THE INTERDEPENDENCE OF THE CORE, THE HEURISTIC AND THE NOVELTY OF FACTS IN LAKATOS'S MSRP

Elie G. ZAHAR*

* Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH.

BIBLID [0495-4548 (2001) 16: 42; p. 415-435]

ABSTRACT: In this paper I try to explain why Lakatos's (and Popper's) conventionalist view must be replaced by a phenomenological conception of the empirical basis; for only in this way can one make sense of the theses that the hard core of an RP (Research Programme) can be shielded against refutations; that this *metaphysical* hard core can be turned into a set of guidelines or, alternatively, into a set of heuristic metaprinciples governing the development of an RP; and that a distinction can legitimately be made between novel predictions and facts to which a theory might have been adjusted *post hoc*. Two basic metaprinciples are finally examined: the (conservative) Correspondence Principle and various (revolutionary) symmetry requirements; both of these heuristic devices illustrate the fundamental role which, according to Lakatos, mathematics plays in the progress of empirical science.

Keywords: Hard Core, Heuristics, Novelty of Fact, Lakatos, MSRP.

CONTENTS

One can learn more about Lakatos's philosophy through comparing it with Popper's position than by reading a comprehensive account like the one given in 'Falsification and the Methodology of Scientific Research Programmes' (MSRP. See Lakatos 1978, pp. 8-101). It will be shown that although Lakatos rightly opposed Popper's refusal to admit a 'whiff' of inductivism into his philosophy, neither he nor Popper is entitled to appeal to any principle connecting 'verisimilitude' with future empirical success. Let us note that already in 1931 Popper had openly subscribed to such an inductive principle (Popper 1979, p. 155). It will however be made clear that given the conventionalist view of basic statements which Lakatos shares with Popper, neither author can provide a *rationale* for the acceptance of any such 'whiff' of inductivism.

An attempt will be made to vindicate the following thesis: provided the conventionalist view be replaced by a phenomenological interpretation of observation reports, one can offer not only a 'rationale' for induction, but also solutions to a number of problems besetting both Popper's and Lakatos's positions. It can then be shown that Popper's demarcation criterion has a naturalistic basis which is indispensable to MSRP; that the hard core of a research programme can be protected against refutation because it is metaphysical in an objective sense; that part of this core can consequently be reinterpreted as the positive heuristic of the programme, i.e. as a metatheoretical constraint which governs the construction of new hypotheses; finally, that the very notion of ad-hocness presupposes the independence of factual statements from theoretical assumptions. Thus the thesis that observation reports, being theory-laden, can be accepted only by convention, is the real villain of the piece. Its removal will neutralize the unacceptable scepticism inherent in the Popper-Lakatos position; and it will also become clear why Lakatos's MSRP constitutes a complement to, rather than a corrective of, Popper's 'logic' of discovery.

1. Conventionalism and Phenomenology

As already mentioned, Lakatos shares Popper's view that *all* synthetic propositions are theory-laden, hence can never be definitively established. He concludes that hypotheses cannot be falsified in the sense of being known to be false. We shall therefore have to examine Popper's conception of the nature of observation.

Popper's demarcation criterion might lead one to believe that the impossibility of verifying universal propositions flows from the physical impossibility of performing infinitely many tasks; while the possibility of falsification rests on the feasibility of observing finitely many items in order to decide the truth-value of a potential falsifier. Popper however goes out of his way to assert that observation -qua perceptual process- bears no epistemological relation to basic statements. Sense-experience may *motivate* us to accept a falsifier but it provides *no reason* for doing so (Popper 1959, p. 105). Basic statements can and should normally be objective propositions about the world. They are not only theory-laden but moreover consist of low-level hypotheses containing dispositional terms; their verification is therefore impossible and has certainly nothing to do with any observer's perceptions. A basic statement is acceptable iff there is *intersubjective agreement* about its truth-value. Such acceptance can be revised and is therefore non-dogmatic; but when carefully examined, such a "revision" simply means that when taken in conjunction with other hypotheses, the previously accepted proposition might be rejected as a result of a fresh agreement about some other "falsifier". At no point does this potentially infinite process involve truly *epistemological* considerations, i.e. considerations linking the truth-value of the basic statement either to the act of observation or to that of reaching a consensus.

This view of the empirical basis threatens to destroy the presumed asymmetry between verification and falsification; for we now have one proposition, dubbed "theory H", confronting another proposition which we call "basic statement B". Despite these labels, all we can assert is the logical incompatibility of H and B. Thus the relationship between H and B is perfectly symmetric. Popper admittedly describes B as a *low-level hypothesis,* but this is of no great help; for since H neither entails nor follows from B, the levels of H and B are not comparable except in some vague intuitive sense. More seriously: nothing in Popper's analysis tells us that the hypotheses impregnating B are less risky than H. All we can claim is that over the past 400 years or so, science has pursued a largely empiricist policy; i.e. scientists have reached agreement over propositions having superficially the same form as "B" rather than over statements resembling "H". But we can provide no explanation of the social success of this empiricist strategy. Worse still: we cannot explain why this policy issued in great technological breakthroughs. The latter are taken to depend on the truth or on the approximate truth of some consequences of our theories; or at least on our having, by and large, rejected false hypotheses rather than accepted false basic statements. Yet truth-considerations have so far played no role in Popper's methodology. Sustained technological progress thus becomes an ongoing miracle.

Let us call Popper's view of basic statements the conventionalist view or conventionalist thesis. This thesis is clearly sociological and it is so because a transcendental critique has, in Popper's opinion, shown its psychologistic rivals to be incompatible with successful scientific practice.

Let me now describe a different position, which I propose to defend; namely the phenomenological view of observation. A singular sentence p' will be said to express a level-0 proposition if p' describes, in the first person, the immediate contents of some speaker's consciousness. Clearly, the truth-value of such a proposition is *logically* independent of all transcendent states-of-affairs, i.e. of all events occurring outside the speaker's consciousness (See Watkins 1984, pp. 79-84). This is why level-0 statements are also referred to as immanent, autopsychological or phenomenologically reduced propositions. Thus according to the phenomenological view of observation, all basic statements can, in the last analysis, be reduced to level-0 sentences; where "be reduced" does not mean: be made logically equivalent to. More precisely: let p be a sentence describing some physical process, e.g.: the current is on; and let p' be the proposition, expressed in phenomenological terms, that a certain spot is seen to move; we shall take p' to describe a sequence of perceptions correlated with p. It can be said that p' is the result of a phenomenological reduction which eliminates from p all references to a mind-independent reality. According to the phenomenological view, all experimental results can be expressed by sentences of the same form as p' and can thus effectively verify or falsify such sentences. The phenomenological thesis presupposes the existence of psycho-physical laws A such that: $\vdash(A\Rightarrow(p\Leftrightarrow p'))$. A will contain clauses about the reliability of the instruments used during some experiment and about the observer's mental and physical state (e.g. the precondition that he is not colour-blind). Let us note that A should be regarded as undergoing the same test as the core hypothesis H to which A is appended. A is therefore contingent, so that p is, not logically but materially, equivalent to p'.

