LAWS, THEORIES, AND RESEARCH PROGRAMS

Theo A. F. Kuipers

INTRODUCTION

Ernest Nagel [1961] has stressed as no one else the importance of the distinction between experimental laws and proper theories, where the latter aim to explain the former by introducing theoretical terms. This ‘law-distinction’ is one of the main dynamic factors in the empirical sciences. It will be dealt with in Section 1. Since there do not seem to be anything like theory-free or theory-neutral observation terms, the law-distinction is explicated on the basis of a theory-relative explication of theoretical and observation terms. It will also be shown how a similar explication of the main points can be obtained by starting from Popper’s so-called empirical basis; this possibility makes it even more surprising that Popper did not pay attention to the law-distinction.

The analysis suggests a disentanglement of the so-called theory-ladenness of observations. In particular, an observation may not only be laden by a theory, even if unladen by it, an observation may nevertheless be relevant to a theory, and even guided by it. After indicating some structural features of proper theories, we will close the first section with a brief presentation of epistemological positions involved in observational and theoretical knowledge claims of increasing strength.

Section 2, as a kind of bridging, but optional, intermezzo of a semi-formal nature, deals with one particular way to represent the structure of scientific theories in some detail. There are two main approaches to the structure of empirical theories. The statement approach conceives theories primarily as sets of statements. This approach has long been considered as the only and obvious approach, e.g., by Carnap and Popper. However, it is also possible to conceive theories primarily as sets of models. One version of this so-called semantic approach is the set-theoretic or structuralist approach. It has been introduced by Suppes [1957] and refined by Sneed [1971], Stegmüller, Balzer, and Moulines.

Its basic idea is that a theory amounts to the specification of classes of set-theoretic structures satisfying certain conditions. After briefly discussing the practical advantages of the structuralist approach in general it will be introduced stepwise; first without the distinction between theoretical and non-theoretical terms, then with that distinction in order to avoid circularity or infinite regress in measurement. The basic outline of the resulting representation of three examples will

Handbook of the Philosophy of Science: General Philosophy of Science - Focal Issues, pp. 1–95. Volume editor: Theo Kuipers. General editors: Dov M. Gabbay, Paul Thagard and John Woods. © 2007 Elsevier BV. All rights reserved.
be given: classical particle mechanics, the periodic table of chemical elements, and psychoanalytic theory. Then some further refinements will follow, viz., absolute versus relative empirical content, various ways of determination of intended applications, relations between theories, theory-nets, and constraints. Finally, we will briefly consider the usefulness of the structuralist approach for non-empirical theories.

In Section 3 we will present the more or less generally accepted view, since the 1980s, introduced by Kuhn and Lakatos, that the development of scientific research takes place by means of encompassing cognitive units, called research programs. We will distinguish four kinds of programs: descriptive, explanatory, design, and explicative. Explanatory programs will be given the main attention, followed by descriptive programs. Explicative programs in the philosophy of science are illustrated in this chapter, for example, the structuralist program in Section 2, and several of the other chapters in this handbook (notably, those of Psillos, Niiniluoto, and Mahner). Computational approaches in the philosophy of science illustrate a particular kind of design programs, viz. designing computer programs that can fulfil certain functions (see [Aliseda and Gillies, this volume]).

The main structural and developmental features of programs will be described, using Dalton’s atomic theory program to illustrate them. In the development of this explanatory program the law-distinction will turn out to play a crucial role.

Finally, we will address the strategic lessons that may be drawn. They involve in the first place the value of programmatic research as such, as well as some specific strategies for the internal development of programs, in particular, idealization and concretization. However, the strategic lessons concerning the interaction between programs, by competition or cooperation, are at least as important.

Apart from some aspects and specific formulations this chapter is essentially based on the works of others, notably Nagel, Popper, Sneed, Stegmüller, Kuhn and Lakatos. In principle, the three sections can be read independently from each other. In particular, as suggested already, Section 2 is an optional intermezzo for somewhat formally interested readers.

1 OBSERVATIONAL LAWS AND PROPER THEORIES

Introduction

In the empirical sciences the informal distinction between observational laws and proper theories plays a crucial role. Observational laws are supposed to describe observationally, usually experimentally, established regularities. Different names for roughly the same concept are: empirical, experimental or phenomenological laws, reproducible effects, inductive generalizations, general facts. Proper theories or systems of theoretical laws (together with definitions and other conventions), on the other hand, are supposed to explain such laws and to predict new ones, by postulating underlying mechanisms. For easy reference, we will call this distinction between proper and ‘improper’ theories, that is, observational laws, the
law-distinction. The law-distinction forms a crucial construction principle for the hierarchy of knowledge and therefore an important heuristic factor in the dynamics of knowledge development. However, it has occasioned philosophers of science much brain racking to explicate the law-distinction in a defensible way.\(^1\) Without doubt, the distinction is strongly related to the distinction between observational (or empirical or experimental) and theoretical terms. Whereