

Running Head: Structural Correspondence: Phlogiston Theory and Newtonian Mechanics

Title: Structural Correspondence, Indirect Reference, and Partial Truth: Phlogiston Theory and Newtonian Mechanics

Author: Gerhard Schurz

Address: Prof. Dr. Gerhard Schurz

Department of Philosophy, University of Duesseldorf

Geb. 23.21, Universitaetsstrasse 1, 40225 Duesseldorf, Germany

E-Mail address: schurz@phil-fak.uni-duesseldorf.de

Abstract: This paper elaborates on the following correspondence theorem (which has been defended and formally proved elsewhere): if theory T has been empirically successful in a domain of applications A, but was superseded later on by a different theory T* which was likewise successful in A, then under natural conditions T contains theoretical expressions φ which were responsible for T's success and *correspond* (in A) to certain theoretical expressions φ^* of T*. I illustrate this theorem at hand of the phlogiston vs. oxygen theories of combustion, and the classical vs. relativistic theories of mass. The ontological consequences of the theorem are worked out in terms of the indirect reference and partial truth. The final section explains how the correspondence theorem may justify a weak version of scientific realism without presupposing the no-miracles argument.

Structural Correspondence, Indirect Reference, and Partial Truth: Phlogiston Theory and Newtonian Mechanics

Gerhard Schurz (University of Duesseldorf)

Abstract: This paper elaborates on the following correspondence theorem (which has been defended and formally proved elsewhere): if theory T has been empirically successful in a domain of applications A, but was superseded later on by a different theory T* which was likewise successful in A, then under natural conditions T contains theoretical expressions φ which were responsible for T's success and *correspond* (in A) to certain theoretical expressions φ^* of T*. I illustrate this theorem at hand of the phlogiston vs. oxygen theories of combustion, and the classical vs. relativistic theories of mass. The ontological consequences of the theorem are worked out in terms of the indirect reference and partial truth. The final section explains how the correspondence theorem may justify a weak version of scientific realism without presupposing the no-miracles argument.

1. NMA, PMA and the Correspondence Theorem

The contemporary debate on scientific instrumentalism versus realism is dominated by the influence of two strong arguments which pull into opposite directions. On the one side we have the *no-miracles* argument (NMA), suggested e.g. by Putnam (1975, 73), which 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. On the other side, there is the *pessimistic meta-induction* argument (PMA), put forth by Laudan (1981), which 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 overthrown in the future, and hence can in no way be expected to be approximately true.

I start this paper with a concession to the anti-realist camp: it is doubtful whether the NMA in its unrestricted form is a reliable argument (additional reasons for this doubt are given below). In this paper I will present a logical argument which is independent from the NMA and allows us to infer, under certain conditions, from the empirical success of a theory its partial truth, *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, 121,

126), however, has objected that there is no evidence for systematic relations of correspondence at the theoretical level. He 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 defi-

dition that T entails a set of conditionals $S(T) = \{S_{i,j}: 1 \leq i \neq j \leq n\}$ of the form

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

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 either "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 exposition 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* 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). Anti-realists should note, however, that such a causal interpretation (which I find quite attractive) is *not necessarily required* by the correspondence theorem.

(*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_{ij}) which follows from T there exists a theoretical *mediator* description $\varphi^*_{ij}(x)$ of the underlying system x such that $(\forall x)(\exists u)(A_i(x, u) \wedge \pm R_j[x, u] \rightarrow \varphi^*_{ij}(x))$ as well as $(\forall x)(\varphi^*_{ij}(x) \rightarrow (\forall u)(A_j(x, u) \rightarrow \pm R_j[x, u]))$ follow from T^* .¹ 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 interpo-

lates 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.

Framed in the explained terminology, the correspondence theorem now asserts the following (proof in Schurz 2008, § 4):

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

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

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.

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. Condition 2 entails that 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. If the potential success of T is preserved by T^* (cond. 1) in a T^* -dependent way (cond. 3), and both theories

are causally normal (cond. 4), the correspondence relation (C) follows. Note that I speak of (strong) *potential* success because the logical derivation of the correspondence theorem is independent from the factual truth values of the considered theories. The logical theorem presupposes that T entails the potential predictions $S(T)$ which are T^* -dependently preserved by T^* , but it does not presuppose that they are factually true (should they be false, then the correspondence relation (C) relates theoretical concepts of two false theories).

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. Phlogiston theory (the example of our next section) was theoretically incompatible with oxygen theory because it assumed the existence of phlogiston which did not exist according to oxygen theory. 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*.

