A translation of: Balzer, W. 1985. Was ist Inkommensurabilität?, *Kant-Studien* 53, 196 -213.

What means Incommensurability?

by

Wolfgang Balzer[1]

Over the past twenty years, the phenomenon of "incommensurability" has stimulated philosophical discussion like few other topics. This is not surprising, since this topic brings together the traditional philosophical questions regarding the possibility of knowledge and the existence of categories – or any categories at all – with more recent, descriptive approaches in the philosophy of the natural sciences. The fact that the discussion has at times taken on a rather polemical tone, however, seems to indicate that there is still no particular clarity on the matter: each author appears to have a different intuition. The spectrum of views ranges from logical distinctions and remarks (Feyerabend, Scheibe) to pragmatically enriched concepts (Kuhn), to the assertion that we are dealing here not with a single concept but with an entire family of concepts (Stegmüller), and to the succinct classification that this is simply a variant of relativism.[2]

At the same time, the underlying phenomena are generally acknowledged. Since Thomas Kuhn's *The Copernican Revolution*[3], and inspired by further writings by Feyerabend and Kuhn, there seems to be agreement that the phenomenon of incommensurability is illustrated by a series of standard examples: Ptolemaic versus Copernican astronomy, Aristotelian theory of motion versus classical mechanics, the theory of impetus versus Newtonian mechanics, the phlogiston theory versus Daltonian chemistry, phenomenological thermodynamics versus statistical mechanics, classical versus relativistic kinematics. If one acknowledges that the examples mentioned are sufficiently similar to one another and sufficiently distinct from other non-examples – which we assume in what follows – then the task is to describe the characteristic feature common to all examples. We believe that previous attempts at explanation or elaboration

[1] Fellow at the Netherlands Institute for Advanced Studies (NIAS) in 1982–83.

[2] See, for example: P. Feyerabend, Explanation, Reduction, and Empiricism, *Minnesota Studies in the Philosophy of Science*, H. Feigl and G. Maxwell (eds.), 3rd printing, Minneapolis 1971; T. Kuhn, *The Structure of Scientific Revolutions*, 2nd edition, Chicago 1970; E. Scheibe, Comparability, Contradiction, and Explanation, in: R. Haller and J. Götschl (eds.), *Philosophy and Physics*, Braunschweig-Wiesbaden 1975; W. Stegmüller, *The Structuralist View of Theories*, Berlin-Heidelberg-New York 1979, as well as Kuhn's latest work, Commensurability, Comparability, Communicability, *Proceedings of the Philosophy of Science Association*, PSA 1982.

[3] Cambridge, 1957.

of the concept of incommensurability have not yet been satisfactory. This is not to say that they had no value as starting points or bases for discussion. But, as Paul Feyerabend puts it in his characteristic way:[4] "Apparently everyone who enters the morass of this problem comes up with mud on his head,..." – that is, at any rate, without a clear, acceptable concept of incommensurability.

At the risk that this might happen also to us, we would like to propose a new approach to explaining incommensurability. Inspired by a discussion with Thomas Kuhn and a paper by David Pearce on reduction,[5] this attempt seems promising to us.

I

An initial reflection on the examples mentioned above and the existing literature leads us to the starting point that incommensurability between two theories exists when both theories compete within a common domain and when both contain identical terms (words) with "locally" non-identical meanings.[6] One encounters, for example, the same word in both theories, but this word means something different in each theory.

It seems reasonable that such a "shift in meaning" constitutes both the core and the main difficulty in the concept of incommensurability. For in order to understand what a shift in meaning is, it seems necessary at first to gain clarity regarding the concept itself. This "natural" line of inquiry is indeed found in many writings on the subject. We believe, however, that one can already grasp the essential aspect of incommensurability without going all the way through the arduous process of explicating "meaning." The essential aspect does not lie in the fact that the meaning of the term differs in the two theories, but rather in the fact that this difference exists despite strong reasons that argue *for* an identity of meaning. The core of the phenomenon lies not in the shift in meaning per se, but in the discrepancy between the difference in meaning and the identity in other, important aspects. To explain such a discrepancy, one does not need a full theory of meaning; rather, one can rely on extension – as a part of meaning. Our further considerations should demonstrate this – and hopefully do so.

Let us consider a few examples. During the transition from the Ptolemaic to the Copernican system, the meaning of the expression "center of the world (of the cosmos, of the universe)" – which plays an important role in both theories – changed. Setting aside all linguistic and practical contexts in which this expression is embedded, we

[4]P. Feyerabend, Changing Patterns of Reconstruction, *British J. Phil. Sci.* 28 (1977), p. 363.

[5]Logical Properties of the Concept of Reduction, *Erkenntnis* 18 (1983). See also D. Pearce, Stegmüller on Kuhn and Incommensurability, *British J. Phil. Sci.* 33 (1982).

