*Slobodan Perović*
Department of Philosophy, Faculty of Philosophy
University of Belgrade

Original Scientific Paper
UDK 167
539.12:167

# THEORY-DRIVEN EXPERIMENTATION IN PARTICLE PHYSICS

**Abstract:**[1] *J. Woodward and S. Schindler agree that experimentation being motivated / driven by the theory it tests (Tt) is an epistemically benign form of theory-ladenness (TL). Despite their agreement, they describe two distinct forms of tested– theory drivenness (TD). I argue that TD Schindler describes is a particularly severe form of TL. I label it strong TD. It kicks in early in the measurement during the operation of the apparatus, preceding the stages at which inferences on the status of the observed phenomena are made. I briefly present a classical toy-case as an instance. The elimination of strong TD by calibrating the instrument based on a different operational theory is arguably accomplishable in the toy-case. Strong TD, however, is ubiquitous in particle physics where, contrary to what A. Franklin and Woodward argue, the experimental environment prevents calibration from eliminating it. Instead, a strategy of incrementally widening experimental loop confronts the problem, e.g. in the discovery of J/Ψ particle. I discuss why the context of the particle physics experiments is conducive to this strategy, whether it eliminates strong TD, and whether it remains a genuine epistemic problem within such a context. Weak TD as sketched by Woodward involves P being predicted by Tt or P being deemed an important physical value as the motivation for performing measurement of P. It is not a form of TL in a traditional sense, but in the context of experimentation in particle physics, I argue that it is an acute socio-epistemic problem, perhaps more acute than the possibility of TL.*

## 1. Theory-drivenness and theory-ladenness of experimentation

In a much-discussed paper on theory-independence of inferences from experimental data to phenomena, Bogen and Woodward (1988) argue that successful experiments establish phenomena, or “features of the world that in principle could recur under different contexts and conditions” (Woodward 2011, 166). Thus, phenomena such as the melting of lead at 327.5°C or the gravitational red shift

---

[1] This paper was presented at the Annual Meeting of the British Society for the Philosophy of Science held in Stirling, Scotland in July 5-6 2012. The work was supported by the project “Dynamic Systems in nature and society: philosophical and empirical aspects” (evidence # 179041) financed by the Ministry of Education, science and technological development of the Republic of Serbia.

are only subsequently explained by the creation of different theories. Briefly stated, experimentally and observationally phenomena are arrived at (i.e. distinguished from the background noise) and deemed genuinely interesting (rather than artifacts of the apparatus) independently of a specific theory that is being tested.

These ideas triggered much debate. In a long passage responding to his critics, Woodward discusses a form of theory-ladenness (TL) based on experiments *being motivated* by the theory they test but dismisses it as epistemically benign:

> Suppose theory *T* explains some phenomenon *P* that a researcher wishes to measure/detect from data *D*. Suppose also *T* provides a motivation for measuring *P*: *the researcher measures P because she wants to test T or because T says P is an important quantity that plays a fundamental role in the nature* – the researcher would not have attempted to measure *P* if he did not regard *T* as serious possibility... There will be an obvious sense in which *T* is "involved" in reasoning from *D* to *P* – the researcher would not have engaged in this reasoning at all or would not have used concepts drawn from *T* to describe its results of this reasoning if she did not accept *T* or at least take it seriously. However, *this sort of involvement of T in data to phenomena reasoning does not necessarily mean that T is being used to explain D or that D cannot be evidence for P unless T is conceived as playing this explanatory role.* (Woodward 2011, 177; emphasis mine)

Here, Woodward describes what he believes is a ubiquitous theoretical motivation for conducting experiments, reaching a conclusion on its epistemic irrelevance with respect to the dilemma of whether inferences rely on the assumptions of the tested theory and exactly in which way.