In order to test a complex system S one of whose components is A, one extracts from S, taken together with some boundary conditions p, a prediction q. Thus: $\vdash(S\Rightarrow A)$ and $\vdash(S\Rightarrow(p\Rightarrow q))$. (The core theory in S will generally entail $p\Rightarrow q$ without the help of A). To p and q correspond autopsychological sentences p' and q' such that: $\vdash(A\Rightarrow(p\Leftrightarrow p'))$ and $\vdash(A\Rightarrow(q\Leftrightarrow q'))$. It follows that: $\vdash(S\Rightarrow(p'\Rightarrow q'))$. (Note that in order to obtain this last implication, we need only assume $\vdash(A\Rightarrow(p'\Rightarrow p))$ and $\vdash(A\Rightarrow(q\Rightarrow q'))$.

S will be refutable by propositions like $p'\wedge\neg q'$ whose truth-value can be effectively decided. Note that both Lakatos and Popper are prepared to accept $p\wedge\neg q$ as a potential falsifier whereas, according to the phenomenological thesis, only statements of the form $p'\wedge\neg q'$ are admissible as observation reports. But contrary to Popper's view, the phenomenological thesis implies that there is practically no possibility of error at the autopsychological level.

Using Popper's own transcendental method, I now propose to show that scientific praxis closely conforms to the phenomenological thesis; moreover, that the latter turns out to contain the valid aspects, while avoiding most of the defects of its conventionalist rival. The phenomenological view will furthermore be seen to entail that repeatability and intersubjective testability are desirable features arising from the nature of the Duhem-Quine problem; but that these desirerata should not be built into the *definition* of a potential falsifier.

Let me start by mentioning an example to which a transcendental critique can be applied. In 1926, D. C. Miller claimed to have performed a variant of the Michelson-Morley experiment, thereby establishing a result which contradicts Special Relativity (henceforth referred to as SR). The experiment was repeated, but no results similar to Miller's were obtained. According to the conventionalist view, Miller's allegedly factual statements have no objective value, hence need not be taken seriously by science. But far from ignoring Miller's claim, M. Born inspected Miller's experimental set up and concluded that the instruments used by Miller were unreliable (Einstein and Born 1969, pp. 107-128). The phenomenological thesis, as opposed to its conventionalist rival, provides a rationale for Born's attitude. Let e be the result announced by Miller and let e' be its autopsychological counterpart. We have: $\vdash[(R\wedge A)\Rightarrow(e\Leftrightarrow e')]$ and $\vdash(R\Rightarrow\neg e)$; hence: $\vdash[(R\wedge A)\Rightarrow\neg e']$; where R denotes SR, and A some complex hypothesis asserting, among other things, that the instruments used by Miller were reliable. Born's decision can be rationally explained only if e' is taken to be true and if Born intended to save R by refuting A. It is obvious that the Duhem-Quine problem plays a central role in the analysis of complex experimental situations; that repetitions and intersubjective agreements are moreover intended to exclude the likelihood of random factors falsifying some of the auxiliary assumptions.

Let me analyse the formal aspect of the above test-structure by abstracting from the particular meanings of R, A, e and e'.

If $\neg e'$ is experimentally verified, then $R\wedge A$ is corroborated; so we can provisionally accept $R\wedge A$ and hence also $\neg e$, since $\vdash[(R\wedge A)\Rightarrow(\neg e\Leftrightarrow\neg e')]$. In these circumstances, scientists do not normally bother to formulate $\neg e'$ explicitly; they short-circuit the autopsychological proposition $\neg e'$ and affirm only its realist counterpart $\neg e$, thus concluding that $R\wedge A$ has been confirmed. This omission of $\neg e'$ gives rise to the impression that $\neg e$ is the only relevant basic statement; but should a refutation occur, i.e. should e' be verified, then the situation might change dramatically. There takes place, on the one hand, a retreat towards the phenomenological kernel e' which, being warranted by the experiment, is immune to doubt; but the "objective" proposition e is no longer accepted since the material equivalence between e' and e follows from a *falsified* hypothesis, namely from $(R \wedge A)$. On the other hand, one tries to identify all the premises used in the derivation of $\neg e'$ from $(R \wedge A)$; i.e. one seeks to determine the distinct components of both R and A. This operation serves to identify all the hypotheses which might be incriminated by the truth of e'; i.e. it determines the extent of the Duhem-Quine problem. Given the logical force of refutations, it is highly desirable to repeat the experiment in order to know whether chance-like events might not have falsified the auxiliary hypothesis A, thus accounting for the result e'. Generally speaking: for any given e', A will be expressible in the form: $A \equiv (A' \wedge A_1)$, where A_1 is a specific proposition about the reliability of the instruments used in some experiment and about the observer's mental and physical health; A_1 may also express the condition that during the experiment in question, only the factors explicitly mentioned by R came into play. Thus: $\vdash [(R \wedge A) \Leftrightarrow (R' \wedge A_1)]$, where $R' \equiv (R \wedge A')$. Having the experiment repeated at different times, in different places and by different observers comes down to modifying A_1 into A_2, then into $A_3, \ldots$, finally into A_n; where $A_1, \ldots, A_n$ describe n independent states-of-affairs. At this point, two possibilities have to be examined:

(a) If each of the n experiments yields a refutation, then in order to rescue R', we have to assume the falsity of A_1 and of $A_2 \ldots$ and of A_n; according to both Duhem and Popper, it can reasonably be conjectured that the fault lies with R' (Popper 1965, p. 243; Duhem 1914, part 2, chapter 2). But Popper did not admit that such "reasonableness" rests, as it obviously does, on the following intuitive probabilistic argument: if, despite all the negative outcomes, we decided to adhere to R', then each of $A_1, \ldots, A_n$ must be considered false; which yields the unique asssignment (t,f,...,f) of truth-values of $(R', A_1, \ldots, A_n)$, where t and f denote the True and the False respectively. Should we however be prepared to give up R', then each of $A_1, \ldots, A_n$ could be either t or f, thus yielding 2^n assignments compatible with all experimental results. Assuming $R', A_1, \ldots, A_n$ to be mutually independent, the chances are that R' is false.