[6]The addition "local" can and will be explained later. In the following we always use the term "Bedeutung" in the way of the English term "meaning", which is in German language normally translated by the Frege's "Sinn". Although we find the German adaptation of the English "reference" linguistically awkward, Frege's terminology would, it seems to us, only cause even more confusion here. "Bedeutungsverschiebung" is therefore synonymous with "meaning-variance" in what follows.

can state – in a rough but nonetheless comprehensive way that captures the most important aspect – that the meaning of this expression changed with the transition from Ptolemaic to the Copernican system. For Ptolemy, our Earth was the center of the cosmos; for Copernicus, our Earth was replaced by a theoretical construct, a point near the Sun (often simply referred to as the Sun itself). In Aristotelian physics, the word "motion" roughly means what we today call "change." In classical mechanics, on the other hand, "motion" is understood in a much more restricted sense. There, motion is just "a change in the distance between particles over time," whereby "particles," in the sense of mechanics, must be idealized or at least capable of being idealized.[7]

With the transition from the theory of impetus to Newtonian mechanics, the meaning of the term "natural motion" changed. In modern physical terminology, one would instead speak of "free particles," but "free particles" or "freely falling particles" do not appear in historical texts (such as those of Buridan). Admittedly, the term "natural motion" does not explicitly play a major role in the theory of impetus, but it is implicitly important there: the only type of natural motion is, in fact, rest (i.e., inactivity relative to the distinguished reference frame "Earth"),[8] in contrast to uniform linear motion, which is also considered "natural" in Newtonian theory. In phenomenological thermodynamics, the word "state" refers to specific, phenomenologically identifiable phases in the course of thermodynamic processes. In statistical mechanics, on the other hand, the word often denotes a class of functions that contain information about the positions and momenta of the molecules involved.[9] Finally, in the transition from Newtonian to relativistic kinematics, the meaning of the word "clock" was changed. While "classical" clocks always have the same duration of tics regardless of their respective states of motion, the duration of tics of "relativistic" clocks depend on the relative state of motion of the clocks.

In all these examples, there are in fact two competing theories, both of which contain a common term (or even several), but this term has different meanings in the two theories.

Now, of course, one might object that our approach – which is limited to the three points mentioned above ("competition," "identical terms," "non-identical meanings") – is indeed heading in the right direction, but that it falls short because it excludes other aspects: for example, it does not address the different perspectives – "worldviews" – associated with both theories, nor the methods and values that are also involved. We have three responses to this. First, the three aspects we have emphasized are sufficient to distinguish incommensurable pairs of theories from commensurable ones. Just as, for example, a physicist in mechanics neglects the colors of particles, we here disregard

[7]The example is presented in T. Kuhn, What Are Scientific Revolutions?, Occasional Paper #18, Center for Cognitive Science, MIT. We do not have a sufficient grasp of the example from chemistry mentioned in the introduction to be able to specify a term in which a shift in meaning occurs.

[8]See also my paper Incommensurability and Reduction, *Acta Philosophica Fennica*, Vol. 30, Nos. 2–4 (1978).

[9]See, for example, K. Huang, *Statistical Mechanics*, New York-London-Sydney 1963, p. 140.

methods, values, and perspectives. Second, it is necessary for any scientific analysis to limit itself to only a very few aspects of the phenomena under consideration. Why should one demand, particularly in the analysis of the concept of incommensurability, that the "real" conditions be captured in every detail, when far fewer demands are made in other fields (e.g., in the natural sciences, in logic, in ethics)? Third, we assert – without further justification – that all other aspects of incommensurability, insofar as they are important for defining and capturing the concept, can be "defined" or at least "pinpointed" with the help of the aspects we have specified. For example, the fact that two theories represent different perspectives on the same reality is expressed (among other things) by the fact that the two languages of the theories are different.

II

For a more precise explanation, we must first establish a framework within which we discuss empirical theories, since incommensurability is, after all, asserted by such theories. It is not necessary here to commit us to a specific, technical concept of theory.[10] It is, however, necessary to distinguish between some essential components of theories, as they can be identified – at least implicitly – in any concept of a theory. We distinguish four components of a theory T: a language L, a class M of models, a class N of non-theoretical structures, and a set I of intended applications. A theory would therefore be (at least) a tuple $T = \langle L, M, N, I\rangle$.

Let the language L be a descriptive language, such that an interpretation of the terms of L, as well as an assignment of truth values to sentences of L – as is known from the various varieties of logic – is possible. In the simplest interesting case, L would be, for example, a finite language of first-order. In what follows, we always assume that the logic and the concept of a proposition are not called into question, so that we can identify the language L with the set of its "non-logical constants" (i.e., the descriptive symbols for individuals, predicates, and relations). Under this assumption, L determines the class of all structures for L, $Str(L)$. A structure for L consists of one or more individual domains (sets of objects), and for every descriptive symbol t of L, in the structure exactly one "denotation" of the appropriate type is found. This mean that t is an individual constant whcih refers to an object, or t is a predicate constant referring to a predicate (and analogously for multiple types, higher levels, and other more complex cases). In what follows, we always denote structures for L by y, y', y_1, y'_1, and the denotation of the term t in L within the structure y by $J_y(t)$, to indicate that t gets in the structure y a specific interpretation.

The class M of models is, in the general case, simply a subclass of the class of all structures for $L : M \subseteq Str(L)$. As a rule, M is characterized by a set α of propositions (axioms) of L: M contains exactly all structures in which all propositions in α true.[11]

The class N of non-theoretical structures is obtained by omitting the "theoretical terms" from all structures for L and grouping the resulting "substructures" into a

[10] An overview of the most important terms can be found in W. Balzer and M. Heidelberger (eds.), *Zur Logik empirischer Theorien*, Berlin 1983.