In a passage reflecting on Woodward's view S. Schindler (2011) suggests a more substantial notion of theory-motivated or theory-driven experimentation that, however, leads him to the same conclusion regarding its epistemic pertinence:

> According to one of these forms [of theory ladenness], ***the design and*** *the* ***conduction*** *of experiments can be motivated by theories* (call this the thesis of "theory-drivenness" of scientific practice). Such a claim is *epistemologically rather unproblematic*. Whether or not the ***conduct*** *of particular experiments has been motivated by the theory, which these experiments seek to test*, is *irrelevant to the (logical) question* of whether the data obtained in these experiments do or do not support the belief in particular phenomena, and whether the latter in turn, do or do not confirm the theory's predictions. (Schindler 2011, 42; emphasis mine)

They agree that there is a genuine logical problem at the core of the question about whether experimentation is theory laden, in other words, whether *P* is derivable as significant (i.e. deemed to stand out from the background noise and distinguishable from the artifacts produced by the apparatus) from data *D*, with or without reliance on *T*.

Woodward argues that although measurements are often theory driven or motivated, this does not necessarily imply the involvement of *T* inferring that *D* is evidence for *P* or that *P* confirms or refutes *T*. While Schindler might be referring to a more substantial kind of theory-drivenness (TD), as he characterizes it in terms of design and conduction which are motivated by the theory that is being tested (*Tt*), he draws the same conclusion as Woodward on its relevance to theory-ladenness of evidence *D*.

Both accounts assume that TL is a question of logical connection between data, phenomena and theory and thus raises a genuine epistemological issue, while TD does not or at least, does not necessarily do so. Thus, TD is a benign form of TL, independent from epistemologically more problematic forms.

Now, even if we consider a more substantial notion of TD, of the kind Schindler hints at, one requiring that the conduct and the design of the apparatus for testing *Tt* be motivated / driven by *Tt*, it is still hard to decide whether the above assumption is correct. The available accounts are insufficiently clear on precisely what the motivation or drivenness of the experimentation by the theory involves.

A clue to a reasonable distinction between the two sorts of TD in Woodward and Schindler's accounts lurks from everyday language. Thus, when I say "My motivation for playing the piano is my love of piano-playing", my motivation gets the ball rolling but is still external to what I do. In contrast, a statement such as "Democratic principles motivate/drive my struggle to conduct the state democratically" indicates that the motivation is internal or intrinsic to the action; the conduct is driven by a particular form of rules or principles. In this way, we can distinguish between weak and strong TD.

Weak TD, which Woodward seems to have in mind, is characterized by testing theory *Tt* which is externally motivated by *Tt* (e.g. it is deemed important to measure *P* predicted by *Tt* in light of *P*'s role in *Tt*) without necessarily presuming that detection procedures and apparatuses are operated by or rely on *Tt*. As such, it may seem epistemically benign; it does not raise an issue pertinent to the debate on TL by directly questioning the logical relationship between data, phenomena and theory.[2] Yet, as I will argue in Section 3, it raises another issue, as important epistemologically as TL, that is not discussed by Woodward.

In contrast, and still following the everyday language analogy, strong TD involves the testing of *Tt* such that *Tt* drives the conducting of the experiment in that *Tt is part of an operational theory (To) of the apparatus*. This TD stems from Schindler's characterization, even though he may have intended to refer to a weak TD like Woodward. It would be unfounded, however, to exclude the possibility of *To* assuming at least parts of *Tt*, if *the conduct and design of the experiment* are driven by the theory they test. In such a case, as we will see shortly, the conclusion that TD is epistemically benign is not justified. Thus, in its strongest form, the involvement of the *Tt* in the conducting of the experiment means that the operational theory *To* of the apparatus assumes either *Tt* or part of *Tt*. And this is, as I will argue shortly, both a very strong sub-category of TL and central to particle physics experiments.

---

[2] We could label weak TD a particularly weak form of TL that does not raise the epistemological issue as do other stronger forms of TL, but it is less confusing if we simply distinguish weak TD from TL.

### *1.1 Strong theory-drivenness of experimentation: a classical case*

A. Franklin et al. (1989) point out, in comparison to the forms of TL where theory enters into the judgments on the significance of phenomena observed and affects analysis after the apparatus has delivered data, "a more difficult problem arises when the apparatus, or part of the apparatus, *depends for its proper operation* on the theory of the phenomena under test." (Ibid. 230) He does not label this sort of TL strong TD but this is exactly the situation where the experiment is strongly driven by the theory it tests, as we characterized it above.

