A translation of: Balzer, W. 1988 *Der Nutzen wissenschaftstheoretischer Analyse: dargestellt an der Frage der Gültigkeit und aus strukturalistischer Sicht*, in: *Why the theory of the sciences?* (eds. P. Hoyningen-Huene und G. Hirsch), Walter de Gruyter, Berlin - New York, pp. 53 -74.

The Utility of science-theoretic Analysis: Illustrated by the Question of Validity and from a structuralist Perspective

by
Wolfgang Balzer

I do not wish to comment here on the *meaning* of the theory of the sciences; other authors in this volume are better equipped to do so. Rather, I would like to limit myself to utility – though I must mention before the caveat that investigations into the utility of certain activities actually fall within the realm of psychology and economics, and I can therefore, as an investigator of the theory of the sciences, offer only a few amateurish remarks here. Distinguishing between 'purpose' and 'utility' is difficult. One must distinguish between the utility of an action or a thing *for an individual* and *for a group* (such as 'society'); here, we are concerned only with the latter meaning. Utility exists when there is a positive influence on material conditions – on everything that makes life easier and safer – and when this influence is not outweighed by parallel negative influences. An action or thing, on the other hand, can also be meaningful even if it concerns exclusively the spiritual entities (such religious or ideological entities). Meaning often depends decisively on the attitude of the person evaluating it, and for different observers, 'the same' thing or action can be meaningful in very different ways. Something similar can be said about utility (keywords: security and nuclear armament), but there is nevertheless a difference in the frequency of cases in which conflicting opinions about utility and harm arise compared to those involving meaning and meaninglessness.

If a theoretical physicist is asked to provide information about the utility of his work, he can point to technical applications of physical knowledge that are generally regarded as useful. This is already more difficult for the economist, whose 'theory of artificial production' is unsuitable because it does not provide precise predictions. He can, however, at least point to some past missteps (such as the Great Depression of the 1920th), which are now understood theoretically to such an extent that they can be avoided under similar conditions. It is even more difficult for the sociologist to demonstrate his usefulness. He might point out, for example, that the game-theoretic analysis of the prisoner's dilemma will influence arms decisions, *provided* that politicians take this analysis into account. This range of examples is intended to help properly classify the theory of the sciences, i.e., roughly *after* sociology – more generally, *after* the social sciences and *before* the humanities. The investigator of the theory of the sciences can point to neither technical applications nor definitive missteps in the past that could have been avoided with his theory, nor can he cite any cases in which his theory would alter the course of events, provided only that the parties involved would took his theory into account.

But what, then, is the use of the theory of the sciences? Or does it have none at all? In what follows, I will attempt to compile some evidence *that* the theory of the sciences is useful. I will begin with general considerations and gradually move on to more specific ones.

The engineer working on the design of a new chip contributes directly to the utility that the chip brings in its application; for the physicist who studied semiconductors, this applies only indirectly, in the sense that his work is reflected in a useful product after further intermediate steps (such as those taken by the engineer). The more theoretical the work, the less direct the connection to useful products or activities, will be. And if demonstrating utility can be difficult even within the natural sciences, this is all the more true for 'softer' disciplines and, in particular, for the theory of the sciences.

Here, a side glance at economics is instructive, where, when determining the utility of a complex commodity, one does indeed take into account the sum of its separately manufactured parts, but not the utility of scientific findings applied in production (at least as long as no patents are involved) nor the utility of infrastructure, such as the school education of workers, their ability to resolve personal conflicts without affecting production, etc. From a more global theoretical perspective – and also when applied to developing countries – this limited economic valuation proves to be inadequate. Nevertheless, economics does not include the aforementioned, admittedly important components in its analysis, because doing so would make it practically unfeasible. The same applies to the cost-benefit analysis of scientific activity. Although it is recognized that theoretical work also contributes to useful commodities and activities, one is often unable to specify such a contribution precisely, simply due to its complexity.

This is precisely the case with the theory of the sciences. Work in the theory of the sciences does not directly translate into useful products. To a certain – rather limited – degree, it influences other activities that are generally considered useful, such as didactics and basic research.

As I understand it, the main concern of the theory of the sciences is to maintain an overview of the entire enterprise of 'science' – or, rather, not to lose sight of it. The theory of the sciences attempts to do this by systematically synthesizing the joint beliefs that appear across the various branches of science into a comprehensive theoretical picture – that is, into a *theory*.[1]

[1] Such a *theory* of the sciences – a theory about science – seems to be finding favorable conditions at present, especially in the German-speaking world.

Now, science is primarily a social endeavor; there are sociological, psychological, and social-psychological aspects, all of which are addressed in the relevant disciplines. A theory that takes all these aspects into account is not currently foreseeable. Some believe that such a theory cannot even exist, precisely because the phenomenon of 'science' is so complex and multifaceted. I do not share this view – first, because using the same reasoning would place all social sciences, and certainly all humanities, in the category of disciplines incapable of theory; second, because we need nothing more urgently than a theoretical understanding of social phenomena (since otherwise we will bring about our own extinction in the near future); and third, because there are already fine theories, for example, in the social sciences, and second, because there is already a stable, 'objective' database for the theory of the sciences, namely in the form of printed books and similar enduring manifestations of scientific activity. It seems to me that this database alone is rich and complex enough to justify a theoretical systematization. The result is a theory about the documented output of scientific activities, and I call this theory the theory of sciences. This theory of science admittedly abstracts from the social and psychological phenomena in science. But abstraction, as is known from all sciences, is always justified when it leads to 'interesting' systematizations (which is certainly the case here). Of course, one can still object that the choice of the term gives the false impression that one is dealing with the overall phenomenon of science in all its aspects. Anyone who considers this objection important is welcome to come up with a different name.

The influence that such a theory of the sciences has on other, more directly useful activities is, as mentioned, only gradual and of the general kind in which a theoretical conception, a hypothetical picture, guides our actions, even though one cannot make accurate predictions with the help of that picture. But since one *must* act, one relies on *some kind of* conception – and preferably the best one available. The theory of the sciences offers such theoretical pictures of the general structure and method of empirical sciences, and since such pictures influence other useful activities, the theory of the sciences is also useful. Actions guided by a theoretical picture will be all the more successful the 'better', more precise, and more adequate the picture in question is. The usefulness of the theory of the sciences will therefore be all the greater, the better and more precise the theoretical picture of science is.[2]

Here, however, the objection arises that in the theory of the sciences there are, in fact, several different and *conflicting* conceptions of the structure and method of science, and that it is therefore difficult to see why orienting oneself toward *some of them* of these could be useful, when orienting oneself toward another, opposing conception would presumably lead to entirely different actions. The following can be said in response: First, the various approaches and schools are *not* as contradictory as the polemical debates between them might suggest. Rather – and I would like to assert this without providing a detailed justification here – they share a broad common

[2] The concept of 'better and more accurate' discussed here is precisely the subject of the theory of the sciences.

basis. To cite just one example: The contrast between the 'statement view' and the 'non-statement view' concerns less the content and more the tools and strategy of the epistemological approach.[3] Second, orienting oneself toward *some kind of* epistemological approach will still yield more useful results than a completely theory-unguided approach. Third, one must regard the current diversity as fruitful for the theory of the sciences itself: The search for the 'correct' metatheory is in full swing, and many competing approaches are – in accordance with Feyerabend – entirely desirable.

These general preliminary remarks can be summarized as follows: the theory of the sciences of any orientation is useful, albeit in a very indirect way, and the greater the benefit of the theory of science, the better and more precise the theoretical picture it provides of the structure and method of science.

This brings me to the second, more theory-of-science-oriented part, in which I would like to highlight the merits of a particular approach – namely, the structuralist one – by examining *one specific aspect* (among a multitude of other possible ones), namely the question of a theory's validity. The claim specifically associated with this is that the picture painted by structuralist theory of the sciences is more precise and better than that provided by other approaches. It follows from what has been said above that the structuralist approach also has an advantage in terms of usefulness. However, the assessment of this claim must be left to the reader's judgment, because I must limit myself here to a (already very condensed) presentation of the question of validity from a structuralist perspective and cannot address alternative approaches.[4]

First, let us briefly review some basic concepts of the structuralist approach.[5] An *empirical theory* T therefore consists of a *formal core* K and a set I of *intended applications* that 'emerge' from real systems. While the core K comprises the concepts of the theory and their basic assumptions, the set I represents reality, the world, or the data relevant to T. The core consists of three parts: a class M_p of *potential models*, a class M of *models*, and the *constraints* Q. The potential models roughly represent the basic concepts, while the models and constraints represent the axioms or hypotheses of the theory.

More precisely, the class M_p is given by a *structure type* τ, *base sets* $D_1, ..., D_k$, *auxiliary base sets* $A_1, ..., A_m$, and *relations* and *functions* (of different numbers of arguments and levels) $R_1, ..., R_n$. The structure type $\tau = \langle k, m, \sigma_1, ..., \sigma_n \rangle$ consists of numbers k ($k > 0$) for base sets and m (≥ 0) for auxiliary base sets and *types* $\sigma_1, ..., \sigma_n$ ($n > 0$) concerning the *arities* and *levels* of the relations.[6] Potential models

[3]See the relevant sections of (Stegmüller, 1973) for a detailed discussion.

[4]To be precise, the approach taken in (Ludwig, 1978) – which is also advocated by Scheibe – must be excluded from my assertion. The view developed below is, in fact, strongly influenced by Ludwig's work.

[5]A comprehensive, up-to-date overview can be found in (Balzer, Moulines, Sneed, 1987).

