

PRESERVATION OF EMPIRICAL SUCCESS AND INTERTHEORETICAL CORRESPONDENCE: JUSTIFYING REALISM WITHOUT THE NO MIRACLES ARGUMENT

GERHARD SCHURZ
University of Düsseldorf

Abstract

This paper utilizes a logical correspondence theorem (which has been proved elsewhere) for a justification of a weak version of scientific realism which does not presuppose Putnam's no-miracles-argument (NMA). After presenting arguments against the reliability of the unrestricted NMA in sec.1, the correspondence theorem is explained in sec.2. In sec.3, historical illustrations of the correspondence theorem are given, and its ontological consequences are worked out in terms of the indirect reference and partial truth. In the final sec.4 it is shown how the correspondence theorem together with the assumption of 'minimal realism' yields a justification of the abductive inference from the strong empirical success of a theory to the partial truth of its theoretical part.

1 THE NO-MIRACLES ARGUMENT (NMA) AND ITS LIMITATIONS

It is a crucial property of scientific theories that they contain concepts, content, and/or models, which represent structures of objects which go beyond what is empirically observable, or 'pre-theoretically given' – so-called *theoretical* concepts, content, or models. This holds quite independently of how one draws the borderline between the observable and the non-observable (by the naked eye, by scientific measurement instruments, or by shared pre-theories). The 'theoretical superstructure' of scientific theories plays a crucial role in explaining and unifying the observable phenomena in the domain of the theory. The class of empirical consequences which follows from the theory (deductively or probabilistically) is called the *empirical* (or *pre-theoretical*) *content* of a theory. The empirical (or pre-

theoretical) *success* of a theory is that part of its empirical content which is verified or confirmed by observations.

Scientific realism is the view that the empirical success of a theory is a reliable indicator of the (approximate) truth of the theory, including the truth of its theoretical superstructure. Thereby, the concept of truth is understood in the ordinary correspondence-theoretical sense, formally explicated in terms of semantical model-theory which goes back to Tarski. In contrast, *scientific instrumentalism* holds that the theoretical superstructure of a theory has merely the instrumental purpose of entailing the observed phenomena in a most simple and unifying way, but there is no reason to assume that this theoretical superstructure corresponds to an unobservable external reality. For example, according to the realistic interpretation of theoretical physics, electrons or quarks do exist, and what physical theories tell about them is approximately true, while according to the instrumentalistic interpretation, electrons and quarks don't exist, at least not literally, and what theories tell about them is not true, although it is a extremely economic way of unifying a variety of empirical phenomena.

The standard justification of scientific realism in the contemporary realism vs. instrumentalism debate is the *no-miracles* argument (NMA), which goes back to Putnam (1975, p.73), and has been used in various ways as a defense of scientific realism (cf. Boyd 1984). This argument says that the empirical success of contemporary scientific theories would be a sheer miracle if we would not assume that their theoretical superstructure, or ontology, is approximately true in the sense of scientific realism. More precisely, the best if not the only reasonable explanation of the continuous empirical success of scientific theories is the realistic assumption that their theoretical 'superstructure' (their non-observable part) is approximately true and, hence, their central theoretical terms refer to real though unobservable constituents of the world. However, there NMA is beset by two (at least) equally strong counterargument, one empirical counterargument and one theoretical counterargument.

The empirical counterargument is the *pessimistic meta-induction* argument (PMA), which goes back to Laudan (1981). The PMA points to the fact that in the history of scientific theories one can recognize radical changes at the level of theoretical superstructures (ontologies), although there was continuous progress at the level of empirical success. On simple inductive grounds, one should expect therefore that the theoretical superstructures (ontologies) of our presently accepted theories will also be over-

thrown in the future, and hence can in no way be expected to be approximately true.