A much discussed toy example of strong TD that allegedly results in a vicious circle is testing the hypothesis that objects expand as their temperature rises. The operational theory tells us how to use and read the apparatus, i.e. how to produce data. In this case, data production depends on the tested hypothesis. We heat an object, measure independently whether its volume increases, and measure whether its temperature rises. If we use a mercury-based thermometer to perform the latter, the measurement is predicated on the assumption that the expansion of mercury is a result, and the measure of the rise of temperature: the experiment assumes exactly the hypothesis that is being tested. More precisely, *To* assumes *Tt*, and, in effect, *such an assumption (mercury expands as its temperature rises) ensures that data (observing mercury reach marks on the container) unequivocally produce a significant P* (measure of the rise in temperature) distinguishing it from the background noise and eliminating the possibility that it is an artifact or something entirely unknown. Thus, how *the apparatus is used and read pre-relates data and the phenomenon of significance in a circular manner.*

*To* deal with this vicious epistemic circle, Franklin suggests fulfilling a condition under which strong TD is eliminated; e.g. using an operational theory *To'* instead of *To* and then trying to establish *P*. Thus, the mercury thermometer can be calibrated against another *To'*-driven thermometer. For instance, we can use a constant volume gas thermometer, whose operation depends on the ratio of pressure and absolute temperature, not on the change of volume. Once calibrated this way, the mercury thermometer can be treated simply as a calibrated measuring device where data are not *Tt* laden; this eliminates strong TD.

Before proceeding with the central issue (strong TD in particle physics), we should point out that strong involvement of *Tt* in the conducting of the experiment (when *Tt* drives its conduction) gives rise to a genuine epistemic problem of TL when *To* assumes *Tt*. Thus, it seems unjustifiable to treat TD as logically distinct from TL. Strong TD is a form of TL; *Tt* kicks in very early in the process, long before inferences from existing data as to the status of phenomena are made and well in advance of confirmation or refutation of *Tt*; thus, the operational themselves procedures raise the epistemological problem.[3]

The toy example points to a severe epistemic problem with strong TD; at the same time, it suggests a way of eliminating it. But is the toy example a rare

---

[3] The debate between Woodward and Schindler focuses largely on such later stages of experiments; this might have contributed to their dismissal of theory-drivenness as epistemically irrelevant.

exception? Is there always a strategy of eliminating strong TD analogous to the toy-case calibration? It is, of course, unrealistic to expect to find exactly the same kind of case, but can we find significantly similar ones?

### *1.2 Strong theory-drivenness in particle physics: a case of J/Ψ discovery*

The experimental procedures and the design of the apparatus in particle physics, often, perhaps even typically, involve strong TD and a seemingly vicious circle that characterizes it. A preliminary phenomenon of interest (*Pp*) is initially determined in such a context, and the strategy of a widening evidence loop is employed. In the later stages of securing a more reliable *P*, it often involves the use of two distinct crosschecked experiments, distinct programs of analysis in an experiment or two distinct detectors.

The experimental process that led to the seminal discovery of *J*/*Ψ* particle illustrates this very well. The physicists in question detected an unexpected peak at 3.1GeV (labeled *Ψ*), and later at 3.6 GeV (labeled *Ψ*'). They very quickly realized these two peaks might turn out to be two instances of a bound state, where c quark and c anti-quark, both postulated by recent theory, are bound. The discovery of the peaks was immediately followed by the spectroscopic exploration of intermediate energy states between 3.1 and 3.6GeV. The tested theory (*Tt*) stated that the two peaks were instances *Ψ* and *Ψ*' of a bound state. A more specific assumption (*Ta*) of this theory was that *Ψ and Ψ' as the bound states will decay in the intermediate region (3.1–3.6GeV) resulting in γ radiation peaks*. Thus, "following the discovery of *Ψ* (3095) and *Ψ*' (3684) several related states, populated by γ-ray transitions, were predicted" (Biddick et al. 1977).