Let us turn to the *ontological interpretation* of the correspondence theorem. The correspondence relation (C) is *not* meant to say that whenever T's intended model is realized, also T*'s intended model 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, we can assign to φ the reference of φ^* in T*'s intended model M*. 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, of course, for the corresponding expression φ^* of T*). This leaves room for either an ontological underdetermination or a non-reference of T's primitive terms. Detailed illustrations of this situation are given at hand of the phlogiston-oxygen example in the next section. Also, note that I use here the model-theoretic notion of 'reference in an intended model of a theory T'. The 'intended' model need not be actual ('real') but may be merely conceptual ('imaginary'). If T's model is merely conceptual but T*'s model is actual, then the reference-shift provides φ with 'real' instead of merely 'conceptual' reference.²

As mentioned above, in its causal interpretation 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 φ . In Schurz (2008, §7.1) it is argued at length that it is exactly this *common cause* property which distinguishes *scientifically legitimate* abductions to theoretical entities from *speculative* abductions: while typical speculations postulate for each new phenomenon a new kind of theoretical cause, good science introduces new theoretical entities only if they figure as common causes of *several* intercorrelated phenomena. 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, whence no correspondence between Aristotelean and Newtonian physics can be established.

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 dispositional properties "if A_i , then R_i ". In other words, T^* entails the same set of bilateral reduction sentences $B(T^*, \varphi^*)$ as T entails for φ (formally, $B(\varphi^*, T^*) = B(\varphi, T)[\varphi^*/\varphi]$). Note that 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 3 is then needed to prove that T^* entails $(\pi^* \leftrightarrow \neg \mu^*)$; whence π^* and $\neg \mu^*$ can be identified and collapsed into the required T^* -corresponding concept φ^* .

A much *simpler* proof of the correspondence theorem would be possible by replacing conditions 2 and 3 by the more direct and stronger requirement that T^* contains a common cause explanation φ^* for the correlated dispositions "if A_i , then R_i ". But this requirement would be too strong: T^* need not explain the correlated dispositions by postulating exactly one common cause; it may also explain them by assuming a more complicated net of causes or hidden variables.³ This is the point where corollary 2 comes into play: what can be proved with conditions 2 and 3 on T^* is only that T^* contains *some* measurable common cause, but not necessarily a *unique* one. T^* may contain several causes φ_1^* , ..., φ_k^* (with $k \leq n$). Corollary 2 tells us that in such a situation T^* entails 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 $\bigvee_{1 \leq i \leq k} \varphi_i^*(x)$. But this formal trick does not remove the possible ontological ambiguity. It does not follow from corollary 2 that on reasons of causal simplicity, T^* will always proclaim only one

instead of several non-identical common causes. For the equivalence of the φ_i^* 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. In the phlogiston-oxygen example of the next section, T^* does indeed contain exactly one counterpart cause of φ . But the Newton-Einstein example of section 4 will be a situation in which T^* contains two counterpart causes for φ which are non-equivalent outside of domain A.

2. *Correspondence between Phlogiston Theory and Oxidation Theory*

The phlogiston theory of combustion was developed by J.J. Becher and his student G.E. Stahl in the late 17th and early 18th century.⁴ According to this theory, 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 oxida-

tion 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, and nobody believes in the existence of phlogiston any more.

The three major domains of application in which both phlogiston theory and oxidation theory were strongly successful were (1.) combustion of organic material (which was thought to underlie also respiration and organic metabolism), (2.) calcination (roasting) of metals, (3.) salt-formation through the solution of metals in acids, and (4.) the inversion of calcination and salt-formation. The shared non-theoretical vocabulary of the two theories, in which these application domains were described, consisted of a shared classification of observable substances (metals, calx, water, acid, salts, ...) and empirically described reactions (combustion, roasting, acidification, salt-formation, ...). The different theoretical concepts and principles of the two theories concerned their theoretical decomposition of observable substances into hypothetical and unobservable components, which are assumed to be exchanged during chemical reaction processes. This led to a rather different theoretical analysis of the chemical reactions in the four domains which is summarized below. Hereby, we use the following *chemical notation*: an arrow stands for a chemical process, the substances mentioned left of the arrow are input substances and those right of the arrow are output substances; "+" (or "-") at a substance-component denotes that the corre-

sponding atomic or molecular group is electropositive (electronegative); '↑' at a substance means that the substance is an evaporating gas; and finally, substances in square brackets "[]" are residuals which did not participate in the chemical reaction but were there from the start. "Ox-T" stands for "oxygen theory" and (later on) for the generalized "oxidation theory", "Phlog-T" for "phlogiston theory".