[6]Types are defined inductively here. 1) Every number $j \leq k + m$ is a type; 2) if σ and σ' are types, then so are $po(\sigma)$ and $(\sigma \oplus \sigma')$. If $u_1, ..., u_{k+m}$ are sets and σ is a type, then the echelon set $\sigma(u_1, ..., u_{k+m})$ is defined inductively: 1) if $\sigma = j$, then $\sigma(u_1, ..., u_{k+m}) = u_j$; 2) if $\sigma = po(\sigma')$, then $\sigma(u_1, ..., u_{k+m}) = po(\sigma'(u_1, ..., u_{k+m}))$, 3) if $\sigma = (\sigma_1 \oplus \sigma_2)$, then $\sigma(u_1, ..., u_{k+m}) = \sigma_1(u_1, ..., u_{k+m}) \times \sigma_2(u_1, ..., u_{k+m})$. A relation R is said to be of type σ over $u_1, ..., u_{k+m}$ if R is an

of a theory now are *structures of structure type* τ, i.e., set-theoretic entities of the form

$$\langle D_1, ..., D_k; A_1, ..., A_m; R_1, ..., R_n \rangle,$$

where $D_1, ..., D_k$ are base sets, $A_1, ..., A_m$ are auxiliary base sets, and R_1, ..., R_n are relations (or functions) of types $\sigma_1, ..., \sigma_n$, respectively, over the sets $D_1, ..., D_k$ and $A_1, ..., A_m$. For M and Q, it is assumed that $M \subseteq M_p$, and $Q \subset \wp(M_p)$; for further assumptions, see (Balzer, Moulines, Sneed, 1987), Chap. II.

Substructures are obtained from the potential models by omitting arbitrary parts such that the result still has the same structure type . If $x = \langle D_1, ..., D_k; A_1, ..., A_m; R_1, ..., R_n \rangle$ is a structure type τ and y has the form $\langle D_1, ..., D_k; A_1, ..., A_m; R_1, ..., R_n \rangle$, then y is called a *substructure of* x (and x a *supplement* of y); in notation: $y \sqsubseteq x$ iff, for all $i \leq k, j \leq m$, and $r \leq n$, $D'_i \subseteq D_i$, $A'_j \subseteq A_j$, and $R'_r \subseteq R_r$, and if y has the structure of type τ. The class of all substructures of potential models is denoted by M_{pp}, where $M_{pp} = \{z / \exists z \in M_p (z \sqsubseteq x)\}$. Substructures can be finite even if the original potential models from which they were derived were infinite. Substructures are well-suited for representing data collected, read, or measured from a real system. This system is represented by an intended application which is, after all, should represent 'the world' – a substructure of a potential models ($I \subseteq M_{pp}$). The assumption that intended applications are substructures of potential models means, first, that one considers intended applications within the terminology of the theory – that is, one ignores irrelevant concepts, even if they were to be realized – and, second, that one has observed and measured certain data on the real systems relevant to the theory, which are 'put together' when describing the corresponding substructure. The assumption *does not* in itself imply the validity of theoretical relationship.

Let us briefly illustrate these components of a theory using the example of wave mechanics, as formulated by Schrödinger in his papers[7] On the one hand, this is intended to provide another positive example of the structuralist concept of theory;[8] on the other hand, it is meant to highlight the relative simplicity of the notation introduced, which becomes evident in this rather complex case.

Some mathematical explanations are necessary to define the models. The *Hilbert-space* $L^2(\mathbb{R}^{3n}, \mathbb{C})$ consists of all functions ϕ: $\mathbb{R}^{3n} \to \mathbb{C}$ modulo equality up to the Lebesgue measure λ^{3n}) for which ϕ^2 is Lebesgue integrable. These functions can be added pointwise and multiplied by complex factors, thereby defining a vector space structure. $\phi^*(a)$ denotes the complex conjugate of $\phi(a)$, and ϕ^* denotes the corresponding function. By defining $\langle \phi, \psi \rangle := \int \phi \psi^* d\lambda^{3n}$, a scalar product (and thus also a notion of distance) is defined in $L^2(\mathbb{R}^{3n}, \mathbb{C})$. With respect to the topology thus given, the space is complete, i.e., every Cauchy sequence of functions converges to a

element of $\sigma(u_1, ..., u_{k+m})$. In the limit case $\sigma = j$, individuals are also referred to as relations here for the sake of brevity.

[7](Schrödinger, 1926). The presentation follows (Zoubek, 1987).

[8]Many other examples have been reconstructed in a similar manner; see, for example, (Balzer, Moulines, Sneed, 1987), Chap. II, as well as the literature cited there.

limit in the set. A *linear operator* W in this Hilbert space is a partial, linear function. W is called *symmetric* if, for all $\langle\phi, \psi\rangle \in Dom(W)$, we have $\langle W\phi, \psi\rangle = \langle\phi, W\psi\rangle$. For further details, see (Achieser & Glasmann, 1968). For ψ: $\mathbb{R}^{3n} \times T \to \mathbb{C}$ and $t \in T$, let ψ_t be the function defined by $\psi_t(a) = \psi(a, t)$. Similarly, for $W : L^2(\mathbb{R}^{3n}, \mathbb{C}) \to L^2(\mathbb{R}^{3n}, \mathbb{C})$, W_t or $W(t)$ is the function for which $W_t(\psi) = W(\psi, t)$.

