Philosophica 32, 1983 (2), pp. 25–42.

SCIENTIFIC CHANGE AND INTENSIONAL LOGIC

Antti Hautamäki

In this paper an analysis of scientific theories and theory change including meaning change is presented by using intensional logic. Several cases of scientific progress are distinguished and special attention is given to incommensurability. It is argued that in all cases the comparison of rival theories is possible via translation. Finally two different forms of theory-ladenness of observation are analysed.

1. One of the most disputable questions in modern philosophy of science is the one of incommensurability. One argument for the incommensurability of successive scientific theories is the claim of meaning change presented by Feyerabend and Kuhn. They have claimed that changes in theory result in changes of meaning : meanings are theory-dependent[1]. The incommensurability-thesis remains, however, vague until some theory of meaning and meaning change is adopted. Only on the basis of a meaning-theory can such related questions like the separation of a factual and a conventional (analytical) component of an empirical theory and the intertranslatability of the languages of different theories be settled[2].

Carnap has proposed that the factual and conventional component in a theory can be distinguished by using *Ramsey sentences*[3]. They can be presented as follows. If P is the conjunction of theoretical and correspondence postulates of a theory T, then ${}^{R}P$ is a Ramsey sentence of T if ${}^{R}P$ is obtained from P by proper simultaneous substitution of predicate variables for theoretical terms and by taking the existential closure of the formula thus obtained. Then the factual part of T is identified with the consequence class of ${}^{R}P$ and its conventional part with the consequence class of ${}^{R}P \rightarrow P$.

Another way to represent the distinction of factual and conventional component of theories is to express the conventional part by *meaning postulates*, which must be separated from the factual and mathematical postulates of a theory[4].

A more general way than these two is to use a set of models to represent the analytical component of a theory. For example P. M. Williams defines a *semantical system* or an interpreted language to be a trible $\mathcal{L} = \langle L,A,M \rangle$, where L is an elementary language, A is a set of structures for L that is closed under isomorphism and $M \subseteq A$.[5] Now the theory of A, Th (A), is the set of *analytic* sentences of L, and Th(M) is the set of *true* sentences of L. Also D. Pearce and V. Rantala represent the analytical component of a theory by a class of models[6].

The above approaches have their advantages but they have their drawbacks, too. First of all they don't refer explicitly to meanings, which makes it difficult to speak about meaning change. I think that a better way to handle meanings is to use possible worlds semantics and the means it offers to represent intensions[7]. It is very difficult to reach the advantages of intensional logic if it is started plainly from the extensional model theory.

In chapter 2 I will give a short presentation of intensional semantics. The crucial thing is to decide how to present theories in possible worlds semantics. This is done in chapter 3. After that I am going to discuss about various relations of theories (chapter 4). Chapter 5 is devoted to the problem of incommensurability. The problem of theory-ladenness of observation and the observational/ theoretical dichotomy is tackled in chapter 6. Let me note that the above problems are considered in this paper from the logical and theoretical point of view. It is a task of another paper to apply this approach to concrete, empirical cases.

2. *Possible worlds semantics of first order languages.*

In this paper I deal only with first order languages, but my treatment is to a great extent independent of this restriction[8]. Every set of predicate and function symbols is said to be a *language*. Formulas of a language L are built by using symbols in L and logical symbols $\sim$, $\wedge$, v, $\rightarrow$, $\leftrightarrow$, $=$, $\forall$, $\exists$, x, y, x, ..., and parentheses (and) in a normal way[9]. The set of *sentences* of L is denoted by S_L.

We interpret first order languages by using possible worlds semantics. A triple $F = \langle W,D,U \rangle$ is said to be a *frame* if W is a non-empty set (of *possible worlds*), D is a non-empty set (of *possible individuals*) and U is a function with domain W such that U(w) (*universum* of w) is a non-empty subset of D ($w \in W$). A *valuation* for L in F is a function f with domain L such that (i) whenever P is an n-place predicate symbol in L and $w \in W$, f(P)(w) is an n-place relation on U(w), and (ii) whenever g is an n-place function symbol in L and $w \in W$, f(g)(w) is an n-place function on U(w). The functions f(P) and f(g) are said to be the *meanings* of P and g relative to F and f. If $F = \langle W,D,U \rangle$ and f is a valuation for L in F, then $\langle F,f \rangle$ is a *model* for L. The notions of satisfaction and truth at w in a model $\langle F,f \rangle$ are defined as usual[10]. We denote the *truth set* of a sentence $\varphi \in S_L$ relative to a model $\langle F,f \rangle$ for L by $[\varphi]_f^F$. The truth set $[\varphi]_f^F$ is said to be the *meaning* of φ relative to F and f (relative to a model $\langle F,f \rangle$). We generalize the notion of truth set to cover sets of sentences in a natural way :

$$[X]_f^F = \bigcap_{\varphi \in X} [\varphi]_f^F \quad \text{where } X \subseteq S_L.$$

From this definition it follows that $[\emptyset]_f^F = W$.