The theoretical counterargument is the *no-speculation* argument (NSA), which has an old tradition in the philosophy of science. In Schurz (2008, §7) have tried to elaborate this argument in a defensible way. The NSA points out that for every possible observation one may construct *ex-post* and *ad-hoc* some speculative 'theory' which just entails ('explains') this observation, but has no other (logically independent) observable consequence. The empirical success of such speculative 'theories' is in no way a reliable indicator of their approximate truth. In the simplest case, a speculative explanation has the following structure:

Speculative 'explanation'-schema
(for particular and general phenomena):

Explanandum: Something happened, or happens regularly.

Speculative 'explanans': Some kind of power wanted that this something to happen, or to happen regularly.

Speculative explanations of this sort have been applied by our human ancestors since the earliest times. All sorts of unexpected events can be explained by assuming one or several God-like power(s). Speculative pseudo-explanations do *not* offer a *proper* unification, because for every particular phenomenon *E* a special hypothetical 'wish' of the God-like power to create *E* has to be postulated (cf. Schurz/Lambert 1994, p.86). On the same reason, speculative pseudo-explanations are *post-hoc* and have *no predictive power* at all, because God's unforeseeable decisions can be known only *after* the event has already happened. Moreover, since the explanandum imposes no constraints on the hypothetical power except that it has created the explanandum in some unexplained (typically super-natural) ways, the same phenomena can be explained by a plentitude of different speculative explanations, which is historically illustrated by the fact that human mankind has invented a multitude of different religious stories about the genesis of the world (Wilson 1998, ch.11, estimates their number as 100.000).

In Schurz (2008, §7) the following criterion for demarcating speculative from scientifically worthwhile explanations (*viz.* theories) is proposed: while speculative explanations postulate for each new phenomenon a new

kind of theoretical cause, scientifically worthwhile explanations introduce new theoretical entities only if they figure as common and unifying cause or explanations of *several* intercorrelated phenomena. It is of utmost importance that this unificatory potential of good scientific theories goes hand in hand with their potential to entail what we will call *strong* empirical success in sec.2: potentially novel predictions by which they can be tested *independently* from those phenomena to which they have been 'fitted' in an ex-post fashion.

2 THE CORRESPONDENCE THEOREM AND ITS CONDITIONS

The PMA and the NSA are strong counterarguments against the reliability of the NMA. Does there exist a justification of scientific realism which does not presuppose the dubious NMA? Even if such a justification does not exist, or exists only for a rather weak version of scientific realism, it would be advantageous to know, by an independent argument, under which conditions the NMA is reliable, and why it is unreliable if it is applied to theoretical speculations. In the remaining part I will propose such an argument. It is a logical argument which allows us to infer, under certain conditions, the partial truth of a theory from its empirical success, *relative* to a theory T^* which preserves T 's empirical success and is assumed as true, or at least closer to the truth than T .

My argument is based on relations of *correspondence* between historically consecutive theories with increasing (or at least not decreasing) empirical success. Boyd (1984) and other philosophers of science have emphasized the existence of such relations of correspondence, which reflect that even on the theoretical level something is preserved through historical theory change and, thus, has a justified realist interpretation. Laudan (1981, pp.121, 126), however, has objected that there is no evidence for systematic relations of correspondence at the theoretical level – positive examples come mainly from the history of mechanics and have exceptional status. Laudan has given a much debated list of counterexamples – examples of scientific theories which were strongly successful at their time but whose ontology was incompatible with that of contemporary theories, from which Laudan concludes that these theories cannot possibly correspond to contemporary theories. In Schurz (2009) it is argued that Laudan is wrong: there are indeed systematic reasons for relations of theory-correspondence, which are based on the fact that under natural conditions the cumulatively

increasing empirical *success* of theories entails constraints on their theoretical superstructure from which one can obtain such relations of correspondence. Schurz (2009, §4) proves a correspondence theorem which presupposes the following conditions to be satisfied for the predecessor theory T and the successor theory T^* (both viewed as sets of sentences of an interpreted language):

(*Condition 1 on T and T^**): The theories T and T^* share a common non-theoretical vocabulary in which their joint empirical (or non-theoretical) success is expressed, and they share a partitioned domain of application $A = A_1 \cup \dots \cup A_n$, with $n \geq 2$, whose (disjoint) subdomains A_i are described by antecedent conditions A_i expressed by the shared non-theoretical vocabulary.