$A_1, \ldots, A_n$ may of course share a false kernel G such that $R' \wedge G$ is testable; which might account for the successive refutations of $(R' \wedge A_1), \ldots, (R' \wedge A_n)$. This is why a caveat of mutual independence has to be entered. Be it as it may, the above –admittedly crude– piece of reasoning provides a rationale for our feeling that, barring miracles, R' must be the culprit.

(b) If $R' \wedge A_1$ is falsified (by e') but all of $(R' \wedge A_2), \ldots, (R' \wedge A_n)$ are confirmed, then, following Einstein, we could conclude that e' must have refuted, not R' but the auxiliary assumption A_1. Given the crucial nature of the experiment, Born however decided to go further than Einstein: he effectively refuted the premise A_1, more particularly: the clause in A_1 stating that the instruments used by Miller were reliable. Be it as it may, both Einstein and Born agreed that Miller truthfully reported what he saw; i.e. they took the truth of the autopsychological statement e' for granted. Thus, repeating the same experiment was not designed to confer objective status on e' but to deal more effectively with the Duhem-Quine problem.

Admittedly, the mere identification of a faulty auxiliary assumption like A_1 does not constitute great progress. Logically speaking, Born saved SR without thereby explaining Miller's results; for even if $\neg A_1$ were established, $\vdash[(R'\wedge A_1)\Rightarrow\neg e']$ need not entail $\vdash[(R'\wedge\neg A_1)\Rightarrow e']$; so e' remains unexplained. To try *directly* to explain e' might moreover be counterproductive and even lead the scientist astray; he may be taken out of the domain of physics into that of psychology; or he may set himself the impossible task of identifying intrinsically random factors. The situation could however change in a dramatically revealing way. Suppose that e' refutes $R\wedge A$ but that we have been unable either to refute A directly or to reproduce e' [i.e. to falsify $(R'\wedge A_2)$, or $(R'\wedge A_3)$,..., or $(R'\wedge A_n)$]. Suppose further that, by a route independent of e', we subsequently constructed a theory $R^*\wedge A^*$ which yields e', explains why e' could not be reproduced and is otherwise observationally equivalent to $R\wedge A$. A^* might e.g. take account of meteorological conditions which skew some experimental results; and we might retrospectively realize that very rare and abnormal weather conditions obtained at the time when the experiment yielding e' was carried out. Although it describes a unique event, e' would clearly be taken to confirm $(R^*\wedge A^*)$ against $(R\wedge A)$. Despite not being reproducible, e' would thus have undermined $(R\wedge A)$. Note that corroboration is logically weaker than refutation and remains provisional. Whereas falsification is irreversible, the confirmation of $(R^*\wedge A^*)$ by e' does not protect $(R^*\wedge A^*)$ against being refuted by the next test. This is why we *risk very little* by regarding the unique event e' as having corroborated $(R^*\wedge A^*)$. It is of course preferable to repeat even a confirming experiment in order to reduce the probability of its outcome being due to random factors; but repetition plays a less important a role in the case of corroboration than in that of empirical refutation.

Let us finally note that the truth of autopsychological propositions is not based on experiencing any feelings of conviction, but on purely phenomenological analyses. Every act of knowledge involves the presence of a subject and of an object. There is a possibility of error as long as the object transcends the subject; for the latter might then be intrinsically incapable of adequately reflecting the attributes of the object. The subject can however carry out a phenomenological reduction which eliminates all *direct* references to any reality external to the the mind. The object will thus coincide with, or become part of the subject: a privileged access is established from the one to the other, so that we can apprehend both the meanings and the intended referents of our (level-0) assertions. Being and knowing are fused into one activity; with the result that all likelihood of error vanishes.

This phenomenological analysis is definitely not reducible to the *psychology* of subjective experiences of certainty. In the course of a dream, we might experience a more acute feeling of conviction about some *allegedly* external state-of-affairs than, later, about its phenomenologically reduced correlate; but phenomenological analysis demonstrates that our second feeling of certainty is well-founded whereas the first was not. As a result, we may of course develop a feeling of conviction about the truth of certain level-0 propositions. Mathematical proofs similarly give rise to feelings of conviction concerning theorems which initially sounded implausible. This does not mean that logic or mathematics is psychologistic. The same applies to phenomenology.

2. Ad-hocness and Parameter-Adjustment

It is well-known that ad-hocness depends on the notion of the *novelty of facts*; which is why the latter has now to be examined in some detail. It has of course to be agreed that the discovery of every new (type of) fact is the discovery of a novel fact. The converse however need not hold; for if we simply equate novelty with *temporal* novelty, we are driven into a paradoxical situation. We should e.g. have to give Einstein no credit for explaining the precession of Mercury's perihelion because this fact had been recorded long before General Relativity was proposed. We should similarly have to say that Michelson's results did not confirm Special Relativity and Galileo's experiments on free fall did not support Newton's gravitational theory. Lakatos, who insisted on methodologies being consistent with the working scientists' *singular* value-judgments, was aware of this difficulty. He tried to avert it through saying that in the *light of a new theory,* known facts 'turn into' novel ones. He was thereby drawing a natural conclusion from Popper's thesis that observations are theory-laden; so a change of theory may well result in a change of fact.

This modified notion is however open to the following fatal objection. Any theory is a nexus of propositions connecting different terms and predicates. We can always define the properties of any physical entity through the relations which it bears to other notions within any theory; so that all of the latter's empirical consequences become novel predictions. This is clearly unacceptable, for it fails to distinguish between an intuitively ad-hoc hypothesis and a genuinely explanatory one. This is why Lakatos later subscribed to what came to be known as the Z-W (Zahar-Worrall) definition of novelty, which can roughly be formulated as follows: a fact will be considered novel with respect to a hypothesis if *it was not used in the construction of the hypothesis.* (See Zahar 1989, p. 16). Note that this definition entails the rejection of the conventionalist thesis; for the latter implies that a fact to which a law is adjusted could be modified by the law and hence cease to be the fact to which it was adapted; which is patently absurd.