Now we define some important concepts. Let $F = \langle W,D,U \rangle$ be a frame and f be a valuation for a language L in F. Then the sets

$$Th_f^F(M) = \{\varphi \in S_L : M \subseteq [\varphi]_f^F\}, \text{ where } M \subseteq W,$$

$$Cn_f^F(X) = Th_f^F([X]_f^F), \text{ where } X \subseteq S_L, \text{ and}$$

$$An_f^F = Th_f^F(W)$$

are called the *theory of M, the set of consequences* of X, and the set of *analytic* sentences relative to F and f respectively[11]. It is meant that analytical sentences are true in every possible world.

3. *Theories*

There is a considerable disagreement as to the structure of empirical theories[12]. According to the 'statement view' theories are deductively closed sets of sentences of some formal language. After the 'structuralist view' of theories, theories are non-linguistic sets of mathematical structures. The 'structuralist view' represents the so called *semantic conception* of theories. The semantic conception has been little concerned with the meaning of scientific terms and meaning change[13]. This is a direct consequence of the conception that theories are non-linguistic entities.

Let me point out that D. Pearce and V. Rantala have proposed a new approach to the analysis of theories, where they can maintain the structuralistic perspective on theories and re-introduce linguistic concepts at the same time[14]. Their definition of theory is based on general model theory. According to them a theory is a quadruple $T = \langle \tau, N, M, R \rangle$ where τ is a similarity type, $N \subseteq \mathrm{Str}(\tau)$, $M \subseteq N$, and R is a collection of relations on N. Here M represents T's 'laws' and N its analytical component[15].

Intuitively a theory consists of a set of concepts and a set of laws. The latter set specifies what is believed about the phenomena in the intended scope of the theory. My idea is to represent both components of a theory semantically as follows. First we construct a frame $F = \langle W, D, U \rangle$ that is sufficiently rich for representing the concepts and laws of the theory. For example, D can contain real numbers and triples of real numbers and so on[16]. The set of concepts can be obtained with a suitable valuation function f in F for some relevant language L. L must contain at least one symbol for each concept of a theory; for example if a theory contains an n-place function then L must contain an n-place function symbol. A valuation function f is suitable if it assigns to every symbol of L an intensional relation or function, that expresses the intuitive meaning of the corresponding concept. It is crucial that f assigns to corresponding symbols different values if the intuitive meanings of two concepts differ. This presupposes that there are in W so many worlds that always when two concepts differ it is possible to find a world, where the references of these concepts differ. So we have a meaning triple :

intuitive concept — term t — meaning of t,i.e.f(t).

In formal presentation the meaning of t is identified with the intuitive concept and so we can say that the range R(f) of f is the *set of concepts* of a theory.

The set of laws of a theory is represented by a set M of possible worlds. Intuitively M is the set of those possible worlds which satisfy the laws of theory. The set M is not given linguistically by some set of axioms. In this respect I follow the structuralist approach[17]. Now I am ready to define what I mean by a theory.

Let $F = \langle W, D, U \rangle$ be a frame. A triple

$$T = \langle L, f, M \rangle$$

is said to be a *theory in F*, if L is a language, f is a valuation for L in F and M is a subset of W. We shall say that L is the set of *terms* of T, f is the *framework* of T and M is the set of T's *models*[18].

The elements of S_L are called *sentences* of T. The set $Th_T = Th_f^F(M)$ is said to be the set of *theorems* of T and the set $An_T = An_f^F$ is said to be the set of *meaning postulates* of T. If $M \neq \emptyset$, T is said to be *consistent*. If T is inconsistent, then $Th_T = S_L$.

The definition of theory given here can be compared to the standard notion of theory in formal logic[19]. There a formal theory is identified with a couple $T = \langle L, X \rangle$, where L is the set of proper symbols of T and X is the set of axioms of T. *Every model* of L in which axioms are true, is a model of T. From these definitions it follows that formal theories cannot represent adequately intuitive theories and their concepts : L can namely be interpreted arbitrarily. It has been said that the axioms of a theory define implicitly the terms of the theory. But this is an illusion, because axioms put only purely formal constraints on models. Also the effort to give meaning to terms by a set of meaning postulates fails, because meaning postulates are only formal sentences[20]. In the approach I have proposed a sentence is a meaning postulate only if it is true in every world and the set of possible worlds is specified without referring to meaning postulates.