(*Condition 2 on T*): (2.1): The predecessor theory T has *strong potential empirical success* w.r.t. partitioned domain $A = A_1 \cup \dots \cup A_n$, which means by definition that T entails a set of conditionals $S(T) = \{S_{ij}: 1 \leq i \neq j \leq n\}$ of the form

$$(S_{ij}) (\forall x): (\exists u)(A_i(x, u) \wedge \pm R_i[x, u]) \rightarrow (\forall u)(A_j(x, u) \rightarrow \pm R_j[x, u]).$$

In words: If R_i has happened in circumstances A_i , then R_j will happen in circumstances A_j .

Notation: x and u are (sequences of) individual variables, x refer(s) to the system under consideration, u are non-theoretical auxiliary variables, e.g. the time variable (but possibly empty), the antecedent conditions A_i describe the conditions of the subdomains, the R_i describe typical reactions of the system x in the subdomain A_i expressed in the shared non-theoretical vocabulary and “ \pm ” means “unnegated or negated” (i.e., $\pm \in \{\text{emptystring}, \neg\}$). Round brackets “ $A(v)$ ” indicate that formula A contains all of the variable(s) in v , while square brackets $R[v]$ mean that R may contain only some variables in v – this is needed for sufficiently flexible theory-reconstructions (see below). The conditionals (S_i) allow one to infer from what has been observed in one domain of application (namely $A_i \wedge R_i$) what will happen in a different domain of application (namely, if A_j , then R_j), without the system x having ever been put into conditions A_j before. Therefore, the conditionals (S_i) enable (potentially) *novel* predictions and, thus, serve as an example of *strong* empirical success. For example, when T is generalized oxidation theory, A_1 may describe the exposition of a metal to air and water and R_1 the end products of the reaction of oxidation, A_2 the exposi-

tion of a metal to hydrochloric acid and R_2 the end products of the reaction of salt-formation, etc.

(2.2): The strong potential empirical success $S(T)$ of T must have been *yielded* by a theoretical expression φ of T , which means by definition that T entails the set of bilateral reduction sentences $B(T, \varphi) = \{B_i: 1 \leq I \leq n\}$ of the form

$$(B_i): (\forall x, u): A_i(x, u) \rightarrow (\varphi(x) \leftrightarrow R_i[x, u]).$$

In words: under empirical circumstances A_i , the presence of φ is indicated or measured by an empirical phenomenon or process R_i .

It is easily seen that $B(T, \varphi)$ entails $S(T)$. I understand bilateral reduction sentences, differently from Carnap, in a non-reductionist sense, as ordinary *measurement* conditions for theoretical expressions: (i) they are not analytically but synthetically true, (ii) they are usually *not* part of T 's axiomatization, but are obtained as consequences of a suitably rich version of the theory, and (iii) their logical form covers all kinds of quantitative measurement laws (via the equivalence of " $\varphi(x) = r_i[x, u]$ " with " $\forall z \in \text{Reals}: \varphi(x) = z \leftrightarrow r_i[x, u] = z$ ") (for details cf. Schurz 2009, §3). The ontological interpretation of the role of φ as described by $B(T, \varphi)$ is the following: φ figures as a measurable *common cause* or *common explanation* of the observable regularities or dispositions $D_i :=$ "if A_i , then R_i ", in the sense that for all $1 \leq I \leq n$, $(\forall x, u): (\varphi(x) \rightarrow (A_i(x, u) \rightarrow R_i[x, u]))$ (cf. Schurz 2008, §7.2).

(*Condition 3 on T^**): The strong potential empirical success of T , $S(T)$, must be entailed by T^* in a T^* -*dependent* way, which means by definition that for every conditional of the above form $(S_{i,j})$ which follows from T there exists a theoretical *mediator* description $\varphi^*_{i,j}(x)$ of the underlying system x such that $(\forall x)((\exists u)(A_i(x, u) \wedge \pm R_i[x, u]) \rightarrow \varphi^*_{i,j}(x))$ as well as $(\forall x)(\varphi^*_{i,j}(x) \rightarrow (\forall u)(A_j(x, u) \rightarrow \pm R_j[x, u]))$ follow from T^* . The need of condition 3 on T^* for the proof of the theorem is obvious, because in order to infer from the entailment of $S(T)$ by T^* something about a correspondence between T 's and T^* 's theoretical part, we must assume that this entailment utilizes T^* 's theoretical part. The justification of condition 3 follows from the fact that $S(T)$ is a strong (potential) empirical success in the sense of novel predictions. From an empirical description of what goes on in domain A_i nothing can be concluded by means of empirical induction alone about what goes on in a qualitatively different domain A_j . For example, from observing the chemical reaction R_i of a given kind of substance

(e.g. a metal) under the influence of oxygen (A_i), nothing can be concluded by empirical induction about the chemical reaction (R_j) of this substance under the influence of hydrochloric acid (A_j). For such an inference one needs a theoretical mediator description φ^* (e.g. the chemical structure of metals) which interpolates between ($A_i \wedge R_i$) and ($A_j \rightarrow R_j$).

(*Condition 4 on T and T^**): The two theories T and T^* must be *causally normal* w.r.t. the partitioned domain $A = A_1 \cup \dots \cup A_n$, which means by definition that: (a) the shared non-theoretical vocabulary of T and T^* divides into a set of independent and a set of dependent parameters (predicates or function terms), (b) the descriptions ' $A_i(x)$ ' of the subdomains A_i are formulated solely by means of the independent parameters (plus logico-mathematical symbols), and (c) no non-trivial claim about the state of the independent parameters of a system x can be derived in T (or T^*) from a purely T (T^*)-theoretical and T (T^*)-consistent description of x . Again, this is a natural condition – for example, nothing can be concluded from the theoretical nature of a certain substance about what humans do with it, whether they expose it to hydrochloric acid or to heat or whatever. (*Remark*: for reasons of simplicity we use ' A ' and ' A_i ' both for an open formula and the set designated by the formula; the context makes it clear what is meant.)

Framed in the explained terminology, the correspondence theorem now asserts the following:

Correspondence theorem: Let T be a consistent theory which is causally normal w.r.t. a partitioned domain $A = A_1 \cup \dots \cup A_n$ and contains a T -theoretical expression $\varphi(x)$ which *yields* a strong potential empirical success of T w.r.t. partitioned domain A .

Let T^* be a consistent successor theory of T (with an arbitrarily different theoretical superstructure) which is likewise causally normal w.r.t. partitioned domain A and which entails T 's strong potential empirical success w.r.t. A in a T^* -dependent way.

Then T^* contains a theoretical expression $\varphi^*(x)$ such that T and T^* together imply a *correspondence relation* of the form

$$(C): (\forall x)(\forall u): A(x, u) \rightarrow (\varphi(x) \leftrightarrow \varphi^*(x))$$

In words: whenever a system x is exposed to the circumstances in one of the subdomains of A , then x satisfies the T -theoretical description φ iff x satisfies the T^* -theoretical description φ^* .

Remark: This implies that $\varphi(x)$ refers indirectly to the theoretical state of affairs described by $\varphi^*(x)$ – provided $\varphi^*(x)$ refers at all, which presupposes that T^* is at least partially true.

Corollary 1: $B(T, \varphi) \cup T^*$ is consistent, and (C) follows already from $B(T, \varphi) \cup T^*$.

Corollary 2: φ^* is unique in domain A modulo T^* -equivalence.

The proof of the correspondence theorem (details in Schurz 2009, §4) proceeds by showing that from T^* 's preservation of T 's strong potential success plus condition 3 on T^* it follows that also T^* contains some expression φ^* whose designatum figures as a measurable common cause of the correlated regularities or dispositions “if A_i , then R_i ”. The requirement of T^* -theoretical mediators of condition 3 enables only the derivation of two unilateral reduction sentences from T^* (for each A_i), one for the positive test condition ($T^* \Vdash (A_i \rightarrow (\pi^* \leftrightarrow R_i))$) and one for the negative test condition ($T^* \Vdash (A_i \rightarrow (\mu^* \leftrightarrow \neg R_i))$). The causal normality condition 4 is then needed to prove that T^* entails $(\pi^* \leftrightarrow \neg \mu^*)$; whence π^* and $\neg \mu^*$ can be collapsed into the required T^* -corresponding concept φ^* . Thus, T^* entails the same set of bilateral reduction sentences as T entails for φ , or formally $T^* \Vdash B(\varphi, T)[\varphi^*/\varphi]$, where “[φ^*/φ]” denotes the operation of replacement of φ by φ^* .

While the conditions 1, 3 and 4 are rather mild, condition 2 on the predecessor theory T is a crucial constraint which excludes pre-scientific theoretical speculations. According to condition 2, the correspondence theorem applies to all and *only* those theoretical expressions φ of T which yield strong potential empirical success by way of bilateral reduction statements. (I speak of strong *potential* success $S(T)$ because the logical derivation of the correspondence theorem is independent from the factual truth values of the considered theories.) Condition 2 on T requires that (strictly) correlated empirical regularities or dispositions “if A_i , then R_i ” are explained within T by an unobservable but measurable common cause or explanation φ . It was argued in sec.1 that it is exactly this *common cause* property which distinguishes *scientifically legitimate* theoretical explanations from speculative abductions. Indeed, the proof of the correspondence theorem would be impossible, if φ were characterized by only *one* disposition, i.e. one bilateral reduction sentence $A_1 \rightarrow (\varphi \leftrightarrow R_1)$. In Schurz (2009, §6) this is demonstrated by the example of Aristotelean physics which introduces a distinct cause for each kind of motion (cf. also van Fraassen

2006, p.281), whence no correspondence between Aristotelean and Newtonian physics can be established.

3 ONTOLOGICAL INTERPRETATION AND HISTORICAL ILLUSTRATION OF THE CORRESPONDENCE THEOREM

It might seem to some readers that the result of the correspondence theorem is too good to be true. So we take pains to explain how the theorem works and to point to its weak spots. Implicit in corollary 1 is the possibility that the two theories T and T^* are mutually incompatible, at the theoretical level, or at the empirical level outside the domain of the shared empirical success. If T and T^* are incompatible, then it would, of course, be a trivial assertion that the union of T and T^* entails a correspondence relation (C), because in that case this union entails everything. Therefore, corollary 1 tells us that the correspondence principle follows in a non-trivial way from a certain part of T , namely $B(T, \varphi)$, which is *consistent* with T^* . Only this part of T , and not the whole of T , is preserved by the correspondence to T^* . In addition, our theorem takes care of empirical incompatibilities by restricting the correspondence between φ of T and φ^* of T^* to a given partitioned domain A , in which T was strongly successful. Outside of the domain A , T may have wrong empirical consequences which are not shared but corrected by T^* .

One example of a correspondence relation between theoretically incompatible theories is the *phlogiston theory* and the *generalized oxygen theory* of combustion. According to the phlogiston theory (developed by Becher and Stahl in the late 17th and early 18th century), every material which is capable of being burned or calcinated contains *phlogiston* – a substance different from ordinary matter which was thought to be the bearer of combustibility. When combustion or calcination takes place, the burned or calcinated substance delivers its phlogiston, usually in the form of a hot flame or an evaporating inflammable gas, and a dephlogisticated substance-specific residual remains. In the 1780s, Lavoisier introduced his alternative oxygen theory according to which combustion and calcination consists in the oxidation of the substance being burned or calcinated, that is, in the formation of a chemical bond of its molecules with oxygen. The assumption of the existence of a special bearer of combustibility became superfluous in Lavoisier's theory. In modern chemistry, Lavoisier's theory

is accepted in a generalized and corrected form, in which the oxidizing substance need not be oxygen, but may be any electronegative substance.

Phlogiston theory was theoretically incompatible with oxygen theory because it assumed the existence of phlogiston which did not exist according to oxygen theory. Nevertheless phlogiston theory was strongly successful in explaining four domains of chemical reactions, namely (1.) the combustion of organic material, (2.) the calcination (roasting) of metals, (3.) salt-formation through the solution of metals in acids, and (4.) the inversion of calcination and salt-formation. In Schurz (2009) it is shown that the theoretical concept which yielded phlogiston theory's strong success in these domains (in the sense of condition 2.2 in sec. 2) was not "phlogiston" itself – this term was empirically underdetermined – but, rather, the concepts of *phlogiston-richness* and *dephlogistication* = *release of phlogiston*. For these concepts, the following correspondence relation with generalized oxidation theory can be derived along the logical route of the correspondence theorem:

Phlogiston-richness of a substance corresponds (and indirectly refers) to the *electropositivity* of the substance.

Dephlogistication of a substance corresponds (and indirectly refers) to the *donation of electrons* of substance's atoms to their electronegative bonding partner.

Let me turn to the *ontological interpretation* of the correspondence theorem. First of all, it is clear from the foregoing that the above correspondence relations do *not* preserve *all* of the meaning of 'phlogistication' or 'phlogiston-richness'. They cannot be regarded as *analytic* truths, but have to be regarded as *synthetic* statements which are true in the domain of applications in which phlogiston theory was empirically successful. Second, the correspondence relation (C) is *not* meant to say that whenever T 's intended model (de-phlogistication) is realized, also T^* 's intended model (generalized oxidation) is realized – this would be a strange scenario of 'causal overdetermination'. Rather, (C) expresses the possibility of a φ - φ^* -*reference-shift*: instead of the reference assigned to φ in T 's intended model M (e.g. phlogiston-richness, or dephlogistication), we can assign to φ the reference of φ^* in T^* 's intended model M^* (electropositivity, or donation of electrons, respectively). Such a φ - φ^* -reference shift will preserve the truth of $B(T, \varphi)$ (proof in Schurz 2009, §4).

It is important, thereby, that the expression φ which yielded T 's strong success need not be a primitive term but may be a *composite* expression (and the same holds for the corresponding expression φ^* of T^*). Whenever T 's expression φ corresponds to φ^* of T^* , but the ontology of the old theory T concerning the entities involved in φ is *incompatible* with the contemporary theory T^* , then it will be the case that φ is not a primitive but a *complex* expression of T , and T will contain certain theoretical assumptions about φ 's *inner structure* or *composition* which from the viewpoint of T^* are false – for example, that ‘dephlogistication’ is a process in which a special substance different from ordinary matter, called ‘phlogiston’, leaves the combusted substance. While T has got a *right* model about φ 's *outer structure*, i.e. the causal relations between the complex entity φ and the empirical phenomena, it has got a wrong model about φ 's inner structure.

This situation is *typical* even for most advanced contemporary theories. For example, we are confident that protons exist because they are measurable common causes of a huge variety of empirical regularities. But concerning the hypothesis about the *inner composition* of protons consisting of *three quarks* things are different: physicists cannot measure quarks in isolation and, hence, are much more uncertain about their reality. In other words, the conception of realistic reference which is supported by the correspondence theorem is compatible with a certain amount of empirical underdetermination even in our most advanced theories. In Schurz (2009) it is argued that the notions of the outer and inner structure of a complex expression or entity φ reflect Worrall's (1989) distinction between ‘structure’ and ‘content’ in an ontologically unproblematic way: the ‘structure’ which is preserved is φ 's *outer* structure, while the ‘content’ which is not preserved is φ 's *inner* structure. Often, the preserved outer structure of a T -expression φ_1 does not only contain φ_1 's relations to observable phenomena, but covers also φ_1 's relation to other T -theoretical terms φ_2 for which a T^* -correspondence can *also* be established. In this sense, the relation between dephlogistication and phlogistication as inverse chemical reactions is preserved in modern chemistry.

The fact that the shifted interpretation of $B(T, \varphi)$ within T^* 's ontology forgets T 's hypotheses about φ 's inner structure and preserves only φ 's outer structure implies that properly speaking, the preserved content-part $B(T, \varphi)$ has to be understood as a Ramsey-type existential quantification in which one quantifies over φ as a whole: if $B(T, \varphi)$ has the form $S(\varphi)$ where φ has the form $f(\mu)$ (e.g. release-of-phlogiston), then what is preserved

within T^* 's ontology is not $\exists X S(f(X))$ but the logically weaker $\exists X S(X)$. It follows from these considerations that the content-part of T which is preserved in T^* is a 'structural' (or quantificational) content-part of T in the explained sense, but not a simple *conjunctive* part of T 's axioms. This contrasts with Psillos' 'theoretical constituents' (1999, pp.80f.), which are conjunctive parts of the given theory. The preservation of theoretical constituents in the sense of Psillos has been criticized by Lyons (2006), while the preservation of content-parts in my sense is not beset by Lyons' objections.

Two further applications of the correspondence theorem are illustrated in Schurz (2009, §5): one example is the caloric theory of heat, with the correspondence between the amount of caloric particles in a substance X and the mean kinetic energy of X 's molecules, and the other example is Fresnel's mechanical wave theory of light, with the correspondence between the oscillation velocity of ether molecules and the oscillation strength of the electromagnetic field in Maxwell's account. In all these historical examples, there exists a *unique* correspondence concept φ^* in T^* . But not always is the situation so nice. T^* need not explain the correlated regularities "if A_i , then R_i " by postulating exactly one common cause (or explanation). It may also explain them by assuming a more complicated *net* of causes or hidden variables. This is the point where corollary 2 comes into play: if T^* may contain several causes $\varphi_1^*, \dots, \varphi_k^*$ (with $k \leq n$) which correspond to φ in domain A , then corollary 2 tells us that in domain A all these causes are equivalent, i.e. (for all $1 \leq I \neq j \leq k$) $T^* \Vdash (\forall x, u): A \rightarrow (\varphi_i^*(x) \leftrightarrow \varphi_j^*(x))$. If we want to have a *unique* formal counterpart of φ in T^* in such cases, we should take the disjunction $\vee_{1 \leq i \leq k} \varphi_i^*(x)$. But this formal trick does not remove the possible ontological ambiguity. For the mutual equivalence of the φ_i^* 's which holds in domain A may fail to hold outside of domain A . It may happen that T^* contains several counterparts of φ , say φ_1^* and φ_2^* , which are not mutually identified by T^* because although they degenerate (extensionally) into one cause in domain A , they split up into two distinct causes outside of the domain A . An example of this sort which is analyzed by an early paper of Field (1973) is the correspondence between Newtonian mechanics and special relativity theory. Field shows that special relativity theory provides two counterpart notions to the Newtonian concept of mass m , (i) rest mass m_0 and (ii) relativistic mass

$$m_r := m_0/\gamma \text{ (with } \gamma := \sqrt{1 - \frac{v^2}{c^2}}).$$

Since rest mass and relativistic mass are approximately identical in the domain in which Newtonian physics was empirically successful, this situation fits nicely with our correspondence theorem.

4 A JUSTIFICATION OF SCIENTIFIC REALISM WHICH DOES NOT PRESUPPOSE THE NMA

Under an additional assumption which I call ‘minimal realism’, the correspondence theorem justifies a weak version of scientific realism without presupposing the NMA or some other form of IBE (inference to the best explanation). Of course, the correspondence theorem alone justifies only a *conditional realism*: if one assumes the (approximate) realistic truth of the presently accepted theory T^* , then also outdated theories T (satisfying the conditions) will contain a (theoretico-structural) content-part which is indirectly and hence partially true. This conditional realism *weakens* Laudan's pessimistic meta-induction. But conditional realism *alone* is not sufficient to justify scientific realism. For someone who, on independent epistemological grounds, does not believe that contemporary or future scientific theories are approximately true, this conditional realism cannot tell anything about the partial truth of earlier theories.

But the situation changes if one makes the following assumption of *minimal realism* (MR):

(MR) The observed phenomena are *caused* by an external reality whose structure can *possibly* be represented in an approximate way by an ideal theory T^+ which is causally normal, entails the observed phenomena in a T^+ -dependent way, and whose language is in reach of humans' logico-mathematical resources.

(MR) is a *minimal* realistic assumption because it merely says that an approximately true theory describing the external reality in a humanly accessible language is possible – independent of whether humans will ever find this theory. The requirement that the T^+ -language must be accessible by humans is necessary because otherwise we could not apply our correspondence theorem to this ideal theory T^+ . The requirement of language-accessibility does not entail that the ideal theory T^+ itself must be graspable by humans. In the contrary, T^+ may be so complex that it can impossibly be understood by human brains, even if aided by super-computers.

Together with (MR), the correspondence theorem *entails* that the abductive inference from the strong empirical success of theories to their partial realist truth is *justified*. For if (MR) is true, then there exists an approximately true *ideal theory* T^+ , which need not be known to us and preserves all of the (true) strong empirical success which our accepted theories have. So the correspondence theorem implies that every (theoretico-structural) content-part of our contemporary theories which satisfies condition 2 (plus 1 and 4) corresponds to a content part of the ideal theory T^+ , and hence is indirectly true. In this way, my account provides an independent and non-circular justification of a weak version of the NMA, or of the abductive inference to the partial truth of strongly successful theories.

REFERENCES

- Boyd, R. (1984): “The Current Status of Scientific Realism”, in: J. Leplin (ed.), 1984, *Scientific Realism*, Berkeley: University of California Press, pp.41–82.
- Field, H. (1973): “Theory Change and the Indeterminacy of Reference”, *Journal of Philosophy* 70, 1973, pp.462–481.
- Laudan, L. (1981): “A Confutation of Convergent Realism”, reprinted in D. Papineau (ed.), 1996, *The Philosophy of Science*, Oxford: Oxford Univ. Press, pp.107–138.
- Lyons, T.D. (2006): “Scientific Realism and the Stratagemata de Divide et Impera”, *British Journal for the Philosophy of Science*, 57, pp.537–560.
- Psillos, S. (1999): *Scientific Realism. How Science Tracks Truth*, London and New York: Routledge.
- Putnam, H. (1975): “What is Mathematical Truth?”, in H. Putnam, *Mathematics, Matter and Method*, Cambridge: Cambridge Univ. Press, pp.60–78.
- Schurz, G., and Lambert, K. (1994): “Outline of a Theory of Scientific Understanding”, *Synthese* 101/1, pp.65–120.
- Schurz, G. (2008): “Patterns of Abduction”, *Synthese* 164, 2008, pp.201–234 (online <http://dx.doi.org/10.1007/s11229-007-9223-4>).
- Schurz, G. (2009): “When Empirical Success Implies Theoretical Reference: A Structural Correspondence Theorem”, *British Journal for the Philosophy of Science* 60/1, 2009, pp.101–133 (online: <http://dx.doi.org/10.1093/bjps/axn049>).
- Van Fraassen, B. (2006): “Structure: Its Shadow and Substance”, *The British Journal for the Philosophy of Science*, 57, pp.275–307.
- Wilson, E.O. (1998): *Consilience. The Unity of Knowledge*, Alfred A. Knopf, New York.
- Worrall, J. (1989): “Structural Realism: The Best of Both Worlds”, *Dialectica*, 43/1–2, pp.99–124.