x is a *potential model of Schrödinger's wave mechanics* ($x \in M$) if and only if there exist $n, P, T, k, p, m, K, w, W$ such that
$x = \langle P; T, \mathbb{R}, \mathbb{C}, \mathbb{N}_n; k, p, m, K, \psi, W\rangle$ and the following hold:
1) P is a non-empty set with exactly n elements
2) $T \subseteq \mathbb{R}$ is an interval
3) k: $\mathbb{N}_n \to P$ is bijective
4) p: $\mathbb{R}^{3n} \times T \to \mathbb{R}$
5) $m : P \to \mathbb{R}^+$
6) $K \in \mathbb{R}^+$
7) ψ: $\mathbb{R}^{3n} \times T \to \mathbb{C}$ satisfies appropriate continuity and differentiability requirements, and for all $t \in T$, $\psi_t \in L^2(\mathbb{R}^{3n}, \mathbb{C})$ and $\int |\psi_t(q)|^2 dq \neq 0$
8) $W = L^2(\mathbb{R}^{3n}, \mathbb{C}) \times \to L^2(\mathbb{R}^{3n}, \mathbb{C})$, and for all $t \in T$, W_t is a linear, symmetric operator.

Interpretation: P is a set of 'micro' particles, T is a time interval. k is a 'counter' for particles. p is a 'density function'. In Schrödinger's first approach, p is regarded as the density of the charge of the particles spread out over space, i.e., integration of p over a small region of space yields the 'proportion' of a particle's charge present in that region. As is well known, this interpretation did not prevail and was replaced by the so-called statistical interpretation, according to which p is a density (in the sense of probability theory) for the probability of the particles being present. m is the mass function, K is Planck's constant (divided by 2π). At every time $t \in T$, ψ_t represents the spatial distribution of the particles. To understand this, one must keep in mind that in *wave* mechanics, the 'particles' are represented by waves – that is, by spatially extended entities. The spatial configuration of these waves is given by ψ_t (at time t). Today, one puts it more cautiously: ψ_t is a summary of all the information that can be obtained about a microsystem through measurement at time t. Finally, the operator W_t corresponds roughly to the concept of force in mechanics: W is the theoretical term used to describe and explain the change of the system over time (represented by the various ψ_t, $t \in T$). Ultimately, these explanations must remain unsatisfactory, because wave mechanics was superseded by the 'classical' formulation of quantum mechanics precisely due to difficulties in interpretation. Note, however, that the mathematical formalism of the classical formulation does not differ dramatically from that of wave mechanics.

x is a *model of Schrödinger's wave mechanics* ($x \in M$) iff there exist $n, P, T, k, p, m, K, \psi, W$ such that
1) $x = \langle P; T, \mathbb{R}, \mathbb{C}, \mathbb{N}_n; k, p, m, K, \psi, W\rangle \in M_p$ and $K \in \mathbb{R}^+$

2) $p = |\psi|^2$
3) for all $t \in T$, ψ_t lies within the domain of definition of W_t
4) for all $t \in T$, the following holds

$$\sqrt{-1}D_{3n+1}\psi)_t = -\sum_{j\leq n}\frac{K^2}{2m(k(j))}(D^2_{3j-2}, D^2_{3j-1}, D^2_{3j})\psi_t + W_t\psi_t.$$

Condition 2) links the 'theoretical' wave function to the 'observable' values of p. 3) ensures that the expression $W_t\psi_t$ in 4) makes sense. Finally, 4) is this the Schrödinger equation, in which the wave function is subject to a dynamic law. For a given operator W, 4) represents a system of differential equations for ψ. Here, $D^2_j\psi_t$ denotes, for $j \leq 3n$, the second partial derivative of ψ in the direction j, and $D_{3n+1}\psi$ denotes the derivative of W with respect to the last argument (' time'). Let us denote the Laplace operator $D^2_{3j_2} + D^2_{3j-1} + D^2_{3j}$ as Δ_i, $m(k(j))$ as m_j, $D_{3n+1}\psi$ as $\delta/\delta_t\psi$, expand the brackets on the right-hand side of 4) to ψ_t, and finally simply omit the index t, then 4) becomes an expression of the kind found in physical formulations (with $i = \sqrt{-1}$): $\frac{\delta}{\delta_t}\psi = -[\sum_{j\leq n}\frac{K^2}{m_j}\Delta_j + W]\psi$.

The expression in square brackets is the 'wave operator', which consists of a 'free' part, $\sum\frac{K^2}{m_j}\Delta_j$ and the 'interaction part' W, just as in mechanics energy consists of kinetic and potential energy.

Wave mechanics exhibits a number of constraints, of which only two will be mentioned here. A set of potential models X satisfies the *identity constraint for* m gdw $X \neq \emptyset$, and for all $x, y \in X$ and all $p \in P_x \cap P_y$, the following holds: $m_x(p) = m_y(p)$. Here, P_x and m_x denote the set of particles and the mass function occurring in $x = \langle P_x, ..., m_x, ...\rangle$ (and analogously for y). X satisfies the *identity constraint for* K gdw $X \neq \emptyset$, and for all $x, y \in X$: $K_x = K_y$.