Let us now give a more formal account of what it is for a hypothesis to be adjusted to known results. Consider a hypothesis $H(a_1,...,a_n)$ containing the free parameters $a_1,...,a_n$; and let $e_1,...,e_m$ be a sequence of empirical results. Since $a_1,...,a_n$ have not yet been fixed, we might initially be ignorant as to whether or not $H(a_1,...,a_n)$ subsumes $e_1,e_2,...,e_n$. Through putting $a=(a_1,...,a_n)$ and $E\equiv(e_1\wedge...\wedge e_m)$, we can, without any loss of generality, assume that n=m=1. We now face the task of so determining a that: [i] $H(a)$ is consistent, and: [ii] $\vdash(H(a)\Rightarrow E)$.

Since scientists take E for granted while provisionally accepting $H(a)$, they will postulate the conjunction $H(a)\wedge E$, from which they then draw conclusions as to the possible values of a. They typically succeed in determining a set K such that: $\vdash[(H(a)\wedge E)\Rightarrow(a\in K)]$. I.e. $\vdash[(a\notin K)\Rightarrow(H(a)\Rightarrow\neg E)]$. Unless $a\in K$, $H(a)$ will thus be refuted by the facts E. But scientists want E to be not merely consistent with, but also explained by $H(a)$. So they will construct another set Q such that: $\vdash[(a\in Q)\Rightarrow(H(a)\Rightarrow E)]$. The next step usually consists in selecting some a_0 which satisfies $\vdash(a_0\in K\cap Q)$. Since $\vdash(a_0\in Q)$, we have: $\vdash(H(a_0)\Rightarrow E)$; i.e. $H(a_0)$ entails E. As for $a_0\in K$, it represents a consistency condition. More precisely: if $\vdash(a_0\in Q)$ but $\vdash(a_0\notin K)$, then by the above: $\vdash(H(a_0)\Rightarrow E)$ and $\vdash(H(a_0)\Rightarrow\neg E)$; which means that $H(a_0)$ is mathematically inconsistent. Thus, choosing $a_0\in K\cap Q$ comes down to constructing a (hopefully) consistent theory $H(a_0)$ which is specifically doctored to subsume the facts E.

In this situation, Popper would deny that H derives any support from E; for since E was known prior to $H(a_0)$, it formed part of the background knowledge B available before $H(a_0)$ was put forward; therefore $p(E,B)=1$. A fortiori: $p(E,H(a_0)\wedge B)=1$. Hence: $C(H(a_0),E,B)=_{Def.}(p(E,H(a_0)\wedge B)-p(E,B))/(p(E,H(a_0)\wedge B)+p(E,B)-p(H(a_0)\wedge E,B))=0$; where $C(H(a_0),E,B)$ is Popper's expression for the degree of the corroboration of $H(a_0)$ by E, given background knowledge B (Popper 1959, Appendix *9).

This argument must be rejected; for as explained above, it fails to demarcate theories which genuinely explain known facts from those merely adjusted post-hoc to these same facts. What matters is not the whole of B but only that part which was *actually used* in constructing $H(a_0)$; where the construction follows the logical pattern described above. The context of the discovery of a hypothesis H may therefore have a role to play in the 'justification' of H. It does not however suffice for a scientist to be *psychologically aware* of a factual result E in order for his conjectures H to become automatically adhoc with respect to E; he must furthermore have made some *objective use* of E in determining H.

Finally: even if $H(a_0)$ were adjusted, in the manner described above, to yield E, $H(a_0)$ could still receive some measure of support from E; since we might, for logical reasons, altogether fail to find any a_0 such that $H(a_0)$ consistently entails E. This will happen whenever, for some K and for all Q, $K\cap Q$ is empty, i.e. $K\cap Q=\emptyset$.There exist two limiting cases in which $K\cap Q=\emptyset$ holds; namely those in which either K or Q is empty. Concerning $K=\emptyset$: consider the fictitious example in which E refers to a single instance of the addition law of velocities and H(a) to SR, where the speed of light a is provisionally treated as a finite adjustable parameter. Then no matter how we vary a, we shall always have $\vdash(H(a)\Rightarrow\neg E)$. I.e. $\vdash[(H(a)\wedge E)\Rightarrow(a\in\emptyset)]$. We can therefore take $K=\emptyset$. Note that we can infer $(\exists x)H(x)\Rightarrow\neg E$ from $H(a)\Rightarrow\neg E$; i.e. $(\exists x)H(x)$ already constitutes a scientific theory since it possesses the potential falsifier E. As for $Q=\emptyset$, it represents the case where H(a) is compatible with, but also irrelevant to E; so no choice of a enables H(a) to decide E in a consistent manner. E.g., take H(a) to be an economic theory in which we vary some parameter a; and let E be the same statement as in the last example. Since economic hypotheses do not bear on any physical laws, there will be no value of a such that $\vdash(H(a)\Rightarrow E)$. Thus $Q=\emptyset$.

The above method of selecting $a\in K\cap Q$ can therefore break down. It follows that H(a) can legitimately derive some support from the mere existence of just one element of $K\cap Q$. Remembering that $a=(a_1,\ldots,a_n)$ and $E\equiv(e_1\wedge\ldots\wedge e_m)$, we conclude that the next -naturally unused- experimental result e_{m+1} might serve to test H(a) severely; for by hypothesis, H(a) will have been determined independently of e_{m+1}; so that if $\vdash(H(a)\Rightarrow e_{m+1})$, then H(a) will have explained e_{m+1} in a non-adhoc way.

Generally speaking, strong evidential support goes hand in hand with overdetermination: if any m experimental results suffice to fix the n parameters $a_1,...,a_n$ occurring in $H(a_1,...,a_n)$, then any number of facts in excess of m can either refute or else strongly confirm $H(a_1,...,a_n)$; for in this hypothetical case, any m data would uniquely determine $a_1,...,a_n$; while the remaining facts could turn out to be either incompatible with, or else subsumed by $H(a_1,...,a_n)$. In the latter case, $H(a_1,...,a_n)$ will have been genuinely corroborated.