We do not meet the problem of empiric interpretation of theoretical terms, because we suppose that every term of a theory is previously understood and so we may construct the framework of the theory for the whole language L. It is just the framework that gives meaning to terms, not axioms or meaning postulates.

The set M of T's models is not specified with a set of axioms. This is an important point, because only in this way we can exclude from the set M some unintended models that are, for example, isomorphic with some models in M. Moreover, in empirical theories the intended models are always partially specified by nonlinguistic means like pointing, measurements or operations. So we can think that M is specified by this kind of non-verbal means. Of course, sometimes the set M is *axiomatizable* in the sense that there is a set X of sentences such that $M = [X]_f^F$. But I do not assume that every theory is so axiomatizable. Every sentence of L, that is true in all worlds in M, is a theorem of T. In this way theorems are semantical consequences of M. If M is not axiomatizable, then we can not express the notion of theorem in terms of deduction.

What is novel in the above conception of theory, is first that theories are presented in the framework of possible worlds semantics.[21]. This is linked with two points. 1. A set of possible worlds or a frame provides a kind of "state space" of the real world or its aspects[22]. We can think that always one state of this space is realized, viz. the actual world. 2. Concepts of theories are interpreted by intensional relations and functions, i.e. by functions that assign to every world a reference of a concept[23]. Another novel feature is that the interpretation of languages is not fixed in the frame, and so there is a great liberty to select an appropriate valuation when a theory is specified. Note that the same symbols can be interpreted in several ways. This makes it possible to study *meaning change* in great details.

4. Relations of theories.

How can different theories in scientific progress be compared ? As we shall see theories are comparable if they can be represented as theories in the *same frame.* But can all theories be represented in the same frame ? This may be practically impossible and thus a more reasonable *assumption* is that two (or more) rival theories within one discipline can be so represented. This can be done as follows. We select a set of possible worlds such that we can present every concept of both theories. This means, firstly that the set D of possible individuals contains all the objects needed in both theories and secondly that if two concepts differ there must be a world where the extensions of these concepts differ, too. Finally we must express the laws of both theories by suitable sets of possible worlds.

But is this assumption realistic in regard to the real scientific progress and its possible revolutions ? I am going to show below that two theories can be incommensurable in a sense that will be specified later although the frame of both theories is the same. Thus my assumption does not exclude a priori any interesting case of scientific change.

Let $T = \langle L,f,M \rangle$ and $T' = \langle L',f',M' \rangle$ be two theories in the same frame $F = \langle W,D,U \rangle$. We shall suppose that T' *supersedes* T in the development of science and we shall indicate this by notation T/T'.

We can compare T and T' in regard to their frameworks and to their sets of models. As to frameworks the following two cases are important : $f \subseteq f'$ and $f \cup f'$ is not a function.

If $f \subseteq f'$ then we say that f' is an *expansion* of f and that f is a *reduct* of f'. If f' is an expansion of f then

1. $L \subseteq L'$, and

2. if $s \in L$ then $f(s) = f'(s)$.

If $f \cup f'$ is a function, it is a valuation for $L \cup L'$ and it is said to be a *common expansion* of f and f'. In expansion new symbols are added but the old symbols retain their old meanings.

If $f \cup f'$ is not a function then there is a symbol s that is common to both theories, i.e. $s \in L \cap L'$, and whose meaning differs in the frameworks f and f' :

$f(s) \neq f'(s)$.

Conversely, if the meaning of a symbol s differs in f and f' then $f \cup f'$ is not a function. But if $f \cup f'$ is a function then the meaning of every common symbol in T and T' is the same. Thus the *criterion* of *meaning change* is that $f \cup f'$ is not a function. But does meaning change exclude the comparison of rival theories ? Not at all, because we can, so to speak, compare theories at *metalevel*, i.e. "from outside"[24]. We can proceed as follows. We construct a language L" and a 1–1 valuation function such that $D(f'') = L''$ and $R(f'') = R(f) \cup R(f')$. Two translation functions Tr and Tr' may be defined as follows : if $t \in L$ and $f(t) = f''(t'')$, then $Tr(t) = t''$, and similarly for Tr'. These functions can be extended easily to functions that translate every sentence of S_L or $S_{L'}$ to a sentence in $S_{L''}$. The above procedure means that we can always reconstruct two theories such that there is no meaning change between them and that the union $f \cup f'$ of their frameworks is a function.

It is best to compare sets of models in relation to different cases of scientific progress. Let me first consider some cases of cumulative change. We suppose that T' supersedes T, where $T = \langle L,f,M \rangle$ and $T' = \langle L',f',M' \rangle$.

a) $f \subseteq f'$ and $M = M'$.

In this case f' is an expansion of f and every theorem of T is a theorem of T' and every theorem of T' that is a sentence of L is also a theorem of T. So we can say that T' is a *conservative extension* of T.

b) $f = f'$ and $M' \subseteq M$.