[11] A precise concept of validity is not required here, but can always be implied.

single class. Behind the distinction between theoretical and non-theoretical terms assumed here lies, of course, the idea that non-theoretical terms have a preferential status over theoretical ones in terms of "directess to reality." Anyone who finds this idea too one-sided can simply ignore the distinction and, in what follows, always identify N with $Str(L)$. For later use, we introduce a function $r : Str(L) \to N$, which assigns to every structure y for L the non-theoretical structure $x = r(y) \in N$ (the "reduct" of y) resulting from y by omitting the theoretical terms (more precisely: their interpretations). A sort of inverse of r is the "complement function" e, which assigns to every $x \in N$ the set of all "theoretical complements" of x, $e : N \to Pot(Str(L))$, $e(x) = \{y \in Str(L)/r(y) = x\}$. The classes $e(x)$ form (for the various $x \in N$) a partition of $Str(L)$, so that we can define an equivalence relation on $Str(L)$ by "$y \sim y_1$ if and only if there exists an x such that y is in $e(x)$ and $y_1 \in e(x)$." Intuitively, two structures y and y_1 are equivalent ($y \sim y_1$) if their reducts are identical.

The most delicate component of a theory consists of the set I of intended applications. Without this component, it is difficult to speak of an "empirical" theory, because in the most extreme case – with a purely formal language L and a well defined model class M – there is initially no reference to reality whatsoever. Empirical theories, however, refer to specific domains of application – namely, their intended applications. It is clear that such a connection to reality generally cannot be addressed without introducing a separate component, but the question of the nature of this component raises problems. Here, we oscillate between the extremes of an intended application as a "thing in itself" – that is, what underlies the application of the theory in a concrete situation as a real, unstructured substrate – and an intended application as a system that is already structured by the notions of theory and the language L. Since it is difficult to speak about a reality that has not yet been captured linguistically, we prefer to situate the intended applications more firmly on the side of linguistically structured systems. Following similar treatments,[12] we assume that every intended application of T is a non-theoretical structure, i.e., $I \subseteq N$. This means that in every intended application x, all non-theoretical terms of L are realized; in x, all non-theoretical expressions of L can be interpreted; in short: $x \in N$.

In this view of empirical theories, both syntactic elements (in the form of L) and semantic elements (in the form of M, N, and I) interact, and we will profit from this in the following.

III

Of the three aspects of incommensurability mentioned above ("competition," "identical terms," "non-identical meanings"), only the second is clear. We know what terms in a language are, and we generally have clear criteria for determining the identity of terms.

[12]Compare J. D. Sneed, *The Logical Structure of Mathematical Physics*, Dordrecht 1971 (2nd ed. 1979), as well as: W. Balzer, *Empirical Theories: Models, Structures, Examples*, Braunschweig-Wiesbaden 1982, specifically pp. 28 ff. and pp. 45 ff.

With regard to the third aspect – namely, the non-identity of a term's meanings – it makes sense to fall back on the extensions of terms. According to both the traditional view (extension of a concept) and the modern view,[13] the extension of a term is an essential component of its meaning. Since we are dealing with a shift in meaning in the present context, it suffices to fall back on the difference of a single component of meaning (the extension), and this fallback works because we have clear identity criteria for extensions (at least if we take the set-theoretical view, which is widespread today, as our basis).

But what is the extension of t in T? Clearly, it is not acceptable to identify the extension with the class of all possible interpretations of t in structures for L ($\{J_y(t)/y \in Str(L)\}$). For if, given a common term t, the type (the number of places, sort, and rank) of t in T and T' is the same, then the class of all set-theoretically possible interpretations $\{J_y(t)/y \in Str(L)\}$ must also be identical to the corresponding class formed with respect to L' ($\{J'_y(t)/y \in Str(L')\}$). Continuing the ideas on the concept of theories[14] developed in this journal, we use the models to determine the extension of t and to define the extension of t in T as the class of all interpretations of t in models of T: $\{J_y(t)/y \in M\}$. This definition is also consistent with the usual distinction between intension and extension of a concept, since the extension consists of the class of all those entities that fall under the concept "in terms of content," i.e., in the sense of the concept's intension. Mere "set-theoretically possible" entities $J_y(t)$, for which y is not a model, do not have the correct intension (which we identify with the internal structure of the models, as expressed in the axioms) and thus do not fall under the concept (do not belong to the extension). An even closer examination of the relationships in the case where the $J_y(t)$ $(y \in M)$ are functions or relations seems to lead to the view of the extension of t as the union of all $J_y(t)$ over $y \in M$: $\bigcup\{J_y(t)/y \in M\}$. In the case of functions, however, this raises certain problems of unambiguity, so we wish to avoid making a strong specification. For the following this is not necessary, insofar as our final explanation will not deal with extensions but only with local differences in meaning, and the latter concept does not presuppose the former.

Let us now tentatively consider whether it is possible to explicate the concept of incommensurability using the two aspects discussed so far: "identical terms" and "non-identical meaning." According to this, two theories T and T' would be incommensurable if there exists, a term t from the corresponding languages L and L', whose meaning (extension suffices for our purposes here) is different in T and T'. It is immediately apparent that this explanation does not achieve the desired result. To see this, one need only consider the case where T and T' deal with entirely different subjects – that is, where there is no real system for which both theories claim to be applicable. Despite such a separation of their extensions of application, it can indeed happen

[13]See, for example, H. Putnam, The Meaning of 'Meaning', in *Language, Mind and Knowledge*, 1975, K. Gunderson (ed.), University of Minnesota Press, Minneapolis.