The spectroscopic analysis turned up an initial **γ** peak in 3.1–3.6GeV; later another four peaks were found (Ibid.; Feldman et al. 1975). The peaks were treated in a preliminary fashion as *debris of a Ψ/Ψ' decay predicted by Ta* as "one could predict the sign and relative magnitude of many transitions from one particle into another" (Gilman 1985, 3). But the debris was *also taken* as *evidence for the existence of the bound states, with Ψ and Ψ' as their instances (Tt)*. Thus, the γ-transitions were seen as evidence of the existence of bound states "from a completely different direction" (Ibid.). In other words, "if the *Ψ* consisted of the combination of two quarks... these quarks passed onto a final state" (Goldhaber 1997, 63) and the observed γ-transitions showed this.

Note the nature of *To* in this case: detected γ peaks in 3.1–3.6GeV range were treated as an indication of the existence, as well as the measure of, relative magnitudes of *Ψ*-decays. Such indirect measurement was possible only based on the *Ta* that bound states will produce γ peaks – only given such an assumption, one could determine with some certainty that the peaks were neither background noise nor artifacts. It seems that we have *an even smaller vicious Tt->data->phenomena->Tt circle than in the thermometer toy-example*. In the toy-example we measure the change in volume of the heated substance independently and introduce a vicious circle only when measuring temperature. *In this experiment, however, we use detection of γ peaks as the evidence for and measurement of Ψ* *decays, as well as evidence of their products (γ peaks themselves) in order to check their mutual dependence (Ψ/Ψ' decays will produce γ debris/peaks).*

The experiment went on to test the finer structure of individual states in *γ transitions* in the intermediate energy region. *Tt* predicted specific *structure of γ transitions* of the debris of the decay of bound states, not just their occurrence. Thus, in the first phase of spectroscopy, immediately following the discovery of 3.1 and 3.6GeV peaks, "the first evidence for such [intermediate] states [χ] was obtained at DORIS and SPEAR [colliding storage rings] and an experiment using NaI (T1) detectors at SPEAR placed upper limits of about 6% on the branching faction of *Ψ'* –> γχ, *subject to certain assumptions about χ decay modes*" (Biddick 1977, 1324; emphasis mine). In the second phase (1975–1983), these assumptions, now involving more detailed predictions as to the structure and the background of individual γ transitions, were tested using the new "Crystal Ball" detector: "Two subsequent experiments ... at SPEAR were accordingly merged and expanded to search for these monoenergetic γ rays" (Ibid.). The search turned up results very close to the predicted values.

In effect, the strategy created a widening and more detailed evidence loop by adding more detailed predictions and their subsequent confirmation without ever relying on factors outside the loop, as for example, an alternative operational theory independent of the *Ta* and *Tt*. The strategy never involved an independently determined factor analogous to the calibration in the toy example.[4] Instead, the reasoning was that if *Ψ* and *Ψ'* were bound states, γ peaks should be detected in the intermediate region (*Pp*); and if such intermediate states were detected, non-random specific structures of γ transitions should also be detected. Thus, the evidence indicating the bound states detected by the initial experiment was strengthened by the discovery of the γ-peaks and was further successfully tested by searching for particular γ decay structures. These two steps involving more detailed predictions of *Tt* and successful testing basedon *To* driven and refined by it, turned a preliminary circularly established *P* (3.1 and 3.6GeV peaks) into a reliable *P*.[5]

### *1.3 The limits of calibrating the apparatus and the widening evidence loop strategy to deal with strong theory-drivenness*

Are experiments in particle physics typically strongly theory driven in a similar way? And is it possible that the ability to eliminate strong TD, as in toy examples, might be an exception with little to do with the experimental practice

---

[4] If anything, it is similar to measuring temperature based on the differential thermal expansion in the toy example. (Chang 2001; Franklin et al. 1989)

[5] Similarly, strongly theory driven programs of data analysis involve cross-checking with alternative programs. In the case of *J/Ψ* discovery, SLAC and LBL both discerned relevant interactions; although they relied on different programs of analysis, each was predicated on assumptions checked by the other program. The outcome involved interactions in the detector, data acquisition (various triggers of recording), data analysis programs and finally, a one-event display that constituted the actual observed phenomenon. Different assumptions of the analysis programs resulted in differences all the way from data acquisition to the one-event displays. At each step, cross checking of alternative programs of analysis based on two different sets of *Tt*-assumptions in *To* was performed.

in particle physics? Perhaps we will never know for sure. Nor we will know whether other cases of strong TD are more severe than the toy example. Certainly, the cases involve more complex apparatuses, and the operational theories and readings are applied in a much more complex experimental environment.