In conclusion, we can say that Lakatos's and Popper's value-judgments, though broadly correct, are too coarse-grained to capture the scientists' intuitive notion of corroboration. Furthermore, even though H might have been doctored to yield E, such doctoring could in principle have proved impossible; which implies that since H has taken a risk –no matter how minimal– it is entitled to receive some measure of support from E.

3. Metaphysical Hard Core and Positive Heuristic

Lakatos claimed that only through a methodological decision does the hard core, or negative heuristic of a programme become metaphysical, that is: untestable (Lakatos 1978, p. 50). He was thereby being consistent both with himself and with Popper; for we have seen that according to the latter, the basic statements through which a scientific hypothesis is tested are stipulated by conventional decision. Having accepted the phenomenological thesis, we are however in a position to assert that independently of any stipulation, every proposition either *is* or *is not* observational. If it is to be held on to come what may, the hard core must be irrefutable, i.e. metaphysical, in the absolute sense.

It is my claim that certain components of the hard core have prescriptive counterparts which can be translated into meta-stataments about scientifiuc hypotheses; for an ontological thesis imposes constraints on the form of every theory which purports to be a true description of reality. Hence *metaphysics can be taken to possess prescriptive import.* It is moreover imperative that such prescriptions be translated into propositions, or rather into meta-propositions; especially if we want the heuristic to operate *deductively* by providing premises for the logical determination of theories. Thus, we have the following scheme:

Metaphysics (Hard Core) → Prescriptions → Meta-statements (Positive Heuristic).

The heuristic may therefore reflect certain aspects of the hard core; so that the distinction between negative and positive heuristic is not as absolute as Lakatos seemed to think. Let me give two examples. Part of the ether programme's hard core was the thesis that physical reality consists of one medium possessing electro-mechanical properties; whence the prescription that theories should derive all phenomena from the states of the ether. The correponding meta-principle is that all laws of nature contain only the concepts of position, time, mass and charge. As for SR, it is based on the metaphysical proposition that no privileged inertial frame exists; which leads to the prescription that all hypotheses should assume the same form in all inertial frames. The corresponding meta-statement is that all laws of nature are Lorentz-covariant.

We shall now examine two fundamental heuristic principles –one conservative, the other revolutionary– which have dominated the development of the physical sciences.

4. The Correspondence Principle

Science is classically portrayed as evolving through a succession of improved approximations to the truth. This cumulative view of progress was challenged by the following observation: at the ontological level, the development of science displays an alarming degree of referential discontinuity: as we move from one hypothesis to the next, the same-sounding word is often found to denote, or rather to conflate, two fundamentally different types of entity. For example: the term 'electromagnetic waves' referred, first to the mechanical states of the ether; secondly, to two oscillating fields deprived of any material support; and finally, to massless particles called 'photons'.

The cumulative view can however be rescued by an appeal to two essential aspects of the development of empirical science; and these aspects turn out to be closely linked. First, it is generally agreed that, at the practical level, our control over physical nature has steadily grown. This progress was secondly taken to have been achieved by the systematic application of a continuity requirement, namely the Correspondence Principle which demands that old laws be either logical consequences or limiting cases of new hypotheses. This essentially mathematical form of continuity also explains why a new theory inherits the degree of confirmation of its predecessors. Let us however note that identity is the most perfect form of continuity; but in the midst of a scientific revolution, identity is the last condition we would wish to impose. In well-specified areas, the new hypothesis is expressly meant to differ markedly from the old one: a new law may be required to avoid, or rather to overcome the known refutations of previous conjectures.

This tension between continuity and saltation can be resolved by means of M. Redhead's notion of one scientific system being *imbedded in* –rather than forming *a substructure of*– another one (Redhead 1995, chapter 4). The following examples will clarify this important distinction. Even though the complex plane **C** extends and hence imbeds the real line **R**, the former does not mirror *all* the properties of the latter, *nor vice-versa;* for the linear ordering of **R** cannot be naturally extended to **C** while **C**, but not **R**, is algebraically closed. Another example is provided by hyperbolic geometry; the latter possesses a model in **C** and can consequently be taken to be imbedded in **C**. It would however be absurd to claim that hyperbolic geometry, which is *non-Euclidean,* constitutes a substructure of the Euclidean plane.

Let me now demonstrate that the possibility of reconciling the gradualist with the revolutionary aspects of scientific progress rests on a topological result; namely on the Heine-Borel theorem which tells us that all closed and bounded subsets of $\mathbf{R}^n$ are compact, where n is an arbitrary positive integer. This illustrates Lakstos's often repeated thesis that mathematics plays a crucial role in the progress of empirical science (Lakatos 1978, p. 51).

Consider the hypothesis:

(1) $(\forall v,a,b)[Q(v,a,b)\Rightarrow(h(v,a,b) = 0)]$, which is taken to have been superseded by:

(2) $(\forall v,a,b)[Q(v,a,b)\Rightarrow(H(\lambda_0,v,a,b)=0)]$, where λ_0 denotes a small constant.

Replacing λ_0 by the variable parameter λ, we shall assume that $H(\lambda,v,a,b)$ and $h(v,a,b)$ are continuous functions of $<\lambda,v,a,b>$ and of $<v,a,b>$ respectively.

Suppose that $[h(v,a,b)=0]$ has been strongly corroborated in some region Δ, i.e. that this law has withstood a number of severe tests in Δ. This number being necessarily finite, we can assume, without any loss of generality, that Δ is of the form:

(3) $\Delta=V\times A\times B$, where V, A and B are bounded closed sets of real numbers.

A could e.g. be taken to be the closed interval [a',a"], where a' and a" are respectively the minimum and the maximum measured values of a. Similar considerations apply to V and to B. When restricted to Δ, the laws (1) and (2) assume the forms:

(4) $(\forall v,a,b)[(Q(v,a,b)\wedge(<v,a,b>\in\Delta))\Rightarrow(h(v,a,b)=0)]$, and:
(5) $(\forall v,a,b)[(Q(v,a,b)\wedge(<v,a,b>\Delta))\Rightarrow(H(\lambda_0,v,a,b)=0)]$ respectively.

In view of the approximate character of all measurements, we can take (4) to be observationally indiscernible from:

(6) $(\forall v,a,b)[(Q(v,a,b)\wedge(<v,a,b>\in\Delta))\Rightarrow(|h(v,a,b)|\leq\mu_0)]$, where μ_0 is some fixed positive real.