Intended applications of this theory include, for example, experiments in which atoms are bombarded with α particles and deflect them. The data obtained in such an experiment (the cross-section spectrum as well as the masses and charges assumed to be known) yield a substructure of a potential model (at least if one introduces the charges via existence quantification).

I now turn to the specific topic I announced regarding the validity of theories. It seems that Popper's insistence on the methodological impossibility of confirming a theory, Thomas Kuhn's plausible argument for the historical nonexistence of decisive, theory-falsifying experiments, as well as a general unease regarding 'the' truth of theories have led to the question of the correctness or validity of a theory within its *domain of intended applications* currently being regarded by many as uninteresting or even as a question that misses the mark of reality. In contrast, a look at the history of science teaches us that the actors involved were always extremely interested in formulating 'correct' theories[9] – a finding I shall take as my starting point for the discussion that follows. The dispute then revolves around what 'correct' means. The

[9]This, in turn, led the realists among philosophers (such as Putnam, following Boyd) to assert that all theories are 'approximately' true somewhere.

simplest solution, consistent with Popper and Kuhn, is to designate a theory as *false* if it has been superseded by a more general, better, or more successful theory, and as correct if it is not false. This criterion is *not purely* historical, because one can impose strict conditions on the concept of 'is more general than', such that the 'old' theory is contained, either partially or approximately, within the 'new' one. Without wishing to criticize this approach, I would like to turn here to a second, more difficult solution, which seems desirable at least as a supplement to the first. This involves a 'direct' definition of what it means for a theory to be valid within its scope of application – 'direct' in the sense of 'without recourse to historical predecessor and successor theories and the associated concept of generalization'.

In a first attempt, one might define theory T as approximately valid within its *domain of intended applications* I if I can be *approximately embedded into models* of T, and the constraints of T are fulfilled.

However, this definition of validity is unsatisfactory. First, it can be objected that I generally includes not only the systems actually studied, but also those that are sufficiently similar to them. Thus, I represents not only the 'available data', but also certain 'hypothetical data' that could be collected from certain real systems if one so desired. However, the intended definition of validity only becomes feasible by eliminating such hypothetical elements. This is easily accomplished. We replace I with the subset I_r of the systems actually studied and define: 'T is approximately valid in L'. This can be specified as follows: We introduce a *uniform structure* U on the class M_{pp} of substructures,[10] $U \subseteq \wp(M_{pp} \times M_{pp})$, whose elements $u \in U$ are abstract degrees of similarity. We define in general, for a class X of potential models ($X \subseteq M_p$): I can be *completed to* X *in degree* $u \in U$ if and only if there exists a class Y of substructures (i.e. $Y \subseteq M_{pp}$) such that the following hold:

1) Y and I are similar in degree u
2) Y can be completed to X.

More precisely, in condition 1), this means that there exists a bijective mapping f from Y to I_r such that all pairs ⟨argument, function value⟩ are similar of degree u ($\langle x, f(x)\rangle \in u$), and 2) means that for every $y \in Y$, there exists a complement $x \in X$ such that $y \sqsubseteq x$. Finally, for a given theory T, a given set $I_r \subseteq I$, and a given uniformity U on M_{pp}, we define: *T is valid in I of degree u*, if and only if there exists a class $X \subseteq M_p$ such that:

1) I_r can be complemented to X in degree u
2) $X \subseteq M$ and $X \in Q$.

That is, the systems actually under investigation are, in degree u, similar to certain set-theoretically existing systems that form a set Y, and the set of these 'hypothetical' systems can be complemented to form a set X of models for which the constraints also hold.

A third reason why this definition of validity is also unsatisfactory lies in the fact

[10] Cf., e.g., (Schubert, 1964), Chapter II.

that nothing further was said or assumed about the nature of the intended applications I_r under investigation. How are the elements of I_r given? The basic idea here, as already mentioned, is that an intended application represents, in a comprehensive manner, the data observed or measured on a real system. Let us consider how data are collected from a real system. In a few cases, the data are collected through sensory perception. In the vast majority of cases, however – and especially in quantitative theories – the data are obtained through measurement. According to the standard literature on measurement[11] measuring the quantity of an object a consists of combining units of the quantity in question to produce an object b that has the same quantity as a – that is, it has approximately the same size or weight as a. The desired quantity of object a is then simply given by counting the required, concatenated units. If measurement in science always would proceed in this way, the process of data collection carried out from intended applications would not pose any problems for the definition of validity: One collects data through direct observation or measurement (in the sense of 'comparison with units'), one aggregates this data into substructures, and one then checks at the desk whether, for a given degree u, the intended applications formed in this way can be supplemented into models of degree u. If so, the theory is valid in I_r to degree u.

However, this scenario is too fine to be true. Normally – that is, in the vast majority of cases – data are not collected in either of the two ways mentioned. The 'normal' procedure is rather as follows. One considers the real system, as given by the measurement process, as a model of one or more relevant, applicable theories, and one uses the hypotheses characteristic of these theories to calculate the value to be measured from other, already known data. In the wave mechanics example, one can determine or measure the mass of a particle, for instance, by determining the charge-to-mass ratio through a scattering experiment and, using the assumed known value for the charge, inferring the mass from this.

