Theory-Ladenness of Observation in the Experimental Context

Slobodan Perovic
Department of Philosophy
University of Belgrade
perovic.slobodan@gmail.com

Abstract

Focusing on the discovery of weak currents, the current debate on the theory-ladenness of observation in modern physics might be too narrow, as it concerns only the last stage of a complex experimental process and statistical methods required to analyze data. The scope of the debate should be extended to include broader experimental conditions that concern the design of the apparatus and different levels of the detection process. These neglected conditions often decisively delimit experiments long before the last stage has been reached, thus predetermining the extent of the dependence of data production on the theory. I explain the nature of these conditions and the theory-ladenness tendencies they produce, noting how they affect the last stage of the data analysis and providing some relevant examples.

1. Introduction: theory-ladenness vs. the bottom-up statistical analysis in High Energy Physics

The notion of the theory-ladenness of observation has been introduced (at least in its strongest form) as a blanket philosophical concept that notes an allegedly vicious circularity in the empirical testing of hypotheses. According to this view, a theory change does not turn on independent, observationally arrived at empirical content (observational terms) on which we can agree; rather, it occurs in a way that renders incommensurate seemingly competing theories.

Relying on the analysis of the discovery of weak neutral currents in high energy physics, Bogen and Woodward (1988), Mayo (1994. 1996) and Galison (1983, 1987) attempt to demonstrate that experimental results escape the theory-ladenness of observation objection, at least in its strongest form. It can be also understood as addressing the revised, weaker objection concerning theory-drivenness of observation.

The evidence for the unification of electromagentic and weak forces in particle physics turned up in the form of weak neutral currents. Such currents, kind of weak currents, do not produce any muons, unlike charged currents, as the interaction between neutron and neutrino is mediated by neutral Z^0 bosons. (The charge of incoming and outgoing particles remains the same due to such bosons.) Thus, the lack of muons in relevant interactions was considered strong evidence for the existence of electroweak forces.

Now, some particle interactions look just as desired evidence (e.g., lack of muons in relevant interactions) but they are only byproducts of interactions with the apparatus that surrounds the detecting area. A goal is to avoid recording such "artifact products" as much as possible and to distinguish them from genuine interactions.[1] Due to the complexity and the sheer number of background interactions the confirmation strategy and the criteria that guide it are usually realized by the computer simulations.

---

[1] This painstaking process can be complicated by the failure of a single circuit in the detector which might be reflected in the recordings as an event of significance.

The analysis of trigger biases (triggering algorithms determine which events of many more that occur in the detector will be recorded and studied) are run under various conditions are performed in order to discard as many of secondary interactions as possible. The Monte-Carlo protocols are used in simulations of particle interactions collisions to distinguish genuine effects and the "artifact products" by offering probabilities that a recorded event is genuine.

Despite differences in their approaches, the above-cited authors agree that physicists who worked on the experiment dealt with physical phenomena arrived at through the use of the bottom-up approach, i.e. extracted from raw data by particular statistical methods and procedures. Thus, the high-level theory had little, if anything, to do with the choice of reliable data that determine the experimental results (i.e. decisions whether enough genuine events was recorded to pronounce the discovery of neutral bosons).

Bogen and Woodward (1988) offer a rather general argument for the above-mentioned point. They introduce the distinction between scientific data and phenomena where inference from data to phenomena is a result of statistical techniques. The reliability of data is secured by experimental procedures which are based on statistical inferences as well as data reduction, exclusion of confounding factors, error and noise control, etc. – all of them employed without input of higher-order theory that they are supposed to confirm. The complex case they think illustrates well their account is precisely the discovery of weak neutral currents.

Mayo offers a more detailed account of actual procedures that supposedly secure statistical inferences. Her account of severe testing aims at providing the effect/noise discrimination and elimination of explanation by artifacts (based on the effects due to the machine itself). Probability is the key to the explanation as it concerns the test procedure, not the hypothesis. More specifically, the analysis that leads to the explanation is based on error probabilities which tell scientists how mistaken they could be given the known state of the world.

Both Mayo and Bogen & Woodward rely on Galison's (1983) historical account of the discovery. It purports that discrimination of muonless and muonful events in 1970s was justified by statistical methods used by scientists within the framework that Galison characterizes as "established physics".