(6) is generally taken to hold not only for the values of v, a and b which have actually been measured, but also for all intermediate values; i.e. throughout the domain Δ. Thus, (6) is regarded as true.

In one of its forms, the Correspondence Principle requires that:

(7) for any $<v,a,b>\in\Delta$, $H(\lambda,v,a,b)\rightarrow h(v,a,b)$, as $\lambda\rightarrow 0$

Given the continuity of H, we also have: $H(\lambda,v,a,b)\rightarrow H(0,v,a,b)$, as $\lambda\rightarrow 0$. By the uniqueness of the limit, (7) entails that:

(8) $h(v,a,b)=H(0,v,a,b)$, for all v, a and b

Let us describe the way in which the new theory H accounts for the empirical success of h in Δ. We provisionally accept (5), from which (6) will be deduced. Since λ_0 is small, there exists a real number $k>0$ such that:

(9) $\lambda_0\in[-k,+k]$.

Let the parameter λ be confined to the closed interval [-k,k]. By the continuity of H and the compactness of $[-k,k]\times V\times A\times B=[-k,k]\times\Delta$, the following theorem about uniform continuity holds good. For any $\mu>0$, there exists an $\eta(\mu)$, dependent only on μ, such that: $|H(\lambda_1,v_1,a_1,b_1)-H(\lambda_2,v_2,a_2,b_2)|\leq\mu$, for all $<\lambda_j,v_j,a_j,b_j>\in[-k,k]\times\Delta$ $(j = 1,2)$ satisfying:

$\sqrt{[(\lambda_1-\lambda_2)^2+(v_1-v_2)^2+(a_1-a_2)^2+(b_1-b_2)^2]}\leq\eta(\mu)$. In particular:

(10) $|H(\lambda,v,a,b)-H(0,v,a,b)|\leq\mu$, i.e. $|H(\lambda,v,a,b)-h(v,a,b)|\leq\mu$,

for all $<\lambda,v,a,b>\in[-k,k]\times\Delta$ such that $|\lambda|\leq\eta(\mu)$.

Taking $\mu=\mu_0$ and $\lambda=\lambda_0$, we conclude that:

(11) for all $<v,a,b>\in\Delta$, $|H(\lambda_0,v,a,b)-h(v,a,b)|\leq\mu_0$, provided $|\lambda_0|\leq\eta(\mu_0)$.

Suppose that the numerical value λ_0 actually satisfies $|\lambda_0|\leq\eta(\mu_0)$. Inequality (11) will then be true for all $<v,a,b>\in\Delta$. Now assume that $[(Q(v,a,b)\wedge\wedge(<v,a,b>\in\Delta)]$. By (5): $H(\lambda_0,v,a,b)=0$. By (11): $|h(v,a,b)|\leq\mu_0$; whence (6), which is observationally equivalent to (4). I.e. the truth of the new theory accounts for the confirmation of the old one in Δ.

There is another form of the Correspondence Principle which actually possesses greater heuristic power than (7). It requires that for any $<a,b>$, we have:

$|H(\lambda,v,a,b)-h(v,a,b)|\rightarrow 0$ as $\lambda v\rightarrow 0$. Fixing v and letting $\lambda\rightarrow 0$, we obtain:

$|H(\lambda,v,a,b)-h(v,a,b)|\rightarrow 0$, i.e. $H(\lambda,v,a,b)\rightarrow h(v,a,b)$, as $\lambda\rightarrow 0$. Thus, we retrieve condition (7). But we can alternatively fix λ at $\lambda=\lambda_0$ and let $v\rightarrow 0$. This yields:

$|H(\lambda_0,v,a,b)-h(v,a,b)|\rightarrow 0$ as $v\rightarrow 0$. By continuity, we must also have:

$|H(\lambda_0,v,a,b)-h(v,a,b)|\rightarrow|H(\lambda_0,0,a,b)-h(0,a,b)|$ as $v\rightarrow 0$. Hence:

(12) $|H(\lambda_0,0,a,b)-h(0,a,b)|=0$. That is: $H(\lambda_0,0,a,b)=h(0,a,b)$ for all $<a,b>\in A\times B$.

From (2) we infer:

(13) $(\forall a,b)[Q(0,a,b)\Rightarrow(H(\lambda_0,0,a,b)=h(0,a,b)=0)]$.

In other words: for v=0, the new law coincides with the old one. As an example, we can cite the case of STR, where: $\lambda_0=(1/c)$, c=velocity of light, and v=speed of the mobile. It is assumed that as $(v/c)\rightarrow 0$, the relativistic equation of motion 'tends' to the Newtonian one; where both v and 1/c are now treated as variable parameters. It follows from the above that the classical law $\vec{f}-m\vec{a}=0$ must strictly hold for v=0. This desideratum, together with the constraint of Lorentz-covariance, enabled Planck to establish the law $\vec{f}=d[m\vec{v}/\sqrt{(1-v^2/c^2)}]/dt$. (See Zahar 1989, chapter 7).

Let us repeat that the above considerations depend crucially on the Heine-Borel theorem, which applies to all closed bounded subsets of $\mathbf{R}^n$. This result leaves open the possibility for the two hypotheses, the old one and the new, to diverge widely outside D; for uniform continuity generally holds only over *compact* subsets of $\mathbf{R}^n$.

5. The Regulative Role of Symmetry Principles

We have seen that both continuity *and* divergence are of the essence of scientific revolutions; but only continuity has so far been shown to possess any heuristic clout. Referring back to the previous section, suppose that (1) has been refuted outside Δ. If no further information is provided, then all we can do is modify (1) in some ad-hoc way by restricting it, say, to the domain Δ. I.e., instead of (1), we could adopt (4) as our basic theory. This method therefore consists in adding extra qualifications to the antecedents of existing laws. But since (4) is logically weaker than (1), falsification would have been evaded only through a *reduction* of the logical content of (1).