In this case Th_T can be a proper subset of $Th_{T'}$ but the meaning postulates of T and T' are the same : $An_T = An_{T'}$. T' is said to be a *factual extension* ot T.

c) $f \subseteq f'$ and $M' \subseteq M$.

Now $Th_T \subseteq Th_{T'}$ and $An_T \subseteq An_{T'}$. In this case we say simply that T' is an *extension* of T. This kind of extension is a form of reduction, viz. homogeneous reduction, i.e. the language of T is a part of the language of T'[25]. In this case it is said that T is *homogeneously* *reducible* to T'. One example is the reduction of Kepler's laws to classical mechanics.

A more interesting form of reduction is the one in which the language of T is not a part of the language of T' :

d) $f \not\subseteq f'$, $f \cup f'$ is a function and $M' \subseteq M$.

The way to reduce T to T' is to replace T' with its conservative extension $T'' = \langle L'', f'', M' \rangle$, where $L'' = L \cup L'$ and $f'' = f \cup f'$. Now every theorem of T is a theorem of T'' and of course every theorem of T' is a theorem of T'' : $Th_T \cup Th_{T'} \subseteq Th_{T''}$. In this case we say that T is *heterogeneously reducible* to T'. An example of this kind of reduction is the reduction of thermodynamics to classical statistical mechanics.

It seems to me that at least in the cases a)—d) the transition T/T' is a *cumulative change*. Let me turn now to *non-cumulative change*. One form of non-cumulative change is that where T and T' are in contradiction with each other. The necessary condition of contradiction is that

e) $M \cap M' = \emptyset$.

If $M \cap M' = \emptyset$, then T and T' are said to be jointly *semantically inconsistent* : both theories can not be true. But if T and T' have a common sublanguage L_o such that f and f' give the same meaning to every symbol in L_o, and there is a sentence φ in S_{L_o} such that

f) $\varphi \in Th_T$ and $\sim\varphi \in Th_{T'}$

then we shall say that T and T' are jointly *syntactically inconsistent*[26]. Note that the meaning of φ is the same in both theories. Thus $[\varphi]^F_f = W - [\sim\varphi]^F_{f'}$ and $M \cap M' = \emptyset$ because $M \subseteq [\varphi]^F_f$ and $M' \subseteq [\sim\varphi]^F_{f'}$. So if theories are syntactically inconsistent they are also semantically inconsistent. The condition f) does not in itself guarantee that theories are jointly inconsistent because in order for two theories to be jointly inconsistent the meaning of φ must be the same in both theories. Without this, the contradiction can be only *apparent*. If f) holds and $M \cap M \neq \emptyset$, then the meaning of some common term must be different in T and T' and the meaning change occurs in the transition T/T'.

In many applications theories T and T' have a common observational part in the sense that $f \cap f' = f_o$ is a valuation for the common observational language $L_o = L \cap L'$. If it is so, the case f) above means that theories T and T' are *observationally inconsistent*. Perhaps a more relevant case is the one where rival theories are not observational inconsistent but only *incompatible* in the following sense[27].

g) $f \cap f' \neq \emptyset$ and there are sentences φ and ψ in S_{L_O} where $L_O = D(f \cap f')$, such that $\varphi \rightarrow \psi \in Th_T$, $\varphi \rightarrow \sim\psi \in Th_{T'}$, and $\sim\varphi \notin Th_T \cup Th_{T'}$.

Now theories T and T' can be consistent but if the sentence φ turns out to be true, T and T' draw inconsistent conclusions from it and so sets $Th_T \cup (\varphi)$ and $Th_{T'} \cup (\varphi)$ are jointly syntactically inconsistent. The condition g) implies that $M \not\subset M'$ and $M' \not\subset M$.

Sometimes two theories are in complete agreement on observational facts in the sense that the set $X_O = Th_T \cap Th_{T'}$ is a complete set of observational sentences (if $\varphi \in S_{L_O}$ then $\varphi \in X_O$ or $\sim\varphi \in X_O$ where $L_O = D(f \cap f')$). Now it is easy to prove a weak version of *Robinson's consistency theorem*. It states that, if T and T' are consistent, then under the above suppositions T and T' are jointly syntactically consistent. For proof, let us assume that there is a sentence $\varphi \in S_{L_O}$ such that $\varphi \in Th_T$ and $\sim\varphi \in Th_{T'}$. Because X_O is complete, $\varphi \in X_O$ or $\sim\varphi \in X_O$. In both cases one of theories T and T' must be inconsistent, contradicting the assumption of the theorem. For example, if $\varphi \in X_O$, then $\varphi \in Th_{T'}$ and T' is inconsistent. Note that it cannot be proved that T and T' are also jointly semantically consistent. The reason for this is that W does not contain every L_O-model in the sense of model theory[28].