Yet it may be precisely the nature of the experimental environment that renders the strong TD the most dominant form of TL[6] in particle physics – it is hard to see, as we will explain shortly, how reliance on tested theory for the operation of the apparatus in such a context can be avoided. Franklin argues that strong TD can be eliminated by replacing *To* with a direct technique such as calibration and deems this a universal procedure of validating scientific results (Franklin 1997). While this might make sense in some cases, in particle physics, due to the experimental environment or context, this one-shot elimination technique seems inadequate at the very beginning and increasingly less employable as the experiment progresses.

We could certainly try to portray the strategy employed in such experiments as relying on calibration, provided we stretch the notion of calibration. Franklin (1997) seems to be doing so and Woodward follows suit (2009, 171–172) when discussing an example of an apparatus in particle physics that produces a known phenomenon recorded with a different detector in similar circumstances, along with the novel phenomenon. The occurrence of an independently recorded surrogate phenomenon along with the phenomenon being searched constitutes a calibration that indicates the apparatus is working properly and is independent of *Tt* that postulates novel phenomena. Discussing the production of surrogates, Franklin notes: "The Cerenkov counter was checked and its efficiency for positrons was measured, by the comparison with a known positron detector in an independent experiment" (Franklin 1997, 32). Similarly, we do not need a theory of a telescope to demonstrate that a patch on a distant building visible through the telescope is really there if we see it with the naked eye (Chalmers 2003, 503). Supposedly, the occurrence of the surrogate ensures that the apparatus works properly and that the novel phenomenon is genuine without involving any assumptions of *Tt*. Supposedly "data themselves" do the job of calibration. (Franklin 1997, 32)

It is questionable whether such a notion of calibration in particle physics is equivalent to the notion we encounter in the toy example involving thermometers or a simpler case of calibrating a telescope.[7] An occurrence of the surrogate phenomenon raises our confidence in the reliability of measurement to some extent, but we are considering its *co-occurrence* with the novel phenomenon – and this is a far more demanding task in terms of estimating reliability. In our example, old operational theories that rely on spectroscopic readings and predict occurrence of the surrogate would be insufficient to decide whether γ-ray peaks are of any theoretical significance because their occurrence still might constitute a peculiar artifact or suggestive background noise, even though the surrogate is

---

[6] Or a particularly strong kind of theory-dependence, as this sort of involvement is sometimes vaguely labeled

[7] See (Chalmers 2003, 504) for complexities of the calibration strategy of electron microscope.

observed; or the co-occurrence of the surrogate and the novel phenomenon could be a result of suggestive background noise or a peculiar artifact. The occurrence of the surrogate indicates it is more likely the apparatus is working properly than if it had not occurred, but it *does not imply enough* about the nature of co-occurrence of the novel and collateral phenomenon, whether they co-occur as a result of genuine processes or as artifacts or a suggestive background noise. And this is precisely the kind of information that we seek in the experiment.

This is especially true of the kind of experiment illustrated in the *J/Ψ* discovery. The experiment was exploratory; it was not a result of a pre-conceived search with past and/or preliminary work to rely on where calibration strategies could be pivotal – or even involved, for that matter. It was not the situation of a phenomenon being tested at energy levels where other phenomena had been tested on another apparatus or detector, with a plethora of accompanying *Tt*-independent operational theories and alternative instruments. The explored energy level had never been tested, and *To* could only emerge in a symbiosis with *Tt*. The role of the surrogate calibration, if any of the procedures could be indeed called that, was limited. In short, it was a pioneering exploration of the entire maze of γ ray transitions and backgrounds that followed the discovery of 3.1 and 3.6GeV peaks. Different structures of the spectrometer (DASP) and NaI (T1) detector would have minimized its potential and thus its role even more.

Various assumptions of *Tt* played the key role in these detecting processes in both strands of the experiment while pursuing the widening evidence loop strategy (Braunschweig et al., 1975; Goldhaber 1997, 66; Feldman et al., 1975).[8] Hence, it was crucial for the physicists to crosscheck with other groups to confirm novel occurrences (Feldman et al., 1975, 821) – an advanced stage of the strategy of widening the evidence loop, as we will see shortly.

