Philosophy of Experiment

Stockholm, 16.-17. November 2023

Abstracts

Sarwar Ahmed

Best explanation for the source of the information: The role of IBE in modern observations

The role of underdetermination appears to be opposite to that of theory-ladenness. Specifically, if an observation is influenced by the theory being tested, how can multiple theoretical explanations equally account for the observation? In this talk, I argue that both of these concepts play a role in the observational process itself. More precisely, I propose that a specific type of underdetermination may arise in conjunction with the complexity of the process and scientific and technological limitations, alongside theory-ladenness in the observational situation. Given that inference to the best explanation (IBE) is pervasive in science (Douven 2022), I posit that this form of underdetermination is resolved by relying on IBE. Making an observation involves determining the observed entity and its properties, which makes IBE superior to other types of inferences for discerning the source of information in an observational setting.

Modern scientific observations are intricate, non-individualistic, non-perceptual, and heavily influenced by theory. It's likely that the information gathered in these observational processes is inconclusive regarding its source. If the source needs to be modelled, then this inconclusiveness may arise partly due to viable alternatives to the standard model (theory) of the source. The degree of uncertainty in source models affects how informative the gathered evidence about the source is. This gives rise to a particular type of underdetermination I term "underdetermination of source by information," which becomes pertinent when inferring the source of the information.

For instance, the LIGO-Virgo collaboration inferred in an observation that the source of gravitational waves is a binary black hole merger (Abbott, B. P., et al. 2016a). However, this inference relies on a substantial amount of background theoretical and empirical knowledge. Furthermore, estimating the properties of the gravitational wave source heavily relies on models representing the phenomenon and the fundamental theory predicting it. Specifically, models like numerical relativity, post-Newtonian approximation, and black hole perturbation theory are employed in the observational process. Additionally, Jamee Elder (2023) contends that these models are assumed to be valid during observation.

This assumed validity of the models, compared to a scenario where they are validated, can introduce a form of theory-ladenness I term "strong theory-ladenness." In observations characterized by strong theory-ladenness, IBE plays a prominent role. To illustrate this, I emphasize its significance in observing binary black hole systems through gravitational waves as a case study.

References:

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

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

Elder, Jamee (2023), “On the “Direct Detection” of Gravitational Waves” (Under review), https://www.jameeelder.com/uploads/1/2/1/6/121663585/jan2023_direct_detection_elder.pdf

Hugo Beauchemin and Kent Staley

What is empirical? A consideration from the triad of measurement, uncertainty, and sensitivity

Philosophers of science have recently rediscovered measurement as an important epistemological issue in philosophy of science. The trend has been to focus first on the characterization of measurement itself, while treating measurement *uncertainty* as a secondary issue of how to characterize the quality of a measurement process or its product. Considerations of *sensitivity* have gone largely unarticulated, while lurking beneath discussions about measurement objectives.

We think this approach misses an opportunity for a better understanding of a constellation of experimental practices surrounding the production and use of measurement results. On our approach, the concepts of *measurement, uncertainty,* and *sensitivity* form a triad, no single element of which can be teased apart and understood prior to and independently of the others. To put the point in rough terms: Measurement refers to a process of producing a result, yielding an uncertainty that is integral to that result. The uses of that result, including its uncertainty, in subsequent inquiries are guided by its sensitivity to the targets of such inquiries. Because the process is an intentional one, the features of the process are chosen with reference to its capacity to produce a result with particular uncertainty characteristics, enabling it have sensitivity with respect to specific anticipated uses.

This production process can be thought of as a process of *contextualization* insofar as the measurement result, characterized by an uncertainty, bearing sensitivity for certain purposes, possesses features imparted to it by the context of its production and use. These same characteristics, however, enable the very same result to be partially *decontextualized*. The process by which uncertainty is evaluated aims to produce an attribute of the measurement result that retains its relevance and meaning independently of an ability to describe the measurement process or the sources of its uncertainty, and that has the capacity to be used in a manner not anticipated in the planning of the measurement.