Combustion and Calcination: While oxygen theory describes combustion or calcination of a substance X as a process of *oxidation*, i.e. a process in which oxygen is *added* to X, phlogiston-theory describes combustion and calcination of X as a process of *de-phlogistication*, i.e. a process in which X gives off its phlogiston:

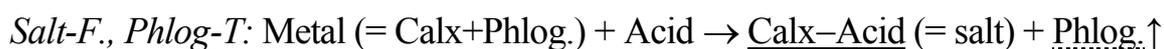
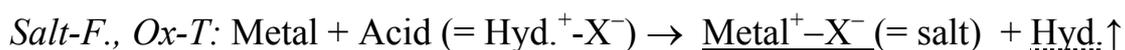


Phlogiston theory described carbonium and metal as phlogiston-rich chemical compounds, while for oxidation theory both were pure substances ('elements'). The similar underlining (drawn-thorough or dotted) of different substance-descriptions given by Ox-T and Phlog-T indicates that they describe the same em-

pirically given substance (this convention is also used below). In the resulting analysis of chemical reactions, end-products and residuals sometimes exchange their role – for example, phlogiston theory identified the evaporating fume (carbon-dioxide) in the combustion of coal or organic material with phlogisticated air and the ash with dephlogisticated coal; while for oxidation theory only carbon-dioxide is the proper end-product, and the ash is a residual of incompletely oxidized coal. Likewise, phlogiston theory identified the upstreaming hot air in the calcination process as phlogisticated air, while for oxidation theory it was a mere byproduct. The electric polarities of the substance-compounds were not part of Lavoisier's oxygen theory, but are the central result of generalized 20th century oxidation theory, according to which oxygen as well as all other oxidizing substances attract electrons in chemical bonds and, thus, become electro-negatively charged.

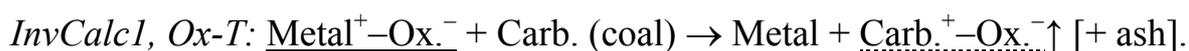
Salt-formation of metals with acids: For modern chemistry, every acid has the qualitative chemical composition 'Hyd.⁺–Y[–]', in which the hydrogenium atom ('Hyd.') gives off its electron to its negatively charged partner Y. If a metal is dissolved in acid, it takes over the role of hydrogenium and forms an ionic bond with the electronegative acid-component Y[–]. Pure hydrogenium evaporates as an easily inflammable gas. In contrast, phlogiston theorists described salt-formation as a dephlogistication of the metal, in which calx and acid combine to form salt, while the evaporating 'inflammable air' (hydrogenium) was identified by Cavendish with pure phlogiston (cf. McCann 1978, 32). This phlogiston-

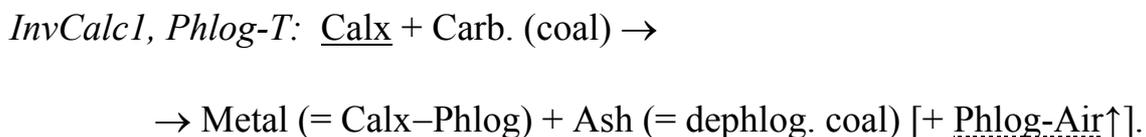
analysis of salt-formation was further confirmed by the fact that by strongly heating the salt one could regain the calx (oxide) of the metal.



Lavoisier believed erroneously that the electronegative acid-component Y must always contain oxygen. This is no longer assumed by generalized oxidation theory – for example, in hydrochloric acid $\text{Hyd}^+ \text{-Cl}^-$ the electronegative chlorine ion Cl^- (which does not contain oxygen) takes over the role of the oxidant (see also Hacking 1983, 9).

It was known from early times on that the process of calcination of metals can be reverted: metals can be extracted from calxes (i.e. mineral ores) through heating in charcoal. This process was also successfully described by both theories. Oxygen theory described this process as *reduction*, i.e. the inversion of the process of oxidation, in which the calx delivers its oxygen to the coal. In contrast, phlogiston theory described this process as *phlogistication*, i.e., the inversion of the process of dephlogistication, in which the (phlogiston-rich) charcoal spends its own phlogiston to the calx to form a metal:



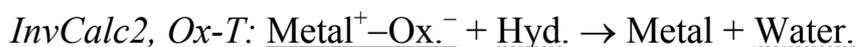


Note that in the analysis of this reaction by Ox-T and Phlog-T, proper end-products and residuals exchanged their role: phlogiston theory identified the evaporating fume (carbon-dioxide) with phlogisticated air (not all of the coal's phlogiston combines with the calx) and the ash as dephlogisticated coal; for oxygen theory, carbon-dioxide is a proper end-product, and the ash is a residual of incompletely oxidized coal.

We now turn to the illustration of our condition (2.1) on Phlog-T. Phlogiston theory entails a variety of potentially novel predictions. For example, phlogiston theory predicts that every substance x which has been gained by the process of inverse calcination from a mineral ore ($A_1x \wedge R_1x$) is a phlogiston-rich metal and hence will yield a salt if dissolved in acid ($A_2x \rightarrow R_2$) – even if x is was a hitherto unknown kind of substance which has never been put into an acid before. This is exactly the logical form of strong potential empirical success $\exists t(A_1xt \wedge R_1xt) \rightarrow \forall t(A_2xt \rightarrow R_2xt)$ as defined in § 1.⁵