Of course, one could decide to use the word 'measurement' only for the procedure described first – that of comparison with units – and to introduce a different word for the procedure described last, in which theoretical assumptions are used. However, this would imply that in science, with the exception of a few branches of experimental psychology, hardly any measurements are would made. It therefore seems more appropriate to me to use the term 'measurement' for the latter, theoretical type of data determination as well. I would also like to emphasize that what has been said above does not imply that it is *in principle* impossible to trace theory-guided measurement procedures back to a comparison with units. I am merely noting that such a traceback hardly ever occurs in scientific practice.

If we use the word 'measurement' in the broader sense described above – that is, as direct observation, as a comparison with units, or as a calculation based on other values using given theories – then the following situation may arise. When determining the data for an intended application x of theory T, it *may* be the case that theory

[11] See, for example, (Krantz et al., 1971).

T itself is used as a hypothesis for the calculation of this data. When collecting data for T, T is thus already being used and, in a certain sense, is already assumed to be valid. Our previous definition of validity therefore allows for a pragmatic circularity: To test the validity of a theory, one must determine the data from which the intended applications consist. However, during this determination, the theory may already be used and thus assumed to be valid. In short: To test the validity of the theory in such cases, the validity of the theory is already assumed.[12]

I would like to refer to the possibility of such pragmatic circularity as the 'test problem'. If such circularity occurs during testing, to what extent can one even speak of a test at all?

One could dismiss the test problem as a pseudo-problem by pointing out that theory-independent data does not exist at all. From philosophy and everyday experience, we know that data is *always theory-laden*. We always 'see' things in the light of certain preconceptions, biases, or theories. We often fail to perceive 'facts' that we do not theoretically expect at all. Far be it from me to dispute all this. But I believe that pointing out such theory-laden nature of data is not sufficient to completely destroy the idea of theory-independent data collections. Such a reference carries the same weight as the claim by some equilibrium economists that their theories, in principle, cover *all* economic phenomena, or as the assertion that all mathematics is ultimately set theory (or category theory). Such claims may be correct in principle, but they are practically irrelevant.

Against the general claim of theory-laden data, we offer the observation that in concrete test situations, a fairly clear distinction can be made between data that 'essentially' depend on a given theory and data that do not (although this does not deny a general philosophical context of the kind in question). Data depend essentially on a theory if the theory is directly used in the measurement or collection of these data. The test problem therefore has two versions. The first, trivial version asks about the possibility of circularity-free tests in the principled sense that the tested theory exerts *no influence whatsoever* on the acquisition of the data. According to the prevailing philosophical view, such tests do not exist. The second, non-trivial version of the test problem, which I address here, consists of the question of whether the test of a theory can be essentially circular, or, alternatively, whether *essentially* non-circular tests can exist at all. Essentially circular means that the data used in the test data used in the test depend essentially on the theory being tested (in the sense introduced earlier, namely that the theory is used in the collection of the data).

This second test problem initially has a superficial solution: When collecting data for a theory, one must simply take care not to use the same theory in the process. As simple and plausible as this methodological rule sounds, it runs the risk of being violated for several reasons. First, when collecting data, it is generally not made explicit

[12]Although this circularity is of the same kind as in (Sneed, 1971), it should be noted that here, and also below, there is no mention of the 'problem of theoretical terms', as formulated by Sneed and also presented in (Stegmüller, 1973).

for which theory this data is intended to serve as evidence. Often this can be inferred from the context of the measurement, or – and this is a common case – one is not yet clear about the theory and is initially simply interested in obtaining reliable data – however that may be achieved. Science does not simply follow Popper's scheme: formulate a hypothesis, test it, reject it, or provisionally accept it. The collection of data often goes hand in hand with the formulation of the hypothesis or even precedes it. In any case, a look at the experimental branches of physics and economics shows that data collection is indeed an independent goal of scientific activity and by no means always serves merely as an auxiliary function for testing given theories.[13] Therefore, as mentioned, the collected data is generally not explicitly labeled as relevant to a specific theory. But if this designation is missing, then it can happen – 'by accident'? – that an invalid valuy sneaks into the data for T, i.e., into the systems of I_r, a value that was measured using T itself.

The probability of such invalid data increases further when one considers that in more mature disciplines, such as physics, measurements can take on considerable complexity and, upon closer analysis, can by analyzed by *measurement chains*, i.e., sequences of interdependent measurement models, each of which captures a measurement process, with up to 10 or more intermediate steps. The longer such measurement chains become, the more likely it is that in one of the intermediate chains – one of the measurement models in the chain – the theory for which the value ultimately measured at the final link of the chain is intended to serve as data is used and assumed.[14]

It is therefore entirely possible – and to a certain extent also probable – that, when testing the validity of a theory, significant pragmatic circles of the kind described will arise. And it is clear at which point in the metatheoretical picture this possibility arises. It arises because, in the process of using the intended applications, one has failed to take into account that the data integrated into the intended applications of a theory are collected independently of the theory to be tested. This identification of the problem also immediately suggests its solution. One simply must formulate the metatheoretical hypothesis that the data summarized in the intended applications of a theory are all determined independently of that theory, in the sense that this theory is not used at any point in their determination.