Franklin and Woodward emphasize calibration, arguing that it eliminates the strong experimental drive by *Tt* but this seems exaggerated. Calibration can play an important role in raising confidence in the apparatus reliability but it is not sufficient to detach To from *Tt* except in special circumstances. The first instance of anything like the straightforward calibration Franklin and Woodward refer to occurred when a new set of detectors ("Crystal Ball") based on initially used NaI(T1) was used to check the structure of γ transitions; only then could the independent surrogates be introduced. Yet before the use of this new set of detectors and based on the initial spectroscopic measurements of γ ray peaks alone, the physicists were convinced that they had discovered bound states (Goldhaber 1997, 61) and even published their results (Krumhansl and Trigg 1974). Subsequent detailed experiments (1974; 1976) on independent detectors exploring the structure and backgrounds of the intermediate states (i.e. looking for monoenergetic γ rays) took about two years to complete and dealt mainly with

---

[8] Given the imperviousness of the experimental environment to the pivotal role of calibration, it may not be surprising that the apparatus turned out to be off by 10MeV when *Ψ* state was discovered; the physicists realized this when they went on to discover *Ψ'* state (Goldhaber 1997, 60).

the nature of bound states and quarks. Thus, physicists judged very early stages of the widening loop that did not involve substantial calibration to be unproblematic.

Moreover, it is questionable whether calibration even in the toy example is a safeguard against strong TD, and whether we can answer this question unequivocally in other experimental contexts. As Chang (2001, 283) suggests, any "measurement will involve us in a circle" as long as we stick to empiricist principles. He uses the history of measuring temperature with gas thermometers to demonstrate this. One way to avoid the circle is to try widening it by using assumptions that drive instruments as autonomously from tested hypotheses as possible, without hoping that this will eliminate the dependence. Duhem (1906) claims that physiologists have an easier time achieving this than physicists because the latter cannot detach their operational theories from the hypothesis they are testing with the apparatus: "The best we can and should hope for is an agreement between many sources" (Chang 2001, 283–4). Thus, in the *J/Ψ* case, in the absence of substantial calibrating techniques in the crucial stages of the experiment, it was extremely important having results delivered at the same time from the other two experiments exploring the same energy range (for *J* – the same bound state as *Ψ*) and by two detectors (spectrometer and Cerenkov counter) at SPEAR (for γ-ray transitions). While they did not provide a calibrating basis, the results from various sources led researchers to think that it *could not have been a mere coincidence the resonances occurred at the same*.

The experimental context lends itself to what Hacking (1983) calls arguments from coincidence: "Arguments from coincidence ... serve to vindicate both the data and the theory used to interpret data. They are effective to the extent that the only plausible explanation of the match between theory and data is the validity of both. Experiments that do narrow down the options so that this is indeed the choice are not easily come by" (Chalmers 2003, 506). These may not be easy to come by in physics but they are common in particle physics. Thus, it may not be detrimental that reliability is partly dependent on circularity, as it might be in the toy case..

Chalmers (2003, 505) states that "experimental scientists have a whole arsenal of techniques for tackling any epistemological problems [because] the use of instruments in science are theory dependent." Such a sweeping generalization might not be prudent given the context of the debate, especially as Chalmers indicates only one such technique. Yet he might have a point: the techniques seem varied; therefore, they might not be instances of one unifying technique. And such techniques do not necessarily tackle the issue as directly as calibration. Nor can they, given the experimental context in particle physics; nor need they necessarily do so. A successful widening of the circle can be regarded as a sign of reliability very early on in particle physics experiments precisely because of a reduced experimental and theoretical context. Given that alternative theoretical accounts of observed events were virtually non-existent in such cases and that even the existing ones were hard to come by, it was reasonable to pursue the strategy involving strong TD. It was reasonable to treat γ peaks as indicators of bound-state decays and their products. It is not clear what else we could and should have done in this experimental and theoretical environment.