[14]On Quine's Observational Sentences, *Kant-Studien* 72 (1981).

that both theories contain common terms. As an example, consider the concept of equilibrium, which appears as a fundamental concept in both thermodynamics and microeconomics.[15] The question of whether the extension of the term "set of equilibrium states (or distributions)" is identical in these two theories may indeed be subtle from a set-theoretical perspective and dependent on the strength of the set theory used; intuitively, however, one would expect both extensions to be different. In that case, thermodynamics and microeconomics would be incommensurable in the sense of the tentative explication. However, this result is unsatisfactory because it is trivial. Theories concerning different subject areas are trivially incommensurable because it makes no sense from the outset to look for a "common measure" among them (why should one?).

We thus see that the first aspect – according to which the theories under consideration must be in a competitive relationship with one another – is essential. At the same time, however, this is also the part that is most difficult to explicate.

IV

The difficulty stems from the fact that one can speak of a competition between two theories only if both have a common domain of application, which may consist of the intersection of the two theories' domains of application. In other words: both theories refer (at least in part) to "the same thing," to "the same aspects of reality." However, if we wish to explicitly include this reality in our analysis, we must speak about it and therefore categorize it in some way using a language. This could be done in two ways, neither of which is without its problems.

First, one could attempt to fall back on a third theory that "encompasses" the two competing theories and thus provides a kind of "neutral language of observation" for describing the "common extension of reference." However, even setting aside the fact that this approach is highly fictional in terms of the progress of science, it does not, in principle, lead to a satisfactory explanation of incommensurability, because one can generally "invent" yet another fictional fourth theory alongside the fictional third theory, which is incommensurable with the third. Such an explanation of incommensurability relative to a given "supertheory" is not realistic, leads to a regress in the manner indicated, and is also – as will become apparent – entirely unnecessary.

The second possibility consists in introducing a "set of real systems," S, whereby these systems need not yet be structured in the language of a theory. One could then say, for example, that two intended applications of the two theories under consideration refer to the same real system, or "capture" it. Although this second option relies on much weaker assumptions than the first option suggested above, it is nevertheless

[15]Compare the axiomatic treatments in W. Balzer, A Logical Reconstruction of Pure Exchange Economics, *Erkenntnis* 17 (1982) and C. U. Moulines, A Logical Reconstruction of Simple Equilibrium Thermodynamics, *Erkenntnis* 9 (1975). The distinction between theoretical and non-theoretical terms identified in the first paper no longer seems correct to us, though this does not affect the axiomatization itself.

unsatisfactory, simply because in concrete cases we cannot specify or "describe" such a "theory-neutral" set S. Mere allusion will not lead us to our goal.

Although we will temporarily introduce such a set S of real systems below for heuristic reasons, we do so with the caveat that this renders the explication inadequate. In general, we adopt as a condition of adequacy for the explication of incommensurability the requirement that, in addition to the two theories, no third entity independent of them may be used. Thus, for example, relations between T and T' are still permitted if they depend on T and/or T'; entities from S, on the other hand, are excluded because they are completely independent of T and T'.

The insight that we consider important and necessary in explaining "two theories refer to the same thing" is the realization that one must – and indeed does – manage without postulating new entities (such as S or a neutral supertheory). Such entities, which provide "objective criteria" for "referring to the same thing," must be replaced by relative criteria that draw solely on the given theories and their structures.

With such a restriction to two given theories, T and T', their competitive relationship can now be identified at both the syntactic and semantic levels. At the syntactic level, the competition between T and T' is generally manifested through identical terms. However, as shown above by the example of the concept of equilibrium, the occurrence of a term in two theories is only a very weak necessary condition and by no means a sufficient one. A more comprehensive characterization of the competitive relationship must therefore take place at the semantic level.

Here, one will bring the intended applications of both theories into play and attempt to compare them. If T and T' refer (in part) to "the same thing," must not their intended applications then overlap: $I \neq I'$? This latter condition would make sense if we conceived of the intended applications as unstructured, "real" systems. However, difficulties arise in light of our earlier assumption about theories, namely that the intended applications are supposed to be non-theoretical structures of the corresponding language. If, under this assumption, $I \cap I'$ is not empty, then there must be a non-theoretical structure for L that is also a non-theoretical structure for L', meaning that the non-theoretical sublanguages of L and L' must be identical. As a rule, however, in the case of incommensurable theories, it is precisely the non-theoretical structures (and thus also the corresponding sublanguages) that are different. Intuitively, two incommensurable theories represent entirely different perspectives on "the same thing," and if such perspectives – in their linguistic form – are already incorporated into the intended applications, then the latter must also be different.

Although, as explained above, we explicitly reject the recourse to a new, third entity that represents "the same," we wish to make temporary use of such an entity for heuristic reasons: we introduce the weakest possible "basis for comparison" for intended applications of T and T' in the form of a set S of real systems. Of course, this already presupposes certain, more general categories without which the concept of a system cannot be conceived. However, since we will later remove the relativization to the basis of comparison of real systems, this presupposition does not constitute an objection to our final explication. The class S contains all those real systems to which both T and T' refer. Using "encompass" as a technical term, we can also say that S consists of precisely those systems that are encompassed by both intended applications of I and of I'. Using the abbreviation "xEz" for "the intended application $x \in I$ encompasses the real system $z \in S$" (and a corresponding definition of E' with I' in place of I), we obtain the following diagram:

$$\begin{tikzcd}
 & I' \subseteq N' \arrow[r] & Str(L') \\
S \arrow[ur] \arrow[dr] & & \\
 & I \subseteq N \arrow[r] & Str(L)
\end{tikzcd}$$