However, this 'solution' immediately raises the question of the status of such a metatheoretical hypothesis: Is it a descriptive, empirical statement about scientific practice, or is it a normative statement that seeks to prescribe or suggest a certain course of action to scientists? Since I regard the theory of the sciences as an empirical endeavor, I do not consider normative statements to be admissible as metatheoretical axioms. However, if one regards the metatheoretical axiom under consideration as an empirical statement about scientific practice, then it is not clear whether it is a correct statement. We are currently hardly in a position to sufficiently substantiate the correctness of this axiom on the basis of real cases of measurements and measurement

[13]This point is also nicely elaborated in (Hacking, 1983).

[14]For a complex example, see (Balzer & Wollmershäuser, 1986).

chains related to the testing of a theory. To do so, it would be necessary to examine at least a few such cases in detail from precisely this perspective.

However, my aim here is *not* to make a definitive statement as to whether science in general avoids the described pragmatic circles. What is important, first of all, is that this statement turns out to be *empirically testable*. This is not as self-evident as it might initially seem to the non-philosopher: Almost all metatheories to date have treated the metatheoretical axiom under consideration *not* as an empirical statement, but as a more analytical component of the concept of empirical science.

The structuralist approach provides a suitable conceptual framework within which the test problem can be analyzed more precisely. To this end, we introduce some auxiliary concepts.

We say that a model $x = \langle D_1, ..., D_k; A_1, ..., A_m; R_1, ..., R_n \rangle$ of a theory T is a *measurement model* for R_i if there are sentences (including law-like ones) in which no symbol for R_i occurs, which are valid in x, that uniquely determine R_i through the remaining components of x (possibly up to given scale transformations) and allow R_i to vary as a continuous function of the 'remainder' of x (relative to suitable, given topologies). Furthermore, R_i (possibly after suitable encoding) should be computable from a finite substructure of the remainder of x (i.e., $\langle D_1, ..., A_m; R_1, .., R_{i-1}, R_{i+1}, ..., R_n \rangle$) (possibly after suitable encoding).[15] By a *term* of a theory, we mean every type σ_i of the i-th relation R_i (for $i \geq n$) that appears in the models. Finally, we say that a theory T' is a *pre-theory* of a theory T if both theories have a common term σ_i such that there are measurable models for R_i in T' but not in T. A pre-theory T' of T is, in other words, a theory in which one can measure a term of T' that cannot be measured in T itself. Such terms do indeed exist,[16] and the concept of a pre-theory thus defined highlights certain dependencies within the network of existing theories: Every theory depends on its pre-theories, but not necessarily the other way round. The data that theory T derives from a pre-theory T' are, in the above sense, essentially influenced by T', but *not* by T. A set of theories together with the pre-theory relation just defined apparently forms a graph. It seems reasonable to start from a given theory in this graph, and proceed to its pre-theories, and from there to further 'second-degree' pre-theories, and so on. In this way, one obtains paths $\langle T_1, ..., T_n \rangle$ in the sense of graph theory, where T_1 is the starting point and all T_i are pre-theories of T_{i-1} (for $1 < i \leq n$).

Let us consider, within this framework, the test problem in its non-trivial version. Can significant pragmatic circularities arise when testing a theory? Can it happen that the theory being tested significantly influences the data against which it is tested? The answer 'yes', which has already been given, can now be made more precise by specifying exactly what such circles look like. Two cases must be distinguished.

A) The description of the intended applications I_r of T includes data that was determined by means of T itself, i.e., such data was obtained through measurement

[15]See (Balzer, 1988) for details.

[16]See (Balzer, 1985) for three concrete examples.

models that are models of T. This is a 'zero-length' circularity. The theory to be tested is used directly for its own testing.[17] Such circles can be easily ruled out at the meta-level by introducing a distinction between T-theoretical and T-non-theoretical terms. T-non-theoretical terms are those terms of T for which there are no measurement models in T. Such terms cannot be further specified with the help of models of T. If one requires that the data used for the test of T concern only T-non-theoretical terms, then circular arguments of the type under consideration cannot occur. For such data cannot have been obtained at all with the help of T and are therefore not significantly influenced by T. As a rule, such data originate from pre-theories of T, and sometimes also from direct observation.

B) In the applications I_r of T, data appear whose determination involves an entire chain of pre-theories, and among these pre-theories, T itself is also represented. In other words: In the graph of T's pre-theories, there are circular paths of the form $\langle T_1, ..., T_n \rangle$, where $T_1 = T = T_i$ for some i such that $1 < i \leq n$, and for all such i, T_i is a pre-theory of T_{i-1}. It is difficult to specify general *and* manageable rules by which such circular paths can be avoided.