A striking example of a successful novel prediction of phlogiston theory has been described by Carrier (2004): after Cavendish had identified inflammable air with phlogiston, Priestley predicted in 1782 that it should be possible to invert the process of calcination by adding inflammable air to a metal calx. He

heated several metal calxes in inflammable air and observed that the inflammable air was almost completely absorbed and that the calxes were slowly reconverted into the metals. This was celebrated as a further strong success of the phlogiston theory. Priestley had also recorded the emergence of water droplets in this reaction, but he assumed that the water was contained in the inflammable air from the beginning (Carrier 2004, 151). Thus, Priestley has performed a new kind of chemical reaction, the inversion of calcination by inflammable air (instead of charcoal), which is described by the two theories as follows:



The inversion principle for chemical reactions, which Priestley has used to obtain his novel prediction, is a theoretical principle which was shared by Phlog-T and Ox-T. It can logically be reconstructed as follows: for each bilateral reduction sentence B_i describing an dephlogistication (oxidation) process, an inverted B_i holds, abbreviated as $\text{Inv}B_i$, which describes a phlogistication (reduction) process ($\forall i \in \{1, \dots, n\}: B_i \leftrightarrow \text{Inv}B_i$). Let $\text{In}(x, r)$ and $\text{Out}(y, r)$ denote that x is an input and y an output substance of a chemical reaction r , and ' $\varphi(x, y)$ ' stand for 'phlogiston moves from x to y '. Then a B_i for dephlogistication of x has the more refined form ($\forall r, x, y$): $\text{In}(x, r) \wedge \text{Out}(y, r) \wedge A(x) \rightarrow (\varphi(x, y) \leftrightarrow B(y))$, where A and B

denote kinds of substances under certain conditions, and the $\text{Inv}B_i$ for phlogistication of x has the form $(\forall r, x, y): \text{In}(y, r) \wedge \text{Out}(x, r) \wedge B(y) \rightarrow (\varphi(y, x) \leftrightarrow A(x))$. Together B_i and $\text{Inv}B_i$ entail the novel prediction $\exists r(\text{In}(x, r) \wedge \text{Out}(y, r) \wedge A(x) \wedge B(y)) \rightarrow \forall r(\text{In}(y, r) \wedge \text{Out}(x, r) \wedge B(y) \rightarrow A(x))$.

In conclusion, Phlog-T was a rather successful theory in the mentioned domains, provided one assumes (as we did) that these domains are described by means of *qualitative* chemistry, i.e., without a quantitative analysis of masses and mole numbers. This was the reason why Phlog-T was the dominating theory in the 18th century, in which chemistry was mainly qualitative. Of course, Phlog-T had also some purely qualitative explanation problems. For example, phlogiston theorists were unable to isolate phlogiston in a domain-independent way: Cavendish's identification of phlogiston with inflammable air (hydrogenium gas) did not work in other domains such as the combustion of coal which has phlogisticated air (carbon dioxide) as its end product: it was never possible to gain inflammable gas from phlogisticated air. But phlogiston theory was not given up because of these problems, and several ad hoc hypotheses were invented to explain these difficulties (see Carrier 2006, §6). Also Lavoisier's oxygen theory faced some qualitative explanation problems. For example, we have mentioned that Lavoisier's assumption that the acid's oxidation power was always the effect of the oxygen contained in the acid was refuted by hydrochloric acids and other halogen acids. These difficulties of Lavoisier's theory have been solved only much later by the *generalized* oxidation theory.

The major reason why most chemists turned from Phlog-T to Ox-T in the very late 18th and beginning of the 19th century did not lie in its problems in the area of qualitative chemistry, but rather in the upcoming new methods and theories of *quantitative chemistry* (see also McCann 1978, 69, 78f). These were, in particular, the exact measurement of the weights of input and output substances of chemical reactions, and the acceptance of the *law of constant proportions* (so named by Proust, but first formulated by Lavoisier), according to which the sum of the weights of the input substances must balance that of the output substances. The problem was that the calcination of metals as described by Phlog-T seemed to violate this law, because assuming that phlogiston has positive or at least non-negative weight, the calx (= dephlogisticated metal) should weigh less than the metal, but it was found to weigh more. Also in regard to these difficulties, phlogiston theorists first reacted with ad hoc explanations. For example, Louis-Bernard Guyton de Morveau argued that phlogisticated metals are lighter than dephlogisticated calxes because phlogiston is lighter than air and, hence, tends to lift the body upwards, like a balloon of helium gas.⁶ This sounded plausible to many chemists, but scientists which were trained in quantitative physics found this argument to be defective: if equal volumes of a metal versus calx are put on two sides of a pair of scales, then the buoyancy force due to the displacement of air is equal on both sides and, hence, plays no role. In order to explain the lower weight of the metal, one would have to assume that phlogiston has negative weight – which sounded very implausible and was not believed by

the majority of (former) phlogiston theorists. A more refined 'double cause' ad hoc explanation was suggested by Pierre J. Macquer in 1779: he assumed that phlogiston is the bearer of light (cf. Carrier 2006, §6). During combustion and calcination oxygen gets fixed as described by Lavoisier, and the fixed oxygen causes the increase of weight in the calx, but Macquer argued that the fixation of oxygen causes the release of phlogiston, which in turn causes the emission of light in the form of a flame. Macquer also argued that his hypothesis was able to explain the reduction of mercury calx with light from a burning glass, which Lavoisier cited as an anomaly for phlogiston theory. Although it was hard to refute Macquer's double-cause hypothesis (oxygen and phlogiston), it could not compete with the much simpler oxygen theory of Lavoisier, which explained the reduction of mercury calx simply as the effect of heat.

We now turn to the illustration of our condition 2.2 on phlogiston theory. Which theoretical expression ϕ can be found in Phlog-T (if any) which is characterized by empirical indicators in the form of bilateral reduction sentences of the form $(\forall x,u): A_i(x,u) \rightarrow (\phi(x) \leftrightarrow R_i[x,u])$ which entail Phlog-T's strong empirical success? Interestingly, this expression is not "phlogiston" itself, because – as we have explained – phlogiston theory did not provide a general (domain-independent) criterion of how phlogiston viewed as a kind of substance could be empirically identified. Thus, the theoretical kind term 'phlogiston' was empirically underdetermined in Phlog-T, and bilateral reduction sentences of the required sort did not exist for phlogiston. Rather, the theoretical expressions of

phlogiston theory which did all the empirically relevant work and were *not* empirically underdetermined were the expressions of *dephlogistication* = the process of releasing of phlogiston and *phlogiston-richness* = the tendency of a substance to release phlogiston, as well as the inverse expressions of *phlogistication* (assimilation of phlogiston) and *phlogiston-poorness* (tendency to absorb phlogiston). For these expressions Phlog-T entails indeed bilateral reduction statements of the required sort, different ones for different domains A_i , namely the following:

B_i's for dephlogistication [resp. for *phlogiston-richness*]: If an input substance x of kind X_i (e.g., a metal) is exposed to the influence of an input substance y of type Y_i at time t (e.g. hydrochloric acid), then x gets dephlogisticated at t [or: x is phlogiston-rich] iff $(\exists z:)$ x and y cause a chemical reaction which produces output substances z of type $Z_i = f_i(X_i, Y_i)$ at t (e.g., the metal dissolves and inflammable air evaporates).

These B_i 's have exactly the general form (B_i) explained in § 1.⁷ Note that while "dephlogistication" refers to a time-dependent process, "phlogiston-richness" refers to an intrinsic, i.e. non-time-dependent property. The B_i 's for 'phlogistication' and 'phlogiston-poorness' are obtained in an analogous way.

Conditions 2 and 3 have already been motivated in the previous section. So we proceed immediately to the illustration of our correspondence theorem.

There does indeed exist a correspondence of Phlog-T with generalized oxidation theory of modern chemistry, which goes further than the identification of phlogistication with oxidation in Lavoisier's sense. To explain this correspondence we need a bit more of modern chemistry (cf. Oxtoby et al. 1999, ch. 3, ch. 6.3). Every substance consists of molecules, and molecules consist of atomic elements bound together by chemical bonds. The *electropositivity* of an element measures its tendency to contribute electrons to its neighboring atoms in electrically polarized or ionic bonds. Conversely, the *electronegativity* measures the tendency of an element to attract electrons from the neighboring atom in polarized or ionic bonds. Metals and hydrogenium are electropositive; carbonium is in the middle of the spectrum, and non-metals such as oxygen are electronegative, with the extremes being the halogens. Oxidation of an elementary substance X (a metal, coal, etc.) in the *generalized sense* consists in the formation of a polarized or ionic bond of X with an electronegative substance Y, in which the atomic elements of X are electropositive and thus donate electrons to the electronegative neighbor Y in the bond. Every process of combustion, calcination or salt-formation consists of such an oxidation process. The inversion of the process of oxidation is called the process of reduction: here the polarized or ionic bond between an electropositive X-ion and its electronegative neighbor is broken, X regains its missing electrons and reappears in its pure elementary form. Therefore we have the following correspondence relations between phlogiston theory and modern chemistry:

Correspondence relations between phlogiston theory and modern chemistry:

Phlogiston-richness of X corresponds (and indirectly refers) to the *electropositivity* of X.⁸

Phlogiston-poorness of X corresponds (and indirectly refers) to the *electronegativity* of X.

Dephlogistication of X corresponds (and indirectly refers) to the *donation of electrons of X-atoms* to their bonding partner in the formation of a polarized or ionic chemical bond.

Phlogistication of X corresponds (and indirectly refers) to the *acceptance of electrons by positively charged X-ions* from their bonding partner in the breaking of a polarized or ionic chemical bond.

These relations of correspondence and indirect reference hold indeed generally, i.e., it is not the case that Ox-T would contain several possible counterpart concepts for "phlogiston-richness" and "dephlogistication" (recall the discussion at the end of § 1). The failure of phlogiston theory was that dephlogistication of a substance x was thought of as a process in which a special chemical part of x, called phlogiston, *leaves* x. The electrons do not leave the substance but just move a little bit to the electronegative neighbors in the molecules. What really is emitted as the end product of an oxidation process (besides the oxidized material) depends on the oxidant, that is, the input substance which causes the oxida-

tion and which spends the electronegative partner – oxidation of coal by oxygen leads to carbon dioxide, while oxidation of a metal with acid leads to hydrogen gas. Therefore we have no correspondence for phlogiston. But we have correspondence for the composite terms of dephlogistication and phlogiston-richness, and these correspondence relations *explain* the strong empirical success of phlogiston theory in terms of modern chemistry.

It is clear from the foregoing that correspondence relations do *not* preserve *all* of the meaning of 'phlogistication' or 'phlogiston-richness'. Therefore, they cannot be regarded as an *analytic* truths, but have to be regarded as a *synthetic* statements which are true in the domain of applications in which phlogiston theory was empirically successful. The fact that the corresponding expressions φ of Phlog-T are not primitive but composite expressions is ontologically crucial. For 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.

I think that 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 B_i 's. 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 (quantum mechanics is a case in point; cf. Ladyman 1998, 418f).

My 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 which is not threatened by the objections of Psillos (1995, 31ff): 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. To elaborate the notions of 'inner' and 'outer structure' in a more general way would be an important task for future work.

3. *Non-unique Correspondence: Classical and Relativistic Mass* ⁹

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 both examples, there exists a unique correspondence concept φ^* in T^* . But not always is the situation so nice. In this section we discuss an example of non-unique correspondence: the correspondence between classical and relativistic mass. In an early paper Field (1973) has used this example to argue for the indeterminacy of the reference of the theoretical concept of the mass $m_N(x)$ of a physical body x in Newtonian physics T , thereby assuming special relativity theory T^* as the true theory. We show now how we can make sense of Field's argument within the framework of the correspondence theorem.

Special relativity theory T^* has two concepts of mass, the rest mass $m_0(x)$, and relativistic mass $m_r(x,f) := m_0(x) / \gamma(x,f)$, with $\gamma(x,f) := \sqrt{1 - \frac{v(x,f)^2}{c^2}}$, which depends on the relation of the velocity $v(x,f)$ of x as measured in inertial frame f to c (the velocity of light). Field (1973, 467-470) argues that some parts of what

we have called the 'outer structure' of Newtonian mass, for example its relation to momentum $p(x,f) = m_N(x) \cdot v(x,f)$, are preserved in T^* if m_N is identified with m_r . Other parts of T 's content, as for example the conservation of m_N in different inertial frames, i.e. m_N 's velocity-independence, are preserved in T^* by identifying m_N with m_0 . This part of T 's content can only be formulated if Newtonian mass is augmented by an additional argument place f for the inertial frame in which m_N is measured; – then it says simply that $(\forall f,f')(m_N(x,f) = m_N(x,f'))$.

Field concludes from this situation that it is objectively undetermined whether Newton's mass refers to rest mass or to relativistic mass. Earman (1977) objected to Field that the four-dimensional coordinate-free reformulation of Newtonian physics and relativity theory provides independent reasons for identifying m_N with m_0 (in an appendix to Earman 1977, Fine added evidence that also Einstein thought that m_N should be identified with m_0). Nevertheless the fact remains that some of T 's content parts are preserved by identifying m_N with m_r but are lost if m_N is identified with m_0 .

In our account, the situation is resolved as follows. First of all, we have to identify the domain of application A_N in which Newtonian physics T had strong success which was preserved in relativistic physics T^* . Obviously, this is the domain in which the velocities are small compared to c . More precisely, the velocities $v(x,f)$ should be small enough to ensure that $\gamma(x,f)$ deviates from 1 by less than a given measurement inaccuracy ε – let us call this upper velocity

bound $v(\varepsilon)$ (calculation yields $v(\varepsilon) = c \cdot \sqrt{\varepsilon \cdot (2 - \varepsilon)}$). Thus, the domain description ' $A_N(x, f)$ ' is defined as ' $v(x, t) \leq v(\varepsilon)$ '. Next, in the case of quantitative theories, correspondence relations do not have the form of equivalences but of identities (recall the remark below the schema (B_i) in § 1), and moreover, they do never hold in a strict but only in an approximate form. Thus, the correspondence relation between m_N of T and the corresponding mass concept m^* of T^* (whatever it is) will have the form

(Correspondence for mass): $(\forall x, f): A(x, f) \rightarrow (m_N[x, f] \cong_\varepsilon m^*[x, f])$.

Here " \cong_ε " stands for ε -approximate identity (precisely defined below), and in the formulation of " $m^*[x, f]$ " we have left it open whether m^* is rest mass $m_0(x)$ or relativistic mass $m_r(x, f)$ (so we have chosen the square bracket notation ' $[x, f]$ ' for the free variables of m_N as well as of m^* ; recall the 'notation' below ' $S_{i,j}$ ' in § 1).