The contextualization that integrates measurement, uncertainty, and sensitivity also incorporates a variety of resources, including theory, simulation, heuristics, rules of thumb, and practical considerations. As a consequence, the presupposition that measurement results constitute an "empirical core" of scientific inquiry in contrast to these "non-empirical" or "quasi-empirical" elements becomes indefensible. That, in turn, provides a new perspective on questions about the confirmation, truth content, or empirical status, of physical theories.

Florian Boge

Deep Learning for Scientific Discovery and the Theory Freedom-Robustness Trade-Off

Machine Learning (ML) systems called Deep Neural Networks (DNNs) are of great promise in science today, in disciplines ranging from social science to High-Energy Physics (HEP). However, they are also subject to astonishing shortcomings: 'Adversarial examples', e.g., are data instances easily classifiable for humans but totally misclassified by DNNs. In HEP, DNNs are expected to foster scientific discovery through the detection of anomalies without reliance on any specific theory or model and here, adversarial vulnerability turns out to be a double-edged sword: On the one hand, it shows that discerning DNNs' credible outputs from flukes requires some skill. But on the other hand, adversarials exhibit DNNs' sensitivity to subtle, often human-inscrutable features that could also be scientifically productive (Buckner 2020) and are, in fact, utilised in anomaly detection. However, as it also turns out, there are other subtle dependencies that scientists should be aware of if DNNs are supposed to foster scientific success. We here offer an analysis of, and a cautionary tale about, DNNs' present utility for scientific discovery in HEP. We will (i) establish adversarials' role in anomaly detection and the great promise for discovery associated with this; (b) introduce a notion of performance-robustness, which DNNs need to satisfy in order to be able to deliver genuine discoveries; and (c) argue that there is a trade-off between performance-robustness and theory-freedom, which spoils overly enthusiastic appraisals of ML driven discovery.

Jamee Elder

Theory Testing in Gravitational-wave Astrophysics

Since 2015, the LIGO-Virgo Collaboration has been observing binary black hole mergers via gravitational waves. These observations are highly theory-laden, leading to a potentially-vicious circularity where general relativistic assumptions may serve to mask phenomena that are inconsistent with general relativity (GR). This talk examines several ways that the LIGO-Virgo observations are used in theory and hypothesis testing, despite this circularity problem. First, I argue that early tests of GR using GW150914 (the first detection) are best understood as a response to the threat of theory-ladenness and circularity. Each test places constraints on the extent to which LIGO-Virgo's theory-laden methods are biasing their overall conclusions. However, remaining degeneracies limit the extent to which these tests constrain deviations from GR. Second, these observations provide a basis for studying astrophysical and cosmological processes, especially through analyses of populations of events. As gravitational-wave astrophysics transitions from new to established science, constraints from early tests of GR provide a scaffolding for these population-based studies. Overall, this talk analyses the ways that theory and hypothesis testing operate in gravitational-wave astrophysics as it gains maturity. In particular, I show how these tests build on one another in an iterative manner in order to mitigate a circularity problem at the heart of the observations and to break degeneracies between changes in source parameters and dynamical theory; and between source properties and cosmology. The picture of that emerges has strong resonances with Hasok Chang's (2004) ``epistemic iteration'', with the field of gravitational-wave astrophysics bootstrapping its way to inferences about black hole populations and cosmology, despite the insecure, theory-laden foundations of inferences about single events.

Enno Fischer

The pursuitworthiness of experiments. An analysis of “no-lose” theorems.

Experiments at the frontiers of fundamental physics often involve expensive research facilities, big research collaborations, and they need to be planned over very long time periods. Therefore, decisions regarding the planning of such experiments can have a major impact on the development of research fields. Philosophers of science have developed various accounts of the epistemic pursuitworthiness of research programs. Unfortunately, most extant accounts of pursuitworthiness give little concrete guidance for scientific practice. Moreover, extant approaches typically concern the pursuit of theories rather than concrete experiments.