This finally brings me to the problem of theoretical terms. It concerns the question of whether, in a given theory, there are terms that have a special status with regard to the validity of the theory and with regard to the verification of that validity – a status that may cause certain difficulties during the verification of validity. I have deliberately chosen this somewhat vague formulation because it makes it particularly clear that a 'problem of theoretical terms' is essentially related to what exactly is meant by theoretical terms. I would like to briefly mention here *only one* specific, new concept – already alluded to – that seems to me particularly relevant with regard to the proposed definition of validity. According to this view, a term in a *theory T is theoretical* if the theory provides means of measuring that term. In other words: The axioms of the theory contain equations that can be used to determine or calculate the term. Or to put it another way: Some of the models of the theory are measurement models for the term in question. This view seems to me to be most consistent with the historical development of science. For theoretical terms are always introduced in connection with a new theory, in which they acquire a new, specific, technical meaning. The new theory usually also provides *at least one* new method of determination or measurement for the term.

It turns out that this idea of theoretical terms can be refined into a purely formal distinction[18] if one takes into account the invariants of the theory in question, i.e., if the measure models have, with respect to the term under consideration, the same precisely definable invariance as the 'normal' models of the theory. It can be shown that the definition in question is equivalent to the following simple characterization: A term t is *theoretical* in theory T if and only if there are measurement models T for

[17] Cases of this kind formed the starting point for Sneed's 'problem of theoretical terms' (Sneed, 1971).

[18] See (Balzer, 1985) and (Balzer, 1986).

the interpretations R_i of t. We thus obtain a formal distinction between the terms of a theory into 'theoretical' and 'non-theoretical' ones. This approach will not lead to a genuine distinction in all cases. Some theories, such as the ideal gas law, are too 'local', too limited in their scope of application, for a distinction between theoretical and non-theoretical terms to be appropriate in their case. On the other hand, theories such as quantum mechanics are so comprehensive in their claim to validity that even for them the distinction described does not seem appropriate. Such theories incorporate – at least in principle – all their pre-theories, so that all measurement possibilities, including those based on pre-theories, ultimately lead to measurement models of the comprehensive theory, and thus all terms become theoretical in the sense introduced.

One must therefore relativize the definition of theoretical terms to the 'correct' applications, namely to theories that on the one hand are more comprehensive than merely empirical hypotheses such as the ideal gas law, but on the other hand are not yet so universal that they 'swallow up' all their precursor theories. For theories of this kind – such as classical particle mechanics, classical collision mechanics, phenomenological thermodynamics, but also microeconomic exchange theory, the formal criterion described above provides a distinction between theoretical and non-theoretical terms that corresponds well with the views of specialists regarding which values one must obtain elsewhere to verify the theory.

If one defines theoretical terms in the manner specified, then the *problem* with theoretical terms lies in the possibility – already mentioned – that, when verifying the validity of the theory, certain measurements themselves utilize this very theory. In other words, the problem lies in the fact that pragmatic circularity may (though not necessarily) arise when verifying the theory. It has already been stated that such circularity can – and should – be ruled out by axiom. And it has been pointed out that, due to the complexity of the issues involved, the empirical verification of such an axiom is difficult.

The point I am making here is that the conceptual framework of structuralist metatheory allows for a natural and non-simplistic analysis of such test situations. The question of whether circularity exists in empirical science is thus made amenable to empirical investigation. If one bears in mind that such questions have thus far been discussed almost exclusively on an a priori level and on the basis of certain metaphysical presuppositions, one recognizes in these admittedly rather abstract considerations a specific contribution to the theory of the sciences, which, in the spirit of my initial remarks, also represents a contribution to the use of the theory of the sciences.

Bibliography

Achieser, N. J. and Glasman, J. M., 1968: *Theory of Linear Operators in Hilbert Space*, Berlin: VEB.

Balzer, W., 1985: A New Definition of Theoreticity, *Dialectica* 39, 127–145.

Balzer, W., 1986: Theoretical Terms: A New Perspective, *The Journal of Philosophy*

83, 71–90.

Balzer, W., 1988: The Structuralist View of Measurement: An Extension of Received Measurement Theories, in C. W. Savage & P. Ehrlich (eds.) *Minnesota Studies in the theory of the sciences*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1992, pp. 93–117.

Balzer, W., Moulines, C. U., and Sneed, J. D., 1987: *An Architectonic for Science*, Dordrecht: Reidel.

Balzer, W., and Wollmershäuser, F. R., 1986: Chains of Measurement in Roemer's Determination of the Velocity of Light, *Erkenntnis* 25, 323–344.

Hacking, J., 1983: *Representing and Intervening*, Cambridge: UP.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A., 1971: *Foundations of Measurement*, New York: Academic Press.

Schrödinger, E., 1926: Quantization as an Eigenvalue Problem, *Annalen der Physik* 79 (Part 1: 361 ff., Part 2: 489 ff.) and 81 (Part 4: 109 ff.).

Schubert, H., 1964: *Topology*, Stuttgart: Teubner.

Sneed, J. D., 1971: *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel.

Stegmüller, W., 1973: *Theorie und Erfahrung*, Second Half-Volume, Berlin, etc.: Springer.

Zoubek, G., 1987: Zur Rekonstruktion der nicht-relativistischen Wellenmechanik, unpublished manuscript in DFG Project Ba 678/31.