In the cases e)—g) we can speak about *scientific revolution*, at least. If in addition $f \cup f'$ is not a function, we can speak about scientific revolution *with meaning change*. Of course, there are many other interesting cases of cumulative or revolutionary change in scientific progress. For example there are various correspondence relations and approximative reductions[29]. I cannot analyze these cases here, unfortunately. But let me handle now the question of incommensurability.

5. *Incommensurability.*

The crucial point in incommensurability is the view that a sufficiently radical shift from a scientific theory to its successor involves a change in the meaning of the terms that are common to both theories. This view is called radical meaning variance view. I am not going to criticize the idea of incommensurability. Instead I try to explicate the notion of incommensurability and to study whether incommensurable theories are incomparable.

We shall say that theories T and T' are (semantically) *in-*

commensurable if the concepts of T and T' are different. More exactly, if T = $<L,f,M>$ and T' = $<L',f',M'>$ then the incommensurability of T and T' means that

$$R(f)\cap R(f') = \emptyset.$$

Incommensurable theories speak about different things. If in addition $L\cap L' \neq \emptyset$, then the meaning of every common term has changed in the transition T/T'. Note that if T and T' are incommensurable then $f\cap f' = \emptyset$, because $f\cap f' \neq \emptyset$ implies that $R(f)\cap R(f') \neq \emptyset$. So incommensurable theories cannot be syntactically commensurable. T and T' are said to be *syntactically commensurable*, if there is a common symbol in L and L', whose meaning is the same in both theories[30].

Does incommensurability exclude comparison of rival theories? No, because one may use the procedure presented in chapter 4 and compare theories by translation functions. Because $R(f)\cap R(f') = \emptyset$, the sets $Tr(S_L)$ and $Tr'(S_{L'})$ are disjoint and so there can be no *syntactical relations* between theorems or analytical sentences of T and T'. But there can be *semantical relations in the following sense. After translation* the sentences of T and T' are all sentences of the constructed language L'', and they can be compared as sentences of L''. Following Williams we can distinguish several cases[31] : let $\varphi^t = Tr(\varphi)$ and $\psi^t = Tr'(\psi)$, then

— φ and ψ are *intertranslatable* if $\varphi^t \leftrightarrow \psi^t \in An^F_{f''}$

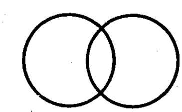
— φ and ψ are *inconsistent* if $\sim(\varphi^t \wedge \psi^t) \in An^F_{f''}$

— φ *entails* ψ if $\varphi^t \rightarrow \psi^t \in An^F_{f''}$.

This proves that there can be even *logical relations* between sentences (theorems) of incommensurable theories. For example, if φ is a theorem of T and ψ is a theorem of T' and if they are inconsistent in the above sense, then clearly T and T' are jointly semantically inconsistent. Similarly T' can be an extension of T or T can be reducible to T' and so on. But it is true that the terms of T cannot be translated into terms of T', if T and T' are incommensurable. In spite of this it can be possible to translate a term of T into a more *complex* expression of T'[32].

There is one case in which logical connections between incommensurable theories are absent. It is the case in which M and M' are overlapping :

$M\cap M' \neq \emptyset$, $M \not\subseteq M'$ and $M' \not\subseteq M$.

M  M'

If M and M' are overlapping, T and T' are not identical, they are not in contradiction, they are not reducible to each other and so on. In this case we can say that T and T' are *independent*[33]. But it seems to me that independent theories belong to different disciplines and that hardly any real transition T/T' in one discipline can be such that T and T' are independent.

6. *Theory-ladenness of observation*

Hanson and Kuhn among others have put forward the thesis that observation is theory-laden : adherents to different theories will observe different things when they view the same phenomena[34]. Let me present an analysis of phenomena. For this purpose I recall that W is a "state space" and that always one state is realized. This state is called an actual world and it is denoted by *a*. We call every subset E of W an *event* (a state of affairs). We say that an event E *occurs* if $a \in E$. An event E is called a *phenomenon relative* to a theory T = $\langle L,f,M \rangle$ if there is a sentence φ of the observation language $L_O \subseteq L$ of T such that $E = [\varphi]_f^F$. The idea is that if a phenomenon $E = [\varphi]_f^F$ relative to T occurs then the observation report φ is justified in principle by observation[35]. If T' = $\langle L',f',M' \rangle$ is another theory then events that are phenomena relative to T and events that are phenomena relative to T' are in general different. At least in this sense observation is theory-laden. But there are other and more interesting senses too.