At this point, the analogy between the toy case and experiments in particle physics breaks down. In the toy case, the substance we experiment with is highly structured; thus, we can fairly easily conceive of alternative hypotheses on the relation between decrease in volume and rise in temperature. The quantity of the substance, its physical state, the nature of the mixture with other substances (e.g. with the glass container) and so on, can all conceivably complicate the relation between the expansion of the substance and the rise in temperature, thereby undermining the reliability of the judgment that the expansion of mercury is the measure of the temperature of the heated substance.

But we do not find anything analogous in particle physics. *The tested objects and their properties are structured in extremely reduced ways that make conceiving alternative theoretical explanations almost impossible.* Fewer assumptions are at work. As Chang tells us, "Tightening the circle, in the sense of involving fewer assumptions" can make the "the refutation more decisive" (Chang 2001, 284) and thus constitutes a more severe test. The complexity of such an experimental context consists, rather, in dealing with the staggering numbers of particle interactions – and this raises a very different epistemological problem (Perovic 2011).

At any point in the analysis, physicists can give up on the preliminary phenomenon, never being sure whether it is a genuine natural process, an artifact or a suggestive background noise, if they do not find an aspect that is closely connected with the *Tt*. There is some, very limited freedom of interpretation within the theoretical framework (e.g. different calculations of possible particle paths based on the conservation of momentum) but given how difficult the framework is to come by and given the nature of the experimental environment, we do not deem it significant, or even an anomaly, without relating it closely to the framework. When 3.1 and 3.6GeV peaks appeared initially in a theoretical vacuum as preliminary phenomenon of some interest, deliberation on whether this constituted a background noise or artifact could not have been started or finalized without *Ta*.

Thus, the strategy increasingly minimizes, if not entirely eliminating, what Collins (1985) calls the "experimentalist's regress" (facts can be generated by 'good" instruments, while instruments are deemed "good" if they can produce the facts). The theory and instruments are increasingly mutually controlling and thus raise each others' power of prediction and accuracy. *To* and *Tt* increasingly cohere. Given the experimental environment, then, the strong TD analogous to the toy case does not seem detrimental or epistemically problematic, as the circularity of inferences seems vicious in a way analogous to the viciousness of the circularity in the toy case only at the initial stage of the experiment and when the experimental environment and the experimental strategy are not considered. It is possible that strong TD reflects epistemic limits of particle physics in such cases. But it might be perfectly epistemically satisfying, given the experimental context, that *the strategy employed incrementally minimizes strong TD as it widens* *the evidence loop by requiring a more elaborate phenomena of To to cohere with increasingly more detailed predictions of Tt.*

## 2. Weak theory-drivenness gives rise to a socio-epistemological problem

Finally, is the weak TD discussed by Woodward as epistemically unproblematic as he suggests? Let us imagine that an experiment is weakly theory-driven but not TL – it is performed because the physicists think the posits of *T* it tests are of the utmost importance for understanding the physical structure of the world but they have found ways to avoid TL of any sort, including the elimination of *Tt* as *To* and distinguishing *P* from the background noise, without the help of *Tt*. This type of TD merely provides the external motivation for the measurements. Ultimately, whether it is epistemically problematic depends on the context of the research.

Let us imagine that the above experiment is performed in a research landscape 1 (RL1) that involves the entire physics community and absorbs all available funding. The experiment consists of performing ever-more precise measurements concerning a particular posit of *T*. In an alternative scenario, in RL2 such *T*-motivated measurements are supplemented by limited resources invested in *T’*-motivated measurements. Now, uncontrolled weak TD in RL1 may turn out to be epistemically detrimental in the long run, as the RL2 has more chance, even though only slightly, of delivering a new experimental and theoretical insight. Meanwhile, in RL3, physicists take any theoretical suggestion including those they deem crackpot and distribute funding to test as many as they can. Is RL3 more likely to deliver valuable insights than either RL1 or RL2?

Of course, these are caricatures, but they have bearing on the real-world research landscape. Physicists normally test various theoretical approaches to the subject of interest, but the real question is *whether theoretical motivation is inescapable, to what extent it is a motivating factor, and whether it is epistemically benign as such*. Scientists are very strategic in their choices of theories and hypotheses they test, although such choices are never easily made in complex experimental environments such as particle physics and can have real impact on the fruitfulness of both theoretical and experimental work.