Physicists often resort to ad-hoc stratagems of a different kind. Starting from an old hypothesis such as (1), they construct a hypothesis of the form $[H(\lambda,v,a,b)=0]$; where λ is assumed to be a parameter so small as to make the new theory empirically indiscernible from (1) in the domain Δ. They then adjust λ so as to make it coincide with some value λ_0 which is chosen in such a way that $[H(\lambda_0,v,a,b)=0]$ subsumes certain results known to have refuted (1); the hope being that $|\lambda_0|$ will be smaller than the $\eta(\mu_0)$ defined above. Ptolemaic Astronomy offers the example of a programme built on the ad-hoc method of introducing epicycles doctored to yield the facts which falsified earlier components of the programme. There was no overall strategy for determining the extra epicycles independently of known observational results (See Lakatos and Zahar 1978). In such cases, the general feeling among scientists is that the new hypotheses possess no greater truth-likeness than the old ones; which shows that despite being common to all programmes, the Correspondence Principle falls short of providing a comprehensive heuristic for scientific research. Additional constraints are needed. These often take the form of symmetry meta-principles under which physical laws are meant to be subsumed. This confirms, once more, Lakatos's views about the essential regulative role played by mathemetics in the development of research programmes.

Symmetries however raise paradoxical problems. No serious difficulties arise in the case of two incompatible theories, which generally have different symmetry groups. Redhead however pointed out that when an old law is logically entailed by a new one, we face a serious problem posed by the so-called Curie-Post Principle (Redhead 1975). The latter asserts that a general proposition transfers its symmetries to its logical consequences and to its special cases. This heredity principle seems prima facie self-evident: if two situations are not told apart by a strong theory, then they cannot *a fortiori* be distinguished by a weaker one; but this has the unfortunate consequence that as long as it progresses cumulatively, physics can display no new symmetries.

The Curie-Post principle clearly implies that no symmetry conditions can help us towards generating new hypotheses; for as long as we try to strengthen existing laws, no symmetry properties *could possibly be* new. Thus it seems as though neither the Correspondence nor any symmetry requirement can enable us to move away from existing theories. The only heuristic method left for strengthening a physical law appears to be that of modifying it, in some ad-hoc fashion, so as to force it to fit recalcitrant facts. Whence the need to demonstrate that the Curie-Post principle can be blocked; for only in this way could the *mathematical conditions* of Symmetry and Correspondence be regulative ideas which help us towards constructing new hypotheses consistent with the old ones.

Let $K(x,a,b,...)$ be a law, where: x is a sequence of spatio-temporal coordinates, while a,b, etc. denote specific physical magnitudes.

A symmetry of K consists of a sequence of functions $<\varphi(x),\alpha(x,a,b,...), \beta(x,a,b,...),...>$ such that, if we put: $x'=\varphi(x)$, $a'=\alpha(x,a,b,...)$, $b'=\beta(x,a,b,...)$, etc, we have:

(14) $\vdash(\forall x)(\forall a,b,...)[K(x,a,b,...)\Leftrightarrow K(x',a',b',...)]$.

For the sake of of simplicity, we shall henceforth write (x,a,b) for $(x,a,b,...)$. Thus (14) is shorthand for:

(15) $(\forall x,a,b)[K(x,a,b)\Leftrightarrow K(\varphi(x),\alpha(x,a,b),\beta(x,a,b))]$.

Any symmetry can be *passively* construed as a change from one frame Γ to another, Γ' say, such that: x and x' are the coordinates of the *same* point P, while a and a' denote the values assumed, at P, by the *same* physical entity when referred to Γ and to Γ' respectively. We could alternatively remain within the same frame Γ and *actively* interpret $x \rightarrow x'$ as a map of Γ into itself which sends a and b into a' and b' respectively.

Note that both construals block Kretschmann's objection that an arbitrary physical law can be rewritten in a generally covariant form; for α and β are defined as quantities dependent *only* on (x, a, b). This caveat forbids the introduction of new entities which might trivially make every law form-invariant under *all* well-behaved bijections. Specifying the arguments x, a and b of the functions α, β,... therefore lends heuristic efficacy to the symmetry principles (See Zahar 1989, chapter 8).

We are now in a position to prove that a logical consequence $E(x,a,b)$ of $K(x,a,b)$ need not possess the symmetry $<\varphi,\alpha,\beta>$ of K. Suppose that:

(16) $\vdash(\forall x,a,b)[K(x,a,b)\Rightarrow E(x,a,b)]$. Hence:
(17) $\vdash [K(x',a',b')\Rightarrow(x',a',b')]$. So, by (14):
(18) $\vdash(\forall x,a,b)[K(x,a,b)\Rightarrow E(x',a',b')]$

(16) and (18) do not however logically entail the relation $\vdash[E(x,a,b)\Leftrightarrow \Leftrightarrow E(x',a',b')]$, i.e the symmetry of E under $<\varphi,\alpha,\beta>$. As a trivial counterexample, consider the case where K is self-contradictory, hence implies *all hypotheses*. E.g. let $K(x,a,b)\equiv[(x=x)\wedge(a=a)\wedge(b\neq b)]$. Since $K(x',a',b')\equiv[(x'=x')\wedge \wedge(a'=a')\wedge(b'\neq b')]$, $K(x',a',b')$ is also inconsistent. Thus (14), i.e. (15), holds good for all φ, α and β; which means that K admits all functions $<\varphi,\alpha,\beta>$ as symmetries. It is however far from true that any $<\varphi,\alpha,\beta>$ defines a symmetry for every physical theory.

Assuming the truth of (14), it is just as important for us to prove that the 'special cases' of the hypothesis $K(x,a,b)$ need not be symmetric under $<\varphi,\alpha,\beta>$. To this end, it proves convenient to regard every symmetry as a map of the set Ω of all the solutions of K into itself. By definition, Ω is the class of all tuples $<A(x),B(x)>$ of functions of x such that:

(19) $\vdash(\forall x)K(x,A(x),B(x))$. By (15):
(20) $\vdash(\forall x)K(\varphi(x),\alpha(x,A(x),B(x)),\beta(x,A(x),B(x)))$.

Substituting $\varphi^{-1}(x)$ for x and then generalizing, we obtain:

(21) $\vdash(\forall x)K(x,A'(x),B'(x))$, where: $A'(x)=_{Def.}\alpha(\varphi^{-1}(x),A(\varphi^{-1}(x)),B(\varphi^{-1}(x)))$ and: $B'(x)=_{Def.}\beta(\varphi^{-1}(x),A(\varphi^{-1}(x)),B(\varphi^{-1}(x)))$.