In the above case of theory-ladenness of observation the observation reports do not depend on the "laws" of a theory. All what is needed is the framework of a theory. So the truth values of observation reports are independent of the truth value of a theory. But let me consider the following case :

— φ is an observation sentence, $\varphi \in S_{L_O}$
— it is consistent with T, $[\varphi]_f^F \cap M \neq O$
— ψ is a sentence of L, $\psi \in S_L$
— it is not a theorem of T, $\psi \notin Th_T$
— $\varphi \rightarrow \psi$ is a theorem of T, $\varphi \rightarrow \psi \in Th_T$
— it is not an analytical sentence, $\varphi \rightarrow \psi \notin An_T$.

If these assumptions hold then ψ follows from φ *assuming* T. I think that it is natural to call the set of all such ψ's the *interpretation* of φ *relative* to T. Now it is evident that the interpretations of an observation report φ may change relative to different theories. I find that the notion of interpretation is a quite good explication of theory-ladenness of observation[36].

A more relevant account is reached when it is discovered that there are two kinds of theoretical terms : those that refer to unobservable objects only, like 'electron', 'gene', and those that also refer to some observable objects, like 'magnetic' and 'intelligent'[37]. Let us say that terms of the second kind are *quasiobservational.* It seems to me that so called natural kind terms, like 'tiger', 'lemon' and 'water' are not observational but theoretical and very often quasiobservational terms[38]. Normally we are using just natural kind terms in our observation reports. The point is that application of quasiobservational terms in some situation depends on theories and observations. If the sentence ψ above contains quasiobservational terms and perhaps some observational terms, then if our observations show that φ is true and we give ψ as our observation report, then our report is true provided that our theory is true. Because ψ is a quasiobservational sentence it can be that there is no direct observational evidence for or against it. So ψ is strongly theory-laden.

Let me make the last point more precise as follows. It can be said that worlds w and w' are *f-identical*, $w =_f w'$, if

1. U(w) = U(w') and
2. f(s)(w) = f(s)(w') for all $s \in L = D(f)$.

That means that w and w' are "isomorphic" L-structures. Now if L_O is the set of observational terms of a theory T = <L,f,M> and if $f_O = f/L_O$ then one way to characterize *T-theoreticity* of T's theoretical terms is this :

(T) there are two f_O-identical worlds w and w' in W for all theoretical terms t in $L - L_O$ such that $f(t)(w) \neq f(t)(w')$.

It follows from (T) that the term t is undefinable from observational terms. This means that there is no formula φ of the language L_O such that

$\forall\bar{x}(\varphi(\bar{x}) \leftrightarrow t\bar{x}) \in An_T$ or $\forall\bar{x}y(\varphi(\bar{x},y) \leftrightarrow y = t\bar{x}) \in An_T$

where x is a sequence of variables and t is a predicate or function symbol, respectively. So it is impossible to know a priori merely on the ground of observation whether t applies or not. In this sense its application presupposes the validity of T. But the principle (T) does not exclude that for some observational formula φ it holds

$\forall\bar{x}(\varphi(\bar{x}) \rightarrow t\bar{x}) \in An_T$.

If it is so then we know a priori that t applies when observational condition φ is satisfied. To exclude also this possibility we state a stronger principle of *T-theoreticity*[39] :

(T') For all theoretical terms t and for all worlds $w \in W$ there is a world $w' \in W$ such that $w =_{f_O} w'$ and $f(t)(w) \neq f(t)(w')$.

The principle (T') implies that *only theory* can provide criteria of the application of theoretical terms[40].

7. Concluding remarks.

The analysis of theories, meaning change, observation and scientific progress presented above is tentative and its main purpose is to show that new insight into the topic of scientific change can be gained by using possible worlds semantics. I think that the most important novel feature in my analysis is that the *valuation function* (the framework of theories) assigns to terms *intensions*, i.e. concepts, not their extensions. With this feature I can express meaning change and what aspects of possible worlds are considered relevant by theories. This means that the role of a framework is twofold : 1. it gives a meaning to every term of a theory and 2. it selects relevant aspects of worlds. A framework represents the conventional part of a theory while the set of models of a theory represents its factual component.

There are many problems that my presentation leaves open. One crucial problem is the physical interpretation of theories. Without it theories are not scientific empirical theories[41]. This problem must be carefully distinguished from another open problem, namely the problem of theoretical-observational-distinction[42]. I have supposed that within a theory we can divide terms into observational terms and theoretical terms. I did not suppose that there is an observational language that is common to all theories. My two principles of T-theoreticity presuppose that the division of terms is given, not absolutely but relative to theories. But is the division of terms possible even in this relative sense ? There is no convincing answer to this question in literature. I believe that by developing further this approach, more light can be thrown on these problems too[43].