Let us now see how a correspondence between T and T^* can be derived along the lines of our correspondence theorem. The domain $A(x, f)$ divides into several qualitatively different subdomains (earthly motions described by ballistic curves, mechanical oscillations, planetary motions, etc.). Within each domain, m_N is characterized by measurement laws (bilateral reduction statements) of the form $(B_i): A_i(x, f) \rightarrow (m_N[x, f] = r_i[x, f])$; for example "if x is put on a balanced beam, then $m_N(x) = k$ grams iff the number of one gram units ($r_i(x)$) on the other

side of the balanced beam is k ". These measurement laws for m_N entail various potentially novel predictions of the (variable-suppressed) form $(S_i:)(A_i \wedge R_i) \rightarrow (A_j \rightarrow R_j)$, for example, "if a body x balances k standard units on a pair of scales, then it will compress a given spring balance by such-and-such amount". All these novel predictions (S_i) are approximately preserved in special relativity theory T^* , in a T^* -dependent way, i.e. by way of T^* -theoretical functional expressions $m^*[x,t]$ which figure as theoretical mediators between $(A_i \wedge R_i)$ and $(A_j \rightarrow R_j)$. Special relativity theory T^* has two such theoretical mediators to offer, rest mass m_0 and relativistic mass m_r . The *key point* to observe here is that within the domain of application $A(x,f)$, both possible theoretical counterparts are approximately equivalent, i.e. it holds that

$$(*) \quad (\forall x,f): A(x,f) \rightarrow m_0(x,f) \cong_\varepsilon m_r(x,f),$$

whereby the precise meaning of " $m_r \cong_\varepsilon m_0$ " as following from the above convention for $v(\varepsilon)$ is given as " $(m_r - m_0)/m_0 \leq \varepsilon$ ". (More generally, ' $x \cong_\varepsilon y$ ' is definable as ' $|x - y| \leq \varepsilon \cdot \min(\{x, y\})$ '.) Indeed, (*) follows already from corollary 2 and the assumption that both T^* -mediators m_0 as well as m_r work in all subdomains of $A(x,f)$ – which is, of course, the case, because the defining condition for $A(x,f)$, namely $1 - \gamma(x,f) \leq \varepsilon$, is mathematically equivalent with $(m_r - m_0)/m_0 \leq \varepsilon$. In the present example, however, T^* contains m_r and m_0 as distinct theoretical con-

cepts: although they are approximately identical in the domain $A(x,f)$ in which T was successful, they depart from each other significantly in situations of velocities which are significantly higher than the threshold $v(\epsilon)$. This is exactly the situation of non-unique correspondence relations between T and T^* , whose theoretical possibility was discussed at the end of § 1. It follows that for the corresponding mass concept $m^*[x,f]$ of T^* one might choose $m_0(x)$ or $m_r(x,f)$ or even an average of the two (which would correspond to a 'disjunction' in the qualitative case).

In conclusion, the considerations of this section have shown that the correspondence theorem can also be applied to quantitative theories such as classical versus relativistic mechanics, provided strict identities are replaced by approximate ones. More importantly, these applications may reveal important insights, such as the existence of several non-identical relativistic counterparts to Newtonian mass, which are approximately identical in the domain of application in which Newtonian physics was successful.

4. Outlook: Consequences for the Justification of Scientific Realism

In this final section I want to give a brief sketch of how the correspondence theorem may justify a weak version of scientific realism. Most scientific realists base their arguments on some version of the no miracles argument (NMA) which was explained in § 1. What my account shares with the critics or scien-

tific realism is the skeptical attitude concerning the *unrestricted* NMA. The major argument against the NMA in my view is the fact explained in § 1: if the theory T does not satisfy the common cause condition (2), but *speculates* for each kind of phenomenon a special cause, then there is *no way* to infer from its empirical success something about the truth-status of its theoretical part. Therefore the unrestricted version of the NMA is *false*. Only a restricted version of it is tenable.

My account does in no place presuppose the NMA. It is based on an *analytic* theorem. Together with empirical-historical evidence concerning the truth of conditions 1-4, the correspondence theorem establishes independently from the NMA or some other form of IBE (inference to the best explanation) *why* a specifically *restricted* version of the NMA can be defended. 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 reality which causes the observed phenomena in a humanly accessible language is possible – independent of whether humans will ever find this theory. 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 strong empirical success which our accepted theories have. So the correspondence theorem implies that every (theoretico-structural) content-part of our contemporary theories satisfying the conditions 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 may provide 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.