In contrast, Schindler (2008, 2011) has recently argued that experimentalists do not judge reliability of data or establish experimental results independently of the background theory. He deems the strong interrelation between theory and experimentation both necessary and desirable, and he demonstrates this, among other ways, by reconsidering the weak currents discovery. He resurrects Pickering's (1984) initial analysis of the case that viewed the 1970s episode in light of the previous attempt at the discovery. Both Pickering and Schindler agree that Galison unjustifiably assumed that what was done in the 1960s was simply mistaken – Galison argue that the physicists involved simply failed to make the discovery even though they believed for some time they made it. In contrast, Pickering's detailed analysis (1984, 1997) suggests that the 1960s group embraced comparatively loose criteria to interpret results of Monte Carlo simulations based on the recordings. This was received with skepticism in the 1960s – hence no discovery was made, but such loosening was the key to the discovery in the 1970s. Fermilab's initial failure to discover weak currents was essentially a result of the agreement to use stricter criteria in interpreting results of Monte Carlo simulations – and such decisions lay in the details of their theoretical understanding.

Pickering's account was vigorously disputed by Galison (1987) and as vigorously defended by Schindler (2010) and Pickering (1990) himself.

Even though this debate might be valuable in its own right (I will not try to resolve it in this paper), it and the philosophical accounts that motivate it neglect the fact that the production of experimental results and thus the extent of dependence of observation on the theory are influenced in a profound way by a broader experimental context.[2] The groundwork for the production of results starts well before the actual theory is utilized in the statistical analysis of data and even before the actual data have been obtained. As a result of this neglect, the focus of the existing debate is somewhat narrow, as it deals exclusively with the last stages (statistical inferences) of the elaborate process of the production of experimental results and possible theory-ladenness in this particular context. Thus, this view is concerned with only one, and possibly not a crucial, layer of the dichotomy between bottom-up and down-top approaches in the complex process of the production of experimental results.

In the experiments as complex as those in high energy physics (HEP) the stages of the experiments on which the proponents of the existing debate focus are greatly constrained by the broader experimental context where some of the key choices in terms of triggering programs[3], scanning, and analysis have already been made once these last stages start unraveling. I will show how such conditions streamline the production of the

---

[2] The authors are aware of this possibility and some of them even hint at it (Bogen and Doodward 1988), but it has not been discussed among the participants of the debate.

[3] Those are the programs that determine when exactly the events will be recorded, in effect choosing to detect only certain events (particle interactions and decays) out of vastly more that take place in the detector.

results to such an extent that the broader experimental context pre-determines the extent to which observations will be a matter of circularity of testing, i.e., determination of data by theoretical predispositions and theoretical reasons for believing in the reality of phenomena. This creates broad tendencies in the production of results in HEP, which typically lean towards one of the two extremes – bottom-up or top-down, much before the last statistical stage of the analysis takes place. I will label these tendencies *"theory-ladenness tendencies"* that range from weak to strong. (We could as well label this notion "theory-dependence tendencies" as well, where the stronger theory-dependence theories would approach the initial, strong understanding of theory-ladenness. But this choice seems rather a rhetorical matter.) This makes it essential to understand broader experimental context and how exactly the last stages of the analysis fit in, if we want to fully understand the nature of dependence of observation on the theory.

A participant in the current, narrow, debate could remark that the debate is concerned with the extent to which analysis of data is dependent on the current theory once we have data – irrespective of the way in which one arrives at such data. But the point is that a possible strong dependence of methodology of data analysis on theory of this last stage can be trumped, severely limited or amplified earlier in the process.[4] It is true that we can argue on the dependence on the theory of data analysis alone, but this is not a good enough reason to neglect the broader and possibly more substantial dependence, or the lack of it, on the theory. Otherwise, there is a danger that, at best, the existing debate remains focused only on a particular piece of a much larger puzzle

---

[4] For instance, the fact that the bottom-up statistical methods are applied might be neutralized by the design of the apparatus that is strongly influenced by the current theory or the lab employing a substantially theory-laden detecting regime (I will offer relevant examples of both shortly), taking it closer to a top-down approach.

concerning the theory-experiment interaction, and, at worst, either side's account obfuscates the key aspects of the experimental process.