The "real, shared background" for I and I' is formed by the set S of real systems, where each $z \in S$ is encompassed by both I and I':

$$\forall z \in S \exists x \in I \exists x' \in I'(xEz \wedge x'E'z).$$

Using E and E', a relationship can be established between I and I' such that two intended applications x and x' are correlated if they both encompass the same real system:

$$x\rho x' \text{ iff } x \in I \wedge x' \in I' \wedge \exists z \in S(xEz \wedge x'E'z).$$

Intuitively, one could say that ρ "identifies" intended applications that encompass the same real system. This identification of intended applications then naturally extends to the full structures of T and T'. Two structures y and y' are "views of the same system" if their reducts (i.e., their non-theoretical substructures) are related to each other via the ρ relation. We denote this identification, "lifted" to the theoretical level, as ρ^+.

$$y\ \rho^+\ y' \text{ iff } r(y)\ \rho\ r'(y').$$

The arrow $\rightarrow$ in the figure above stands for "embedding." N is embedded in $Str(L)$ if that every element of N consists of a substructure of a structure for L.

By referring to a minimal basis of comparison in the form of a set S of real systems, a relation between ρ and ρ^+ arise in natural form, so that intended applications and any other structures of two theories can be "identified". The problem with this picture, of course, is precisely the set S. Since we do not assume any linguistic or language-based characterization of S, it is not clear how S can be specified at all in concrete cases. The possibility of merely hinting at it is largely fictitious, and any other method employs language in some form. In accordance with the maxim "Whereof one cannot speak, thereof one must be silent", we therefore simply omit the set S as well as the associated relations of encompassing E and E' retrospectively. We are then confronted with two relations, ρ and ρ^+, between the intended applications and the structures of both theories. Their interpretation remains unchanged: "$y\ \rho^+\ y'$" (or "$x \rho x'$") expresses that the two structures y and y' (or x and x') encompass the same real system.

However, since these relations cannot be defined as above without S, E, and E', the question arises as to what extent the specified interpretation of ρ^+ and ρ can be ensured in some other way, where "in some other way" means "without resorting to S". The question is whether ρ and ρ^+ can be characterized "internally" – that is, solely by reference to $T$and T' – in such a way that they cannot be interpreted as anything other than identifications. It is quite clear that this question cannot be answered with a definitive "yes." But one can attempt to find internal characterizations that narrow the range of possible interpretations of ρ and ρ^+ further and further in the direction of identification.

V

In fact, such characterizations have already been studied extensively under the heading of "empirical equivalence." One finds several different concepts of equivalence in the literature,[16] but none of them seem entirely satisfactory to us, at least with regard to our topic. We therefore wish to briefly outline a concept of empirical equivalence that seems particularly suitable for establishing equivalence between incommensurable theories. Presumably, the following considerations can also be carried out using other concepts of equivalence. One need only "adapt" the concept of theory to the concept of equivalence in each case.

Let $T = \langle L, M, N, I\rangle$ and $T' = \langle L', M', N', I'\rangle$ be two empirical theories. For a relation ρ between non-theoretical structures ($\rho \subset N \times N'$), we define for the "full" structures the induced relation ρ^+ by

$$y\ \rho^+\ y' \text{ if and only if } r(y)\ \rho\ r'(y'),$$

where r and r' are the functions that "chop off" the theoretical terms. We say that T and T' are empirically equivalent with respect to ρ if the following three conditions hold: (1) ρ is a function from N to N' ($\rho : N \to N'$), (2) I' is the image of I under ρ ($I' = \rho(I)$), (3) for all y, y': if $y\ \rho^+\ y'$, then ($y \in M$ if and only if $y' \in M'$).

The following ideas underlie this definition. The relation ρ "identifies" the (more precisely: certain) non-theoretical structures of both theories. To every non-theoretical structure of T, exactly one corresponding non-theoretical structure of T' is assigned. This assignment is, of course, made with a view to identifying the intended applications that refer to "the same real system". In condition (2) of the definition, we require

[16] Some of these, for example, can be found in W. Balzer and J.D. Sneed, Generalized Net Structures of Empirical Theories, Part I, *Studia Logica* 37 (1977), Part II, *Studia Logica* 37 (1978).

that ρ indeed induces an identification of the intended applications: for every intended application of T, there corresponds exactly one intended application of T' under ρ, and in this way, all applications of T' are obtained. In light of our earlier explanations, it seems that condition (2) alone would suffice, so that ρ need not be defined for all elements of N. However, we insist on the stricter condition (1) for two reasons. First, this condition precludes a characterisation of ρ for intended applications in a non-verbal manner – such as by pointing on something. A relation ρ that satisfies condition (1) can only be characterized at the linguistic level (i.e., non-ostensive). Second, a ρ defined for all non-theoretical structures offers certain formal advantages that we do not wish to forgo (and which manifest themselves in the form of theorems such as Theorem 3 below). According to condition (3), the "induced identification" ρ^+ must respect the models of both theories. If a model y of T ($y \in M$) is identified by means of ρ^+ with a structure y' of T', then y' must also be a model of T' ($y' \in M'$), and vice versa.[17]

Two empirical theories are empirically equivalent if there exists a ρ such that T and T' are empirically equivalent with respect to ρ, that is, if there is an "identification relation" ρ between the non-theoretical structures that matches intended applications on one side with intended applications on the other, and in which the induced identification ρ^+ maps models onto models.

It can now be formally proven that the concept of empirical equivalence of theories also leads to an equivalence of the corresponding empirical claims of both theories (Theorem 2 below). The empirical claim of T is: $I \subset r(M)$, where $r(M)$ denotes the class of all non-theoretical structures obtained from genuine models by omitting the theoretical terms.[18] Intuitively, the claim states that every intended application can be completed into a model by adding theoretical terms. Theorem 1 below is a generalization of Theorem 2. Furthermore, we can prove that the induced relation ρ^+ is compatible, in a precise sense, with the distinction between theoretical and non-theoretical terms (Theorem 3). To do so, let us recall the equivalence of structures introduced above, which is defined by the equality of their reducibles: $y_1 \sim y$ iff $r(y_1) = r(y)$. Using this equivalence, the compatibility of ρ^+ mentioned above can be formulated as follows: When two structures ($y \; \rho^+ \; y'$) are identified, two structures y_1, y'_1 that are equivalent to them ($y_1 \sim y$, $y'_1 \sim y'$) are also identified with each other ($y_1 \; \rho^+ \; y'_1$). Intuitively, this compatibility condition states that theoretical terms under ρ cannot be identified "further" than is possible through the identification of the corresponding non-theoretical terms with ρ plus the determination of the theoretical terms by the non-theoretical ones. One could concisely (though not entirely correctly) say that ρ^+ relates "entire equivalence classes" of theoretical complements to one another.

[17] Unlike the treatment in Balzer and Sneed, Generalized Net Structures, we do not consider "constraints" here. Including them would complicate matters considerably, but would not alter the result.

[18] See also Balzer and Sneed, Generalized Net Structures, regarding the concept of an empirical claim.

Theorem 1: If T and T' are empirically equivalent with respect to ρ, and $X \subseteq N$, $X' \subseteq N'$ such that $X' = \rho(X)$, it holds: $X \subseteq r(M)$ if and only if $X' \subseteq r(M')$.

Theorem 2: If T and T' are empirically equivalent with respect to ρ, then so are the corresponding empirical claims, i.e., the following holds:
$I \subseteq r(M)$ if and only if $I' \subseteq r(M')$.

Theorem 3: If T and T' are empirically equivalent with respect to ρ, then ρ^+ is compatible with the distinction between theoretical and non-theoretical terms, i.e., for all y, y', y_1, y_1', the following holds:
if $y\ \rho^+\ y'$ and $y_1 \sim y$ and $y_1' \sim' y'$, then $y_1\ \rho\ y_1'$.[19]

In the context of incommensurability, condition (2) of the above definition of empirical equivalence will generally not be satisfied, because incommensurable theories usually overlap only in a subset of their intended applications. Thus, one cannot identify all intended applications of both theories with one another. In the above definition, I and I' must then be replaced by two subsets of intended applications

$$I_1 \subseteq I \text{ and } I_1' \subseteq I'$$

with the interpretation that for every structure in I', there is a corresponding interpretation in I_1' that captures the same real system, and vice versa. I_1 and I_1' represent the "common, real, intended domain" of T and T'. The problems that, as indicated above, could be solved by introducing a third theory or a set S of real systems, resurface with the distinction between I_1 and I_1' required here, and it seems as though the introduction of I and I' violates the adequacy condition introduced earlier, according to which, when explaining incommensurability, one must not use a new entity independent of T and T'. In fact, however, there is no violation, since I_1 depends on T and I_1' on T'; and the manner of characterizing I_1 and I_1' is likely, in principle, to be fraught with exactly the same problems as the characterization of the set of intended applications of a theory in general, so that the manner of distinguishing between I_1 and I_1' does not introduce any substantially new elements – at least none that are independent of T and T'.

We can now say that two theories T and T' are in competition with each other if they are empirically equivalent with respect to a relation ρ and under suitable restrictions on the domains of intended applications in the sense described above. At the moment, we see no way to precisely define "suitable restrictions" on the intended applications. Replacing "appropriate" with a mere "there exists" would allow for absurd counterexamples. Formally, one could add a "maximality condition" in addition to the existence condition, according to which I_1 and I_1' are the "best" possible

[19]To prove Theorem 1, one also needs the condition – which is not mentioned formally but is implicit in the content – that for every non-theoretical structure there exists a "full" structure, such that the former appears as a reduction of the latter. The theorem then follows from condition (3) of the definition of empirical equivalence with respect to ρ. Theorem 2 is a trivial consequence of Theorem 1 and condition (2) of the definition. Theorem 3 follows directly from the definition of ρ^+.

restrictions of I and I' that satisfy the remaining conditions for a fixed ρ. But this merely shifts the problem to the choice of the "correct" ρ. We believe that further conditions for ρ can and will be found that improve the explication of "competition"; we simply do not have these conditions at the moment. Thus, a residual issue remains to be addressed pragmatically; however, given the current state of the philosophy of science, this does not constitute a special case nor an objection to our explanation. On the contrary: an explanation that could be made fully precise at the outset would raise strong suspicions of its inadequacy.

VI

Let us summarize the considerations so far. Incommensurability consists in the presence of identical terms with locally non-identical meanings, whereby the theories under consideration must be competitors. We have analyzed the two non-trivial aspects here – non-identical meaning and competition – to a considerable extent. The task now remaining is to assemble these isolated explanations into an intuitive, homogeneous overall picture, in which the "locality" of the difference in meaning will also be addressed.

We begin with two theories, T and T', that compete with each other – that is, they refer (in part) to "the same thing". At the syntactic level, this manifests itself through terms that occur simultaneously in the languages of both theories. At the semantic level, there is an identification relation ρ, which links the non-theoretical structures of both theories, thereby relating subsets of the intended applications to one another and models (via ρ^+) to models.

Due to the existence of ρ, it can even be proven that the two empirical claims of the theories are equivalent when restricted to suitable subsets I_1, I_1' of intended applications to be used in the identification (Theorem 1 above). One is (almost) tempted to say that both theories are empirically equivalent and are therefore merely two equivalent "theoretical pictures of the same reality." Why, then, can we still speak of incommensurability here?

Well, one observes – and only upon closer inspection (which is consistent with the historically "late" discovery of the phenomenon) – that despite all their similarities, the meaning of a common term (or several) in both theories is indeed different: the term has different extensions in both theories. But it is not only the extensions – in the sense of sets of theoretically possible interpretations – that differ. It turns out that an even sharper form of difference in meaning occurs: "local" difference in meaning.

This occurs when, even for two models linked by ρ^+ ($y \; \rho^+ \; y'$), the interpretations of the common term t are different: $J_y(t) \neq J'_{y'}(t)$. That is, although both models refer to "the same thing," although both describe the same real system, different things are assigned to the term t in the two models. The interpretations $J_y(t)$ and $J'_{y'}(t)$ either involve "objects" of entirely different types (which is expressed by different set-theoretic types of the two sets $J_y(t)$ and $J'_{y'}(t)$), or – given that the same objects appear in both $J_y(t)$ and $J'_{y'}(t)$ – they establish relationships between these objects in different ways.

Here, one can distinguish between a weak and a strong form of local difference in meaning – and accordingly, one could also differentiate within the concept of incommensurability. If local difference in meaning exists *for all* model pairs linked by ρ^+, one speaks of *strong* local difference in meaning:

$$\forall y \forall y'(y \in M \wedge y' \in M' \wedge (y\ \rho^+\ y') \rightarrow J_y(t) \neq J'_{y'}(t)).$$

If the difference applies to *at least one* such pair:

$$\exists y \exists y'(y \in M \wedge y' \in M' \wedge (y\ \rho^+\ y') \wedge J_y(t) \neq J'_{y'}(t)).$$

this constitutes *weak* local difference in meaning. For the time being, as long as no detailed counterexamples exist, we prefer the strong form.

Consequently, two theories T and T' are incommensurable if they are in competition, if they contain common terms, and if these terms are strongly and locally different in meaning. In other words: there is a common term t (or several such terms) in the languages of both theories, and there are identifications of the non-theoretical structures of both theories as well as of the models; and within an "overlap region" – including the intended applications – t has a different interpretation (meaning) in each of the two models linked in this way. We believe and assert that this interpretation captures the core of the phenomenon of[20] "incommensurability".

VII

Before we conclude by returning to the standard examples, we should address one more problem that arises in the examples and is attributable to the fact that the theories under consideration are presented in axiomatized form. In an axiomatized theory, it has been possible to select a few "basic concepts" from the totality of all terms relevant to the theory, with the help of which the remaining terms can be defined; furthermore, a relatively small set of axioms has been selected from the class of all propositions assumed to be valid in the theory, from which the remaining propositions can be derived. When describing an axiomatized theory, therefore, one need only specify the basic concepts and axioms, and the examples mentioned at the beginning are usually presented in axiomatic form as well.

[20] Our explanation aligns quite well, in terms of wording, with a passage from Feyerabend: "....... in most cases it is impossible to relate successive scientific theories in such a way that the key terms they provide for the description of a domain D', where they overlap and are empirically adequate, either possess the same meanings or can at least be connected by empirical generalizations" (Explanation, Reduction, and Empiricism, 1971, p. 81). But Feyerabend makes no effort to elaborate on this idea further. On p. 74 of the same work, there is a completely different approach, which is brought much more to the fore through his treatment of examples. Furthermore, our explanation exhibits clear differences in content from the picture suggested in the above quotation. A more precise distinction, however, requires considerable effort and cannot be made in a footnote.

However, this raises the problem that our concept of incommensurability becomes dependent on the specific axiomatization chosen. For example, if T and T' are incommensurable in certain axiomatizations in the sense explained above, it may nevertheless be possible to find another axiomatization of T', say T'', such that T and T'' have *no* terms in common *at all.* T and T'' are then not incommensurable in our sense, because we have excluded this case of complete disjunction as trivial. As long as we do not have axiomatized theories before us, this problem does not arise, because there is then no other "version" T'' of T'. T'' will already be identical to T'.

In the case of axiomatized theories, however, the problem has a simple solution: we need only take into account the various possible axiomatizations. To do this, we move from T and T' to classes Θ and Θ' of empirically equivalent axiomatized theories. The elements of Θ then are all different, empirically equivalent versions of "the same theory." We call two such versions, $T \in \Theta$ and $T' \in \Theta'$, incommensurable if Θ and Θ' are incommensurable, and we define this, in turn, by the fact that there exist elements $T_1 \in \Theta$ and $T_1' \in \Theta'$ that are incommensurable in the sense defined above. In other words, we call two axiomatized empirical theories T_1 and T_1' incommensurable if there are empirically equivalent and incommensurable (in the sense of Section VI) versions T of T_1 and T' of T_1'. This definition may hold even though T_1 and T_1' initially have no basic concepts in common at all. An empirically equivalent version T' of T_1' can exist, that is formulated using one or more basic concepts of T_1. Consider, for example, the transition from classical to special relativistic kinematics. In the most "natural" axiomatization possible, the two theories have no basic concepts in common.[21] It is possible to reformulate both theories equivalently so that both operate with an undefined basic concept of "clock".[22]

Often, however, one will be able to avoid the laborious path of empirically equivalent versions by including defined terms in the analysis alongside the basic concepts. Our explanation from Section VI can then simply be adopted by allowing both basic concepts and defined terms as "terms." Of course, when using a formally structured semantics, the interpretation $J_y(t)$ of the defined term t must also be defined accordingly. It often happens that in "disjoint" axiomatic theories, the "common" terms necessary for incommensurability can be introduced by definition. Given two axiomatized theories T and T', one can therefore speak of incommensurability if it is possible to define common terms in T and T' such that the theories T_1 and T_1', extended by these terms, are incommensurable in the sense of Section VI. However, it is not entirely clear here in exactly which cases the two terms to be introduced by definition can be regarded as identical. Since this leaves room for ad hoc introduced inadequacies, we shall regard this version of incommensurability only as a particularly practical "approximation" (and not as a genuine explanation), which leads to a considerable

[21] Compare W. Balzer, Empirical Theories, Chapters III and IV.

[22] This possibility is hinted at in J. L. Synge, *Relativity: The Special Theory*, Amsterdam 1956, p. 38, and J. L. Synge, *Relativity: The General Theory*, Amsterdam 1960, pp. 105 ff. However, we are not aware of any detailed elaboration of this idea. We are indebted to Joseph D. Sneed for his extensive discussions on this point.

simplification in clear-cut cases. For example, in the most recently mentioned example of kinematics, there is agreement that in each of the two axiomatizations discussed in our book (*Empirische Theorien*), one can define new terms for which the word "clock" is used in both. Following such a definitional extension of the two theories, incommensurability follows without the need to elaborate full, empirically equivalent reformulations of the theories.

VIII

Now that we have covered all these concepts, let us return once more to the examples that provided the impetus for this discussion in the first place and for the treatment and delineation of which the entire apparatus is intended to serve. The reader should not expect exact proof that incommensurability in the explicit sense actually exists in all examples. The various cases are too difficult for this, and the theories involved have unfortunately not yet been presented ("logically reconstructed") with sufficient precision. Our attempt to clarify the theory of impetus[23] still seems inadequate to us with regard to the concept of impetus; in the thermodynamic example, a first step was taken by comparing the respective[24] "static subtheories"; and in classical and relativistic kinematics, the ρ relation is still missing.[25] All other theories mentioned are still awaiting reconstruction.

Nevertheless, we will attempt to demonstrate that our explanation is, if not accurate, at least plausible, by offering a few remarks on the examples.

In the example of Ptolemaic and Copernican astronomy, the identification relation ρ will consist of a mapping of the "particles" involved (Earth, Sun, planets, possibly fixed stars) and their orbits. The common term is "center of the world (of the cosmos, of the universe)", and the local difference in meaning lies in the fact that, in the Ptolemaic model, this term denotes the Earth (or a point in space occupied by the Earth), whereas in the corresponding Copernican model, it stands for a point in space that is by no means occupied by the Earth.

In the case of Aristotelian theory of motion versus Newtonian mechanics, and in the case of phlogiston theory versus Daltonian chemistry, any remarks we might make would be highly speculative, so we prefer to refrain from them.

When comparing the theory of impetus and Newtonian mechanics, ρ consists of an identification of the spacetime framework and the position functions. How "impetus" is to be incorporated here remains unclear at the moment. The term "natural motion" can be introduced by definition in both theories. A particle in the theory of impetus performs a natural motion if its position function has only one function value (one point of space); in mechanics, on the other hand, the trajectory (i.e., the graph of the position function) of the particle can be a straight line. It is clear that one can specify

[23]W. Balzer, Incommensurability and Reduction.

[24]See T. A. F. Kuipers, The Reduction of Phenomenological to Kinetic Thermostatics, *Philosophy of Science* 49 (1982).

[25]Compare W. Balzer, Empirical Theories, pp. 240 ff.

two models with identical position functions, such that a particle in these models performs a natural motion in the sense of Newton, but not in the sense of the impetus theory. Both models contain the same object o, but the term under consideration "the natural motion of o" has different local meanings in the impetus model and in the Newtonian model.

The relationship between phenomenological thermodynamics and statistical mechanics involves, among other things, a mapping of phenomenological states to "statistical" states, which constitutes a part of ρ. It is already clear here, for ontological reasons, that the term "state" has different meanings (denotations) in two models, however they may be identified. On the phenomenological side, it stands for an entity that is abstract, but is not further analyzed. On the statistical side, by contrast, it denotes a class of "conjugate coordinates." Regardless of the exact structure of ρ, there will thus be a local difference in meaning.

Finally, as regards the case of classical and special relativistic kinematics, one can, for example, take the reconstructions given in our book (Empirische Theorien, p. 240.) as a basis and derive them "with limit processes" from the corresponding space-time identification.[26] In both theories, "clocks" can be defined as particles whose trajectories have a specific mathematical form. It then becomes apparent that even in ρ-coupled models, the two types of clocks can be different.

[26]See W. Balzer, Empirical Theories, p. 220.