A detailed elaboration and defense of this justification of a weak version of

scientific realism must be left to another paper. But let me emphasize that the correspondence theorem need not be interpreted in this realistic way. It makes also good sense without the assumption of minimal realism. In the non-realistic interpretation of the correspondence theorem, the ideal theory 'T⁺' would no longer reflect the structure of an external reality, but would be an *idealized construction* expressing that the historical succession of more and more successful theories converges to some fictitious end (e.g., in the sense of Charles S. Peirce). Such a notion of convergence by correspondence makes sense in my account, because the notion of correspondence between theories can be shown to be *transitive*.¹⁰ Convergence to a limit theory T⁺ by way of correspondence relations has to be understood in terms of less-and-less partial truth: that part of the theoretical superstructure which matches the structure of the ideal theory T⁺, or in realists terms, the structure of external reality, becomes more and more comprehensive. In this sense, my account satisfies Carnap's requirement of *metaphysical neutrality*.

Acknowledgements: For valuable comments I am indebted to Bas van Fraassen, Kevin Kelly, Hans Rott, Michael Friedman, James Ladyman, Brandon Fitelson, Patrick Suppes, and Edward Zalta.

References

- Boyd, R. (1984). The current status of scientific realism. (In J. Leplin (ed.), *Scientific realism* (pp. 41-82). Berkeley: University of California Press.)
- Carrier, M. (2004). Experimental success and the revelation of reality: the miracle argument for scientific realism. (in M. Carrier et al. (eds.), *Knowledge and the world: challenges beyond the science wars* (pp. 137-61). Heidelberg: Springer.)
- Carrier, M. (2006). Antoine L. Lavoisier und die Chemische Revolution. (In P. Leich (ed.), *Leitlinien naturwissenschaftlichen Denkens*. Duisburg: Königshausen & Neumann.)
- Field, H. (1973). Theory change and the indeterminacy of reference. *Journal of Philosophy*, 70, 462-481.
- Earman, J. (1977). Against indeterminacy. *Journal of Philosophy*, 74, 535-538.
- Hacking, I. (1983). *Representing and Intervening*. (Cambridge: Cambridge Univ. Press)
- Ladyman, J. (1998). What is structural realism? *Studies in the History and Philosophy of Science*, 29/3, 409-24.
- Laudan, L. (1981). A Confutation of Convergent Realism. (Reprinted in D. Papineau (ed.), *The philosophy of science* (pp. 107-38). Oxford: Oxford Univ. Press.)
- Oxtoby, D. et al. (1999). *Principles of modern chemistry*. (Orlando: Saunders

College Publishing)

McCann, H. Gilman (1978). *Chemistry transformed: the paradigm shift from phlogiston to oxygen*. (Norwood, N.J.: Ablex Publ. Corporation)

Psillos, S. (1995) Is Structural Realism the Best of Both Worlds? *Dialectica*, 49/1, 15-46.

Putnam, H. (1975). What is Mathematical Truth? (In H. Putnam, *Mathematics, Matter and Method* (pp.60-78). Cambridge: Cambridge Univ. Press.)

Schurz, G. (2008). Patterns of abduction. *Synthese*, 164, 201-234.

Schurz, G. (2009). When empirical success implies theoretical reference: a structural correspondence theorem. To appear in: *British Journal for the Philosophy of Science*.

Worrall, J. (1989). Structural realism: the best of both worlds. *Dialectica*, 43/1-2, 99-124.

Notes

¹ The existence of a $\varphi^*_i(x)$ which is independent from j is required in Schurz (2008, §4, 'definition'); but in §3 of the same paper it is shown that the requirement of i,j -dependent mediators is sufficient, because $\varphi^*_i(x)$ is definable as $\bigwedge \{\varphi^*_{i,j}: 1 \leq j \leq n, j \neq i\}$.

² Apart from that qualification, I do not assume a particular theory of reference.

³ This critical remark has been independently made by Bas van Fraassen and Kevin Kelly.

-
- ⁴ in his 1697 paper "Zymotechnia Fundamentals", Georg. E. Stahl coined the term "phlogiston" as a renaming of 'terra pinguis' introduced by his teacher Johann. J. Becher. In his "Physical Education" 1667, Becher modified the traditional (Aristotelean and alchemist) four-element model (earth, water, air and fire) by replacing 'air' and 'fie' by 'terra pinguis' as the 'fiery element' which is released when substances are burned (see "prelude to modern chemistry", <http://hilltop.bradley.edu/~rbg/com.html>).
- ⁵ Novel predictions of scientists will often have the apparently different form 'an x of kind α will exhibit reactions $R_j(x)$ under conditions $A_j(x)$ ', where ' α ' is a theoretical statement. But the *empirical indicators* for ' x is of kind α ' will (after some transformations) again have the form 'under conditions $A_i(x)$, $R_i(x)$ has happened'.
- ⁶ Cf. Wikipedia (English), entry "phlogiston" (en.wikipedia.org/wiki/Phlogiston).
- ⁷ I am indebted to Edward Zalta for getting the free variables of this example right.
- ⁸ James Ladyman pointed me towards the correspondence for phlogiston-richness.
- ⁹ I am grateful to Hans Rott for pointing me to this example.
- ¹⁰ This possibility was pointed out to me independently by Michael Friedman and Brandon Fitelson. A detailed proof is left to further papers.