University of Helsinki

NOTES

[1] See Kuhn (1970) pp. 266–267.

[2] Cf. Suppe (1974) pp. 202—203.

[3] See Carnap (1958), (1963).

[4] See Carnap (1952).

[5] See Williams (1973).

[6] See Pearce & Rantala (1981).

[7] Outlines of possible worlds semantics can be found in Montague (1974), Introduction and chapters 3.-5.

[8] For most parts of our analysis it is enough that any set of sentences based on symbols in L and a satisfaction relation between sentences and possible worlds are given, cf. Feferman (1974).

[9] See Montague (1974) pp. 97–98.

[10] See Montague (1974) pp. 99—103. Note that we do not use intensional operators in this study, although they are very useful in the analysis of scientific concepts, cf. van Fraasen (1979).

[11] Cf. Williams (1973).

[12] See Suppe (1974) pp. 16–61 and 221—230, and Suppe (1979).

[13] See Suppe (1979) pp. 320 and 326. Note that Suppe's notion of semantic conception of theories is controversial.

[14] See Pearce & Rantala (1981).

[15] I shall apply the analysis of Pearce and Rantala by identifying N with a set of possible worlds, taking τ to be a set of predicate and function symbols and omitting R entirely. The set R can be incorporated into possible worlds semantics like Montague does, cf. Montague (1974) pp. 113—115, where Montague introduces a set of accessibility relations.

[16] By a suitable choice of the domain D we can present even within first order logic non-trivial scientific theories, cf Montague's article 'Deterministic Theories' in Montague (1974).

[17] Cf. Stegmüller (1979). In fact I follow more closely Pearce and Rantala (1981), for in their presentation M is a set of structures of a given type τ.

[18] Rantala & Pearce suppose that M is closed under isomorphism,

but I find this supposition too strong, for the domains of isomorphic models can be fully different. We suppose only that if a world w∈M and w' is identical with respect to f, with w (i.e. U(w)=U(w') and f(s)(w)=f(s)(w') for all s∈L) then w'∈M.

[19] See for example Margaris (1967) chapter 3.

[20] For this reason I do not accept Przelecki's and William's effort to give meaning to theoretical terms by meaning postulates, see Przelecki (1969) and Williams (1973).

[21] As far as I know there is no other attempt to define generally a concept of theory within possible worlds semantics. Cf. however Bressan (1974), where he presents a general modal language and some of its applications to physics. But Bressan does not handle problems of scientific change.

[22] Van Fraassen and Suppe construe theory structures as configurated state spaces, but their presentation is purely "semantical" without linguistic elements like terms or sentences, see van Fraassen (1980) and Suppe (1974).

[23] Agazzi proposes the replacement of the extensional viewpoint with a more effective intensional one, but he advocates the introduction of meanings for 'basic predicates' by means of operational definitions, see Agazzi (1976).

[24] Cf. Przelecki (1979) p. 352.

[25] For different kinds of reduction see Krajewski (1977).

[26] Cf. Williams (1973a) p. 362.

[27] Cf. Przelecki (1977) pp. 359—360.

[28] Cf. the formulation and the proof of Robinson's consistency theorem in Kleene (1967) pp. 368—369.

[29] For these see Krajewski (1977) chapter 4.

[30] One can feel that this definition of commensurability is too weak, because it says nothing about common domain or common references of terms. But I find the talk about common references problematic, for we do not know what the actual world is and so we can not know what the actual references of terms are. The subject matter of theories can be identified only intensionally : for example, two theories are both dealing with microparticles and their subject matter is partially the same if both theories have the *same concept* of microparticles. It is impossible to know the "real" reference

class of this concept. But note that although both theories have the same concept of microparticles the "laws" of theories can be different, i.e. they "believe" different things about microparticles.

[31] Williams (1973a) pp. 360—361.

[32] Cf. the critique of meaning variance in Levin (1979).

[33] I mean that T and T' are independent if they are incommensurable and M and M' are overlapping. Such theories speak about different things and they can be both true or both false or one can be true and the other false. The heliocentric theory and quantum mechanics are independent in this sense.

[34] See Suppe (1974) pp. 191—199.

[35] We can also say that if an event is a phenomenon relative to T, it is in principle *observable.* What a scientist observes is normally only a part of all observable occuring phenomena and so observation is a matter of selection. This selection does not only depend on theory but also on practical interests.

[36] Normally scientists are interested in the interpretations of observations and they are disposed to "see" and report their observations in terms of the elements of interpretation, i.e. their report is ψ not φ. Cf. Krajewski (1977) pp. 60—63.

[37] See Przelecki (1969) pp. 48 and 85.

[38] See Putnam (1975) chapter 12 and especially p. 225.