2. The broader experimental conditions: design of the apparatus, data analysis and theory

Detection in HEP experiments relies on, broadly speaking, two distinct systems of detection, semi-automatic and automatic (generic). (Perovic 2011) They have developed as two reasonable and distinct responses to the need for balancing efficiency in handling complexities of experiments with a direct involvement of physicists in detection. The computerized production, readings and statistical analysis of experimental results are much more dominant in some experiments while they are used only very selectively in others. Also, detection process in some experiments has been serviced by specialized trained technicians – the so-called "scanners" – while detection in other experiments was dominated by physicists involved in the entire discovery process.

The increase in the automation was motivated by the expectation that such an approach would offer "experiments a thousand fold increase in cost-effectiveness. The combined reduction in cost and increase in computing power would allow experiments to use less biased trigger assumptions, record more data on tape, and simultaneously accelerate the data analysis leading to publication" (Hoddeson et al., 2008, 274-5). As a result, the recording of events has become mediated by the program that, in effect, reconstructs (provides estimates) rather than straightforwardly detects particle events.

Now, the more automated approach to detection typically goes hand in hand with the experiments that aim at confirming a specific hypothesis, which requires analysis of specific tracks rather than broad scanning and exploratory analysis that does not rely directly on current theory. Not surprisingly, such experiments are typically based on theoreticians' current preferences.

In one such typical and famous experiment that resulted in the discovery of W and Z bosons, the experimental apparatus consisted of very specialized detectors (to which we will turn our attention shortly) as the expected choice of tracks and triggers to be analyzed was very specific. Moreover, the experiment was really an accurate determination of the masses of particles whose existence was assumed based on the established theory; the masses of the particles were already determined fairly reliably and precisely (Amaldi 1987, 3; Blondel 1994, 418). And proton and anti-proton collisions that were analyzed in the experiment were treated as a test of the standard electro-weak model. And the experiment was designed directly on theoreticians' request as the pressure to discover W and Z was strong at the time. (Darrilat 2004, 2)

The search for the proton decay had a very similar structure (Perkins 1984) although it turned out to be somewhat futile. It was initiated by a few leading theoreticians, despite the fact that many experimentalists regarded it as completely uninspiring at the time. The half-life of proton decay was found to be very close to the value suggested before the theoreticians' initiative, close to infinity.

Now, even though in such confirmatory experiments the details of the statistical tinkering will decide whether (or rather when) the decision will be made that the discovery occurred – and this is the domain where exploratory statistical analysis can play a role, they already have a strong tendency towards theory-ladenness (i.e., strong theory-dependence) of data production. The last stage can rectify this only within a domain severely limited by current theory.

Actually, the broadest conditions, yet perhaps the most constitutive of the theory-ladenness tendencies, emerge already at the level of the design of the apparatus. Such tendencies exhibited at the level of the design of the apparatus might be hardest to counter in later stages of the experiment.

For instance, the explored energy domain explored by the apparatus in the W and Z bosons experiment was so narrow that one of two competing detectors, namely UA2 "could not measure particle charges except for limited regions where W decay asymmetry was maximal" (Darriulat 2004, p. 6). In the same sense the UA1 was a "clean machine" (Ibid.) geared for only most accurate measurements within the existing theoretical framework. Thus, already the design of the apparatus based on very specific current theoretical expectations renders this experiment close to the top-down approach and inherently conducive to theory-ladenness of observation. And it minimizes the potential impact on theory-ladenness of the tinkering with statistical methods.

There is an even more drastic case of such broad initial limitations that can be overlooked if we focus on theory-dependence in the statistical analysis alone: the LHC is apparently bound to miss a part of the spectrum that potentially contains Higgs boson as predicted by SUSY, a major alternative to the Standard Model (Cobal 2006, 271). Thus, even if, for instance, the bottom-up side of a hypothetical debate on theory-ladenness of data analysis at LHC turns out to be correct, it would be a very limited assessment given the data domain to which the statistical analysis could have been applied: it neglects the fact that the data production was severely limited from the outset by the design of the apparatus, which perhaps resulted in eliminating a crucial domain (for the discovery of Higgs boson) of data production.