It is often true that “the researcher would not have attempted to measure *P* if he did not regard *T* as a serious possibility” (Woodward 2011, 177). This particular theoretical approach to experimentation is very apparent in particle physics where the largest portion of resources is often channeled into a few long-running experiments.

At the opposite end of the spectrum of RL1, experiments check hypotheses and theories but lack sufficiently convincing theoretical reasons for doing so. In fact, the initial discovery of *Ψ* peaks was a result of explorative analysis based on theoretically unmotivated hunches during the exploration of narrow energy domains (Goldhaber 1997, 66). And the possible epistemic problem that weak

TD of RL1 might lead to was the reason for the exploratory experimentation in the first place. It is also why LHC has been testing SUSY and other alternative approaches, as well as the Standard Model.

It is difficult to recognize whether weak TD is epistemically problematic in a given context – this requires serious socio-epistemological analysis – but it is fairly easy to see why theoretical motivation for experimentation can be epistemically problematic. This is not a trivial problem; it is often difficult to widen the range of alternatives because of limited resources or the dominance of a particular theoretical approach. Weak TD is an acute epistemic threat – as acute as the threat of strong TD and other forms of TL. Therefore, strategies must be developed to avoid it.

A more general point relates to the epistemological relevance of weak TD. Debates on TL are predicated on an assumption that a single observation has an epistemic status *per se*, in isolation from observations in other experiments. Hence, debates on TL typically focus on the relationship between such isolated observations and theory. This assumption might be adequate for the purposes of partial analysis of the experimental process that focuses on the relationship between tested theory and experimental results; but the analysis of the experimental process must be complemented by studying more general contexts and considering the impact of theoretical motivations, as experimental results are open to the threat of weak TD as much as they are open to different forms of TL.

## References:

Biddick C.J. et al. 1977. “Inclusive γ-Ray Spectra from *Ψ* (3095) and *Ψ*’ (3684) Decays.” *Physical Review Letters* 38 (23): 1324–1327.

Bogen, J. and J. Woodward 1988. “Saving the Phenomena”. *The Philosophical Review* 97: 303–352.

Braunschweig W. et al. 1975. “Observation of the two photon cascade 3.7 → 3.1 + γγ via an intermediate state Pc.” *Physics Letters* 57B (4): 407–412.

Chalmers A. 2003. “The Theory-Dependence of the Use of Instruments in Science.” Philosophy of Science 70 (3): 493–509.

Chang H. 2001. “Spirit, air, and quicksilver: The search for the “real” scale of temperature.” *Historical Studies in the Physical and Biological Science* 31 (2): 249–284.

Collins, H. 1985. *Changing Order: Replication and Induction in Scientific Practice*. London: Sage Publications.

Duhem P.M.M. 1906. *La Théorie physique: son objet, sa structure*. Translation:

*The Aim and Structure of Physical Theory*. Princeton: Princeton University Press, 1991.

Feldman G.J. et al. 1975. “*Ψ* (3684) Radiative Decays to High-Mass States.” *Physical Review Letters* 35 (13): 821–824.

Franklin A et al. 1989. “Can a Theory-Laden Observation Test the Theory?” The *British Journal for the Philosophy of Science* Vol. 40 (2): 229–231.

Franklin A. 1997. “Calibration.” Perspectives on Science 5.1 (Spring): 31–80.

Gilman F. 1985. “The November Revolution.” *Slac Beam Line* (January): 3–5.

Goldhaber G. 1997. “From the Psi to Charmed Mesons: Three Years with the SLAC_LBL Detector at SPEAR.” In L. Hoddeson, ed. *The Rise of the Standard Model*. Cambridge: Cambridge University Press. 57–78.

Hacking I. 1983 *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Krumhansl J.A. and George L. Trigg. 1974. “Editorial: Publication of a New Discovery.” *Physical Review Letters* 33: 1363.

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.

Schindler S. “Bogen and Woodward’s data-phenomena distinction, forms of theory-ladenness, and the reliability of data. *Synthese* 182: 39–55.

Woodward J. 2009. “Data, Phenomena, and Realiability.” *Philosophy of Science* 67: S163-S179.

Woodward J. 2011. “Data and Phenomena: A Restatement and Defense.” Synthese 182 (1): 165–179.