[39] This definition is an analogy of Sneed's notion of T-theoreticity. Sneed's idea is that exactly those functions whose values cannot be calculated without recourse to a theory T are theoretical relative to T. See Stegmüller (1976) pp. 40—46.

[40] This does not mean that there are no observational criteria of application. The point is that observations do not alone without recourse to the validity of a theory suffice to the application of theoretical terms. More exactly, although the formula $\forall \bar{x}(\varphi(\bar{x}) \rightarrow t\bar{x})$ is not analytical it can surely be a *theorem* : according to the theory φ is an observational criterion of t's application. Cf. Padoa's principle in Kleene (1967) pp. 362—364.

[41] Cf. Suppe (1979) pp. 324—325.

[42] For this distinction see Causey (1979).

[43] I wish to thank Veikko Rantala for comments on an earlier version of this paper.

BIBLIOGRAPHY

AGAZZI, E., "Logic and Methodology of Empirical Sciences", in E. Agazzi (ed.), *Modern Logic — A Survey, Historical, Philosophical, and Mathematical Aspects of Modern Logic and its Applications*, Reidel, Dordrecht : Holland, 1981, pp. 255—282.

BRESSAN, A., "On the usefulness of modal logic in axiomatizations of physics", in K. F. Schaffner & R. S. Cohen (eds.), *PSA 1972*, Reidel, Dordrecht : Holland, 1974, pp. 285—303.

CARNAP, R., "Meaning postulates", *Philosophical Studies* 3, (1952), pp. 65—73.

CARNAP, R., "Beobachtungssprache und theoretische Sprache", *Dialectica* 12, 1958, pp. 236—248.

CARNAP, R., "Replies and Systematic Expositions", in P.A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*, Open Court, La Salle, 1963.

CAUSEY, R. L., "Theory and Observation", in P.D. Asquith & H.E. Kyburg (eds.), *Current Research in Philosophy of Science*, PSA, Michigan 1979, pp. 187—206.

FEFERMAN, S., "Two Notes on Abstract Model Theory", I, *Fund. Math.* 82, 1974, pp. 153—165.

KLEENE, S.C., *Mathematical Logic*, John Wiley & Sons, New York, 1967.

KRAJEWSKI, W., *Correspondence Principle and Growth of Science*, Reidel, Dordrecht: Holland, 1977.

KUHN, T.S., "Reflections on my Critics", in I. Lakatos & A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge Univ. Press, Cambridge, 1970, pp. 231—278.

LEVIN, M.E., "On theory-change and meaning-change", *Philosophy of Science* 46, 1979, pp. 407—424.

MARGARIS, A., *First Order Mathematical Logic*, Ginn-Blaisdell, Waltham, Massachusetts, Toronto, London, 1967.

MONTAGUE, R., *Formal Philosophy*, Edited and with an introduction by Richmond H. Thomason, Yale Univ. Press, New Haven and London, 1974.

PEARCE, D. and V. RANTALA, "On a New Approach to Metascience", *Reports from the Departement of Philosophy*, University of Helsinki, No 1, 1981.

PRZELECKI, M., *The Logic of Empirical Theories*, Routledge & Kegan Paul, London, 1969.

PRZELECKI, M., "Commensurable Referents of Incommensurable Terms", in I. Niiniluoto & R. Tuomela (eds.), *The Logic and Epistemology of Scientific Change*, North-Holland, Amsterdam, 1979, pp. 347—365.

PUTNAM, H., *Mind, Language and Reality*, Cambridge Univ. Press, Cambridge, 1975.

STEGMUELLER, W., *The Structure and Dynamics of Theories*, Springer-Verlag, New York, 1976.

STEGMUELLER, W., *The Structuralist View of Theories*, Springer-Verlag, Berlin, 1979.

SUPPE, Fr. (ed.), *The Structure of Scientific Theories*, Univ. of Illinois Press, Urbana 1974.

SUPPE, Fr., "Theory Structure", in P.S. Asquith & H.E. Kyburg (eds,), *Current Research in Philosophy of Science*, PSA, Michigan, 1979, pp. 317—338.

VAN FRAASSEN, B.C., "Modality", in P.D. Asquith & H.E. Kyburg, (eds.), *Current Research in Philosophy of Science*, PSA, Michigan, 1979, pp. 282—290.

VAN FRAASSEN, B.C., *The Scientific Image*, Clarendon Press, Oxford, 1980.

WILLIAMS, P.M. (1973), "On the Conservative Extensions of Semantical Systems : a Contribution to the Problem of Analyticity", *Synthese* 25, 1973, pp. 398—416.

WILLIAMS, P.M. (1973a), "On the logical relations between expressions of different theories", *BJPS*, 24 (1973), pp. 357—408.