It is important to realize though that the dependence of design of the apparatus on the theory is often part of broader, long-term research strategy which precedes it. Thus, often the same event can be explored with hadron colliders or linear accelerators of very different design. The latter favors exploratory detection as it produces very few background interactions, while the former requires heavy automation of the detection process due to heavy background. The linear accelerators however cannot as yet explore as high energy domains as hadron colliders due to technical constraints. Thus, the physics community is faced with numerous tradeoffs between exploratory and confirmatory searches, energy domains and saturation of the backgrounds at every level, which determine to a great extent whether and how much each stage of detection can be autonomous from the existing theory.

In contrast to confirmatory experiments and the excessive use of automation that goes along with it, the semi-automated systems are better suited for exploratory experiments that tinker with the triggering conditions, scanning and analysis. Such experiments are closer to a bottom-up approach as they are less conducive to theory-ladenness of data and circularity of testing even irrespective of the extent of tinkering with statistical methods that takes place in the last stage of the analysis: the choice of energy regime, triggering algorithms, phenomena to investigate and detecting methodology already streamline production of data in a direction of strong autonomy from current theory. Actually, instead of hypothesis directly following current theory, a cluster of possible hypothesis loosely tied to or challenging the current theory is being presumed. And the experiment is often a hypothesis-formative, rather than a hypothesis-conformational tool.

A shining example of this experimental approach is the discovery of J/psi particles. In 1974 the existence of the charm quark was established in the form of J and psi particles (neutron mesons with 3-4 GeV masses) as a combination of a charm quark and its anti-quark. (Goldhaber 1997) The experiment was designed as a scanning of broad energy range without any theoretically-driven expectation of “narrow structures” that would indicate in advance possible existence of desired particles in the domains explored. An “inconsistency”, or a peak in cross-sectio at 3.1 GeV, completely at odds with current theory, was noted early and, in contrast to the methodology of confirmatory experiments, explored meticulously (Goldhaber 1997, 58-9). Actually, the physicists decided to introduce a new technique: to change the energy pattern and explore the cross-section as a function of colliding energy while looking for predicted resonances - and this tinkering, on a phenomenon detached from the current theory, resulted in the discovery (Ibid., 59-62).

Similarly, the co-discovering team at Bevatron that worked with much larger backgrounds, due to the nature of their apparatus, developed triggers and analysis procedures suited for a broad search. (Ting 1977, 238) The guiding idea was precisely to set up the apparatus for the search of narrow resonances not predicted by the existing theory. (Ibid. 241)

Thus, if the use of Monte-Carlo protocols and the statistical analysis turns out to be of the kind that the bottom-up side argues for, it was nevertheless only a minor generator of the autonomy of the data production in the overall experiment compared to the previous stages. If it turns out that the statistical analysis was guided by the theory – whatever that theory that was employed in the analysis might be (at best a provisional hypothesis loosely tied to the current theory) - it is really a minor point as the entire data-preparation process resulted in choice of data that were fairly detached from current theory to start with.

3. Theory-ladenness of data production and detecting strategies

Thus, the upward/downward theory-ladenness tendencies of data production depend neither exclusively nor primarily on the last stages of the analysis. The reasons deciding such tendencies are not uniform and it is not always easy to say which tendency is predominant in the experiment and the community developing it. The experimental background varies across periods and labs; different physicists opt for different approaches for various reasons which eventually place them closer to the bottom-up or top-down end of the spectrum of approaches.

Sometimes one approach to detection is favored over another as a deliberate effort to pursue a particular view of science. For instance, R. R. Wilson, director of Fermilab in the 1970s, discouraged the computerization of Fermilab believing that individual physicists should be involved in every stage of the discovery process. (Hoddeson et al., 2008, 341). This approach was motivated by his ideal of "science … pursued by lone independent explorer" and a physics lab as a place where an individual physicist has control over the entire experimental process.

