-Szabó (Institute of Philosophy, Hungarian Academy of Sciences)
- Daniel Kostić (Université Bordeaux Montaigne, SPH - sciences, philosophie, humanité - EA 4574, France)
- Philippe Huneman (IHPST (CNRS/ Université Paris I Sorbonne))
- Maria Kronfeldner (Central European University)
- Sebastian Lutz (Uppsala University)
- Edouard Machery (University of Pittsburgh)

- Olga Markič (University of Ljubljana)
- Tomáš Marvan (Institute of Philosophy, Czech Academy of Sciences)
- Marcin Miłkowski (Institute of Philosophy and Sociology, Polish Academy of Sciences)
- Patricia Palacios (University of Salzburg)
- Slobodan Perović (University of Belgrade)
- Stathis Psillos (University of Athens)
- Julian Reiss (Durham University)
- Rush Stewart (Ludwig Maximilian University of Munich)
- Elena Trufanova (Institute of Philosophy, Russian Academy of Sciences)
- Giovanni Valente (Politecnico di Milano)
- Özlem Yilmaz (Istanbul Technical University)

Local organizing committee

- Duško Prelević (Chair)
- Drago Đurić
- Slobodan Kanjevac
- Vlasta Sikimić
- Danijela Marković
- Andrea Berber

Acknowledgement

The EENPS 2021 Belgrade conference is supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (project: Logico-Epistemological Bases of Science and Metaphysics, No. 179067 at the Faculty of Philosophy, University of Belgrade).

Author Index