Taken together, (19) and (21) tell us that if $<A(x),B(x)>$ is a solution of the law K, then due to the symmetry $<\varphi,\alpha,\beta>$ of K, $<A'(x),B'(x)>$ will constitute another solution of the same law. We can in fact interpret this symmetry as the application of Ω into itself defined by:

$<A(x),B(x)>\rightarrow<A'(x), B'(x)>$, where:

(22) $<A'(x),B'(x)>=_{Def.}<\alpha(\varphi^{-1}(x),A\varphi^{-1}(x),B\varphi^{-1}(x)),\beta(\varphi^{-1}(x),A\varphi^{-1}(x),B\varphi^{-1}(x))>$.

It is easily seen that this map does not necessarily send a 'special case', i.e. solution of a given type, into another one of the same type. For example: let Ω_0 be the subset of Ω consisting of all tuples of the form $<f(x),g(x)>$, where $g(x)=0$=zero function. Thus:

(23) $\vdash(\forall x)K(x,f(x),0)$.

The symmetry $<\varphi,\alpha,\beta>$ sends $<f(x),g(x)>$ into $<f'(x),g'(x)>$, where: $g'(x)= =\beta(\varphi^{-1}(x),f(\varphi^{-1}(x)),g(\varphi^{-1}(x)))=\beta(\varphi^{-1}(x),f(\varphi^{-1}(x)),0)$ since g is the zero function. There is however no reason to suppose that g' vanishes identically; so that we might well have $<f'(x),g'(x)>\notin\Omega_0$. Here is an important example given by M. Redhead (Redhead 1975, pp. 103-104): apply Maxwell's equations to a static system of charged particles; taking $f(x)$ and $g(x)$ to be the electric and the magnetic fields respectively, we conclude that $g(x)=0$. Let $\varphi(x)$ be any non-trivial Lorentz transformation. We know that Maxwell's theory is Lorentz-covariant and hence form-invariant under the symmetry φ. As already explained, φ can be *passively* viewed as a change from a stationary frame to a mobile one, in which the particles are no longer at rest. The moving frame will therefore contain both an electric and a non-vanishing magnetic field; i.e. $g'(x)\neq 0$.

The heredity principle is consequently false. This entitles us to search for hitherto unknown symmetries to be imposed on new hypotheses; where the latter *might*, for all we know, turn out to be consistent with our past conjectures. But how are we to find such symmetries? Going back to (15), it is obvious that if K, φ, α and β were all unknown functions, then it would be impossible for us even to begin our search. We could however start by considering a *given* theory K in order then to determine the class of its symmetries $<\varphi,\alpha,\beta>$; *or* from some *available* $<\varphi,\alpha,\beta>$, which could then be heuristically exploited in the construction of an appropriate K, i.e. of a theory covariant under $<\varphi,\alpha,\beta>$. And note that even within the same programme, these two approaches need not be mutually exclusive. For example: K could form part of a more general theory; it could e.g. be the set of Maxwell's equations taken within the context of classical physics. After examining the symmetries of K, we might find that they do not subsume the whole system in which K is imbedded. Thus, Classical Mechanics is not Lorentz-covariant since it does not possess all the Maxwellian symmetries.

The latter can be put to heuristic use by being turned into adequacy requirements to be imposed on *all* the laws of nature. Einstein and Poincaré started from a known **K**, namely from Maxwell's equations; they determined the symmetry group *L* of **K** and then used *L* in order to generate new hypotheses. This is also how Planck modified the old laws of motion into Lorentz-covariant relations. Another heuristic device consists in strengthening a covariance requirement by enlarging an existing symmetry group. For example: both for scientific and for philosophical reasons, Einstein decided to construct a generally covariant theory of gravitation. His field equations had therefore to keep their form, not only under the Lorentz transformations, but also under all well-behaved bijections (See Zahar 1989, chapters 5, 7, 8). Such a move seems incidentally to be forbidden by the Curie-Post principle, which is one reason why the latter had to be confined within strict limits.

6. *Conclusion*

Because of Lakatos's flamboyant style and his contempt for 'bourgeois formalism', there has been a tendency to under-estimate his creativity and his achievements. But he had the merit of reminding us of the necessity of accepting a synthetic –though irrefutable– principle of induction without which the reliability of scientific theories cannot be explained. More importantly: without renouncing the objectivist aspects of Popperian philosophy, he demonstrated that far from having to be relegated to the subjective domain of psychology, heuristics forms an essential, and perhaps the most interesting component of rational activity.

BIBLIOGRAPHY

Duhem, P.: 1914, *La théorie physique: son objet, sa structure*, Paris, Marcel Rivière.

Einstein, A. and Born, M.: 1969, *Briefwechsel 1916-1955*, München, Nymphenburger Verlagshandlung.

Kretschmann, E.: 1917, 'Über den physikalischen Sinn der Relativitätspostulate, A. Einsteins neue und seine ursprüngliche Relativitätstheorie', *Annalen der Physik* 53, 575-614.

Lakatos, I.: 1978, *Philosophical Papers 1*, Cambridge, Cambridge University Press.

Lakatos, I and Zahar, E.: 1978, 'Why did Copernicus's Programme supersede Ptoplemy's?', in Lakatos 1978, pp. 168-192.

Popper, K. R.: 1959, *The Logic of Scientific Discovery*, London, Hutchinson.

Popper, K. R.: 1965, *Conjectures and Refutations*, London, Routledge.

Popper, K. R.: 1979, *Die beiden Grundprobleme der Erkenntnistheorie*, Tübingen, Mohr (Siebeck).
Redhead, M.: 1975, 'Symmetry in Intertheory Relations', *Synthese* 32, 77-112.
Redhead, M.: 1995, *From Physics to Metaphysics*, Cambridge, Cambridge University Press.
Watkins, J. W. N.: 1984, *Science and Scepticism*, Princeton, Princeton University Press.
Zahar, E.: 1989, *Einstein's Revolution*, La Salle (Ill.), Open Court.

Elie Zahar is Reader Emeritus in Logic and Scientific Method at LSE and Research Scholar at the HPS Department in Cambridge. He has published two books: *Einstein's Revolution* (1989) and *Essai d'épistémologie réaliste* (2000). His main interests are in the philosophy of science, in the foundations of mathematics and in their role in scientific discovery. Among his publications are: 'The Mathematical Origins of General Relativity and of Unified Field Theories' (1979), 'Logic of Discovery or Psychology of Invention?' (1983), 'Les fondements des mathématiques d'après Poincaré et Russell' (1987), 'Natural Axiomatization' (1991), 'Intentionality or Platonism?' (1993), and 'Poincaré's Philosophy of Geometry' (1997).