At other times, the approach is largely driven by the culture of a laboratory. At CERN, the trend of rapid automation of detection was favored from the outset, driven by the hierarchical structure of the lab and detachment of experimentalists and theoreticians. Very early on experimentalists were turned into serviceman of the apparatus and external theoreticians were practically excluded from the experimental process. (Perovic 2011)

At still other times, a detecting regime is a result of necessity. For example, the discovery of Ω mesons was a direct result of the application of a "pedestrian" regime of scanning and analysis when the computing power was already allocated to other projects. (Maglic et al. 1961) And in the case of the discovery of the second element of the meson pair, namely the ρ meson, the scanning machines were used full time on different problems deemed more important by the senior members of the lab. So J.A. Anderson and his collaborators (Anderson et al. 1961) made measurements directly on the scanning table. They applied the Chew and Low extrapolation method to predict dipion resonance (i.e., ρ meson).

Finally, to turn our attention to the much-discussed discovery of neutral currents, it was a standard confirmatory experiment that involved the elaborate study of a well-known phenomenon without much innovation in experimental apparatus and techniques. The crux of the experiment lay in its final stage - the assessment of the statistical data with the aim of distinguishing artifact and genuine events that essentially span over a decade. The production of data and the design of the experiment did not involve anything like exploratory methodology that we have encountered for example in the J/psi discovery. And it may not be surprising that CERN triumphed in the race, given that its organization favored confirmatory experiments. Thus, even though the outcome of the debate over the theory-ladenness of the last stage of the experimental analysis might be valuable in this case, the broader conditions render the experiment as strongly theory-laden in the first place which is yet another indication of the necessity to broaden the parameters of the debate.

References:

Amaldi U. et al. (1987) "A Comprehensive Analysis of Data Pertaining to the Weak Neutral Cuurrent and the Intermediate Vector Boson Masses," *Phys. Rev.* D36 1385.

Anderson J.A. et al. (1961) *Phys Rev. Letters* 6, 365.

Blondel A. (1994) "Precision Tests of the Standard Electroweak Model at LEP," in Peach K.J. and Vick L.L.J., eds., *High Energy Phenomenology*.

Bogen J. and Woodward J. (1988) "Saving the Phenomena," *The Philosophical Review*, Vol. 97, No. 3, 303-352.

Cobal M. (2006) "Physics Potential of the Atlas Experiment," Sidhart F. Honsell and de Angelis A., eds., Frontiers of Fundamental Physics, Springer, 267-272.

Darriulat P. (2004) "The Discovery of W and Z, a personal recollection," *Eur. Phys. J. C* 34, 22-40.

Galison P. (1983) "How the first neutral-current experiments ended," *Reviews of Modern Physics*, 55, 477 – 509.

Galison P. (1987) *How Experiments End*, Chicago: University of Chicago Press.

Goldhaber G. (1997) “From the Psi to Charmed Mesons: Three Years with the SLAC-LBL Detector at SPEAR,” Hoddeson L. et al., eds., *The Rise Of The Standard Model*, Cambridge University Press, 57-78.

Hoddeson L. et al. (2008) *Fermilab: Physics, The Frontier & Megascience*, University of Chicago Press.

Maglic B.C. et al. (1961) *Physical Review Letters*, 7, 178.

Mayo D. (1994) “The New Experimentalism, Topical Hypotheses, and Learning from Error,” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1994, Volume One: Contributed Papers, 270-79.

Mayo D. (1996) *Error and the Growth of Experimental Knowledge*, University of Chicago Press.

Perkins D.H. (1984) “Proton Decay Experiments”, *Annual Review of Nuclear and Particle Science*, Vol. 34, 1-50.

Perovic S. (2011) “Missing Experimental Challenges to the Standard Model of Particle Physics”, *Studies in History and Philosophy of Modern Physics,* 42(1), 32-42.

Pickering A. (1984) *Constructing Quarks*, The University of Chicago Press.

Schindler S. (2008) “Rehabilitating Theory: Refusal of the ‘Bottom-Up’ Construction of Scientific Phenomena”, *Studies in History and Philosophy of Science – Part A,* 38 (1), 160-184.

Schindler S. (2011) “Bogen and Woodward’s Data-Phenomena Distinction, Forms of Theory-Ladenness, and the Realiability of Data”, *Synthese*, 182 (1), 39-55.

Ting, Samuel C.C. (1977) “The Discovery of the *J* particle: a personal recollection,” *Reviews of Modern Physics*, Vol. 49, No. 2, 235-249