In my talk I explore first steps towards a systematic approach for evaluations of the pursuitworthiness of experiments. A starting point is the analysis of “no-lose” theorems. These are claims to the effect that—no matter what the result of an experiment is—there will be a significant and guaranteed epistemic gain. Typically, these claims are made regarding the potential discovery of new phenomena that are theoretically established but lack empirical confirmation. In such cases the potential discovery of the new phenomenon would be a clear epistemic gain, confirming the theoretical expectations. According to the “no-lose” framing, also a non-discovery would be an important outcome because it would have significant consequences for how the theoretical proposal is to be evaluated.

For example, before the discovery of the Higgs boson, physicists had very clear expectations that such a particle would be discovered. Moreover, physicists assumed that the less likely case of a non-discovery would also have major implications for our understanding of the Standard Model of particle physics. According to this reasoning, the search for the Higgs was worthwhile because of the high expectations of finding the Higgs. But even in the less likely case of a non-discovery the project would have brought a considerable epistemic gain.

While the Higgs search was a success, there are also instances in which “no-lose” reasoning has failed. One example is low-energy supersymmetry. Supersymmetric particles could not be confirmed to date despite high theoretical expectations. And, contrary to “no-lose” arguments that have been forwarded since the 1980s, it is not so clear what the epistemic gain of the non-discovery is or may be. Further no-lose arguments have been formulated regarding future experiments. For example, there are attempts to formulate no-lose arguments regarding the discovery of dark matter candidates at a proposed future circular collider (FCC) at CERN.

In the talk I will first analyze the examples with a particular focus on (1) the overall logic and (2) the kinds of theoretical assumptions underlying “no-lose” reasoning. Based on these considerations I will then explore the viability and relevance of such arguments: how much weight should physicists give to no-lose reasoning in the planning of large-scale experiments? On the one hand, such arguments may give a sense of guaranteed payoff. On the other hand, an overly strong focus on such theorems may unduly disregard the autonomy of experimental inquiry, such as in exploratory forms of experimentation.

Till Grüne-Yanoff

Simulations Are Not Experiments

Engineers routinely call certain computer simulations “experiments”. In this paper, I argue against this equivocation: for both conceptual and methodological reasons, one should distinguish experiments from computer simulations.

In conversation, engineers often report about an experiment, only to later reveal that they described a simulation in which some intervention was modelled. Peer-reviewed articles bear titles like "Finite Element 3D Simulation Experiment on Shield Excavation" or "Design and Simulation Experiment of Rigid-Flexible Soft Humanoid Finger". A cursory search of the term "simulation experiment" on google scholar reveals more than 260000 entries. Many scientists and engineers, it seems, do not draw a clear distinction between simulation modeling and experimentation practices.

There are of course important similarities between models and experiments. After all, modelers set variables and parameters in a model in a similar way as experimenters establish experimental control of background variables in an experiment. They also manipulate the model in ways akin to experimental intervention, and they are interested in observing what differences our manipulations make to the model result, similar to experimental observation of effect of intervention (Guala 2002, Morgan 2003, Mäki 2005).

But there are important differences. One is conceptual. Experiments manipulate material objects. Simulations manipulate representations. Two qualifications are in order. First, experiments of course often manipulate material objects that serve as representations: animal models, for example, or scale models in a wind tunnel. But in each of these cases, the model experimented on has important material qualities that purportedly qualify it as a representation of the target. By intervening on these material qualities, and learning about them through controlled manipulation, the experiment continues "interrogating nature" (Regnault, cited in Acherman 2010, 401). Simulations, in contrast, represent the whole experimental process, and therefore interrogate representations, not nature. They are "model experiments", not models in experiments (Morgan 2002). Second, simulations of course have a material aspect: they are run on computer hardware. Some have argued that this makes them material experiments (Parker 2009). Against this, I argue that in contrast to genuine material experiments, it is not the material but the formal properties of simulations that vouch for the representational quality. In other words, it is not the circuity of the computer that justifies taking the simulation as a good representation of its target, but the imagery or symbols produced by this circuity – and there might not be any isomorphism between these two.

The other important difference is methodological. Modelling and experimenting differ with regards to the sources and types of the most troubling errors. In experiments, the most troubling errors affect internal validity. Internal validity issues challenge the validity of inferences from experimental observations back to the experimental system. They operate under great uncertainty about what factors are influencing the observed outcome, and how they affect it. To establish internal validity, experimenters need to carefully design and control their experiments. In simulations, modelers determine initial conditions, set the model's parameters and program all processing rules. With the exception of programming errors, they can generally trust that the model results are indeed generated by those parameters and variables. Certainty about factors and their mechanisms is much higher than with experiments. Thus, internal validity is less of a worry for simulations than for experiments.

For modelling, in contrast, the most troubling errors concern the representational quality of the models. This is often not a problem for experiments, in particular for those that are performed on the target themselves. If one investigates the target by performing an experiment on it directly, and one is confident about the internal validity of one's inferences, then one knows that the inference is applied to the target. In contrast, when one models a target and learns about the model by manipulating it, one draws on the representational quality of the model to justify applying what one learned about the model to the target. One therefore needs to justify why the model result should be believed to also be applicable to the target, and this is often difficult and prone to error.

For these reasons, I conclude, one should distinguish between experiments and simulations, and press engineers and scientists to disambiguate their language.

REFERENCES

Guala F. (2002) Models, simulations, and experiments. In: Magnani L., Nersessian N. (eds) Model-based reasoning: Science, technology, values.. Kluwer, New York, pp 59–74

Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303-315.

Morgan, M. S. (2002). Model experiments and models in experiments. In *Model-based reasoning: Science, technology, values* (pp. 41-58). Boston, MA: Springer US.

Morgan, M. (2003). „Experiments Without Material Intervention. Model Experiments, Virtual Experiments, and Virtually Experiments."S. 216–235 in: H. Radder (Hg.) The Philosophy of Scientific Experimentation. Pittsburgh.

Parker, W. S. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese*, *169*, 483-496.

Reif-Acherman, S. (2010). Henri Victor Regnault: Experimentalist of the Science of Heat. *Physics in Perspective*, *12*, 396-442.

Giora Hon

Elements of the practice of experimentation: Commitment, methodology and technique

Science is an activity, and the scientist is an agent who pursues some practice, which in one way or another engages evidence. In science, claims to knowledge are typically supported by arguments that engage evidence at some point in explanation, in prediction, or indeed in any mode of presenting data and its interpretation. Thus, the practice of science includes at least three elements so that an argument can be formulated: presuppositions, rules of inference, and conclusions. Corresponding to these three elements in practical terms are commitment, methodology, and technique. I apply this analysis to the practice of experimentation and show how these elements feature in carrying out an experiment. The fundamental claim is that every experiment has an argumentative structure, and that it therefore exhibits, albeit implicitly, these three salient elements, which will be illustrated with several brief case studies.

Milena Ivanova

What Makes an Experiment Beautiful?

Scientific products are often celebrated for their aesthetic dimension and compared to works of art. Scientists themselves, like artists, are praised for their creativity, originality and aesthetic sensibility. In this talk I explore the aesthetic dimension of scientific experiments, from experiments performed in the early years of the Royal Society, to contemporary experiments involving complex technologies and set ups, and ask: what makes an experiment beautiful? By focusing on historical case studies as well as qualitative data collected from interviewing over 200 contemporary scientists, I identify what is aesthetically valued in the lab and the role beauty plays in experimental practice.

Dana Matthiessen and Nora Mills Boyd

Observations, Experiments, and Arguments for Epistemic Superiority in Scientific Methodology

Philosophers of science have inherited a distinction between *observation* and *experiment* that purports to track an epistemic difference. The distinction turns on understanding experiment as active manipulation. In contrast, observation is cast as characteristically non-manipulative. In virtue of this difference, some claim that experimentation is epistemically superior to observation, all things considered. This has two consequences: first, it entails that a researcher deciding between physically non-manipulative or manipulative methods that are in other ways equal should opt for the manipulative. Second, it entails that sciences in which researchers lack the ability to physically manipulate their targets of inquiry are in a worse epistemic position than those who can. We will argue against these claims.
While there can be practical grounds for drawing a conventional distinction between observation and experiment, any such distinction does not, as a general matter, track a difference in the epistemic merits of scientific methods.

To better understand scientific methodology, we propose shifting the focus from physical manipulation as highlighted by the observation/experiment distinction to an alternate set of features that cross-cut this distinction. This accounts for the epistemic boon of manipulation where appropriate, but without attributing the success of these methods to manipulation per se. In that sense, our account gets at more basic features of empirical methods to account for their superiority.

We will first provide evidence that a view of experimentation as epistemically superior to observation recurs in the philosophy of science literature and identify common underlying assumptions (including Hacking 1989, Okasha 2011, Zweir 2013, and Currie & Levy 2019). We then provide an argument for this view in (logically) stronger and weaker forms, which grounds the superiority of experiment in its fine-grained control over the production of data and its capacity to enable causal inference. We dismiss the stronger form. We then consider the weaker form, which claims that experiment is "in principle" superior to observation (including under some ceteris paribus assumption), and criticize this as well. Finally, we defend an alternate set of features that matter for the epistemic superiority of data gathering methods. Methods with features such as higher **signal clarity**, better **characterization of backgrounds**, and/or increased **discrimination and variability of precipitating conditions** will be epistemically superior to alternatives in which these features are lower, worse, and/or diminished, all other things being equal. We introduce these parameters with the aim of demonstrating, first, that they are relevant to claims comparing the general epistemic merits of empirical methods and, second, that these parameters cross-cut the traditional distinction between observation and experiment.

References:

Currie, Adrian and Arnon Levy. 2019. "Why experiments matter." *Inquiry* 62(9-10): 1066-90. https://doi.org/10.1080/0020174X.2018.1533883
Hacking, Ian. 1989. "Extragalactic Reality: The Case of Gravitational Lensing." *Philosophy of Science* 56(4): 555-81. https://doi.org/10.1086/289514
Okasha, Samir. 2011. "Experiment, Observation and the Confirmation of Laws." *Analysis* 71(2): 222-32. https://doi.org/10.1093/analys/anr014
Zwier, Karen R. 2013. "An Epistemology of Causal Inference from Experiment." *Philosophy of Science* 80(5): 660-71.

James Mattingly

Classifying experiments by way of information bottlenecks

I will extend the analysis I undertook in Information and Experimental Knowledge (Chicago, 2021). There I argued that all knowledge-generating experiments are in fact analogical experiments. That is, lab experiments, computer simulations, thought experiments, natural experiments, etc. are investigations of and attempts to gain knowledge from systems that are more or less tightly connected analogically to some system(s) of interest. (This includes even investigations of the future states of some system by examining its past and present states.) An important lesson was that there is no *in principle* distinction between normal experimental modes and other seemingly less empirically grounded modes; the difference is epistemic and the latter modes produce experimental knowledge as robust as the former.

Once we accept that, as I think we should, we naturally continue to wonder what separates the various types of experiment from each other and makes some seem somehow more empirical, for example. What distinguishes these various practices from each other are *information bottlenecks* that are associated to and characteristic of these practices. Here I first propose a taxonomy of experimental practices in terms these information bottlenecks. Categorizing experimental practices this way rather than seeing them as distinct types of practice makes clear their unity and can give insight in how to mitigate their limitations. I illustrate the way these bottlenecks interfere with the logical relations between the inferential rules that map features of the experiment to features of the world. These rules are what in Barwise and Seligman are termed the “local logics” of the information-bearing maps between the world and the lab.

Analysis of these local logics and their relations provides crucial insight into the mathematical constraints governing how one (perhaps broadly construed) system can bear in- formation about very different systems and the requirements for each of the various modes of experiment to generate experimental knowledge thereby. The local logic of the proximal system governs the inferential relations between observations and the dynamics of that system, and the challenge for experimentalists is to generate observational data that results in a sufficiently rich local logic as well as to assess the strength of analogy between the local logics of the proximal and distal systems (what Barwise and Seligman called an infomorphism). Natural experiments provide an interesting example and include, e.g., astrophysical experiments. When constrained to the use of passively generated data about some proximal system(s) there are two distinct kinds of possible information bottlenecks that can prevent the acquisition of experimental knowledge of the distal system(s) of interest: first is that there may be insufficiently many tokens (observations) to generate an interesting and properly constrained local logic of the distal system; second is that the data may well not be typical in the way required to generate univocally the infomorphic map between the proximal and distal systems. Knowing, as we now do, much more about typicality in the observable universe and having access to ever increasing quantities of astronomical data can generate robust experimental, astrophysical knowledge.

Slobodan Perovic

Theoretical and Observational Explanations in Cosmology Following the Landmark Discovery of Cosmic Microwave Background (CMB) Radiation

Various and often conflicting theoretically and observationally (experimentally) motivated approaches in cosmology gave rise to numerous alternative explanations of the CMB, most of which have since been largely forgotten. These explanations, which either modified or bypassed the now-standard Hot Big Bang model, were formulated by some of the era's most prominent physicists, including those who played pivotal roles in shaping the emerging orthodoxy. The tension between cosmological inferences drawn from firmly established terrestrial physics and more speculative theoretical inferences based on novel data underscored this development. It also prompted a reevaluation of the distinction between theory and fact, as well as an early socio-epistemic questioning of reliability of the observational process. A closer examination of this development offers a valuable opportunity for gaining a deeper understanding of the methodological and epistemic framework underpinning modern cosmology.

Collin Rice (Colorado State U.)

Balancing Experimental and Mathematical Constraints in Mesoscale Modeling in Physics

As physicist Herbert Bernard Callan notes: “It should perhaps be noted that the choice of variables in terms of which a given problem is formulated, while a seemingly innocuous step, is often the most crucial step in the solution.” (1985, 465). Numerous reasons might be offered for the inclusion of a particular set of variables/parameters within a scientific model or theory. For example, Jim Woodward (2016) has argued that variable choice ought to be guided by the *pragmatic* aims of causal representation and explanation. In contrast, in *A Middle Way,* Robert Batterman (2021) argues that mesoscale variables and parameters often ought to be selected because they are the most natural variables with which to characterize many-body systems. As Batterman tells us, he is, “addressing *metaphysical* concerns about the proper way to carve nature at its joints” (2021, 121, my emphasis). While both proposals are useful in certain contexts, they are difficult to apply in numerous areas of science where scientists have little hope (or desire) to provide causal explanations or to discover the system’s natural kinds/variables. This is certainly true in various parts of cosmology and astrophysics, but also in various areas of material science.

As a result, I first argue that we require alternative justifications for the selection of a set of variables/parameters to include in a scientific model. Fortunately, Batterman’s work suggests an alternative by proposing that physicists ought to focus on mesoscale features because, “these quantities provide a rather direct connection with measurements we can actually perform on many-body systems” (Batterman 2021, 66). More specifically, “One of the most important aspects of the hydrodynamic description in terms of correlation functions is its rather direct connection with experiment." (Batterman 2021, 15). Multiscale modelers in other areas of science agree: "The starting point for a 'middle-out' approach to modeling biological systems may be influenced by a number of factors, including the ready availability of relevant experimental data" (Walker and Southgate 2009, 451).

Using cases from material science, astrophysics, and statistical ecology I argue that the reasons for preferring middle-out theorizing strategies and mesoscale variables/parameters often comes from *both* experimental and mathematical modeling constraints. Consequently, even if these variables/parameters do not yield causal explanations and should not be interpreted as natural kinds, scientific modelers can still have good reasons for preferring the inclusion of the variables/parameters involved in employing these mesoscale modeling techniques. This is because these variables/parameters often provide a direct connection to the existing (or possible) experimental results and allow for the application of the available mathematical modeling techniques. Finally, I argue that these experimental and modeling constraints ought to be given a kind of priority because they have normative force across a wider range of modeling contexts than other justifications for variable choice. The resulting view helps clarify the relationships between these different modeling/theorizing constraints, when those constraints warrant various conclusions concerning the truth of physical models/theories, and the balance between empirical and non-empirical motivations for selecting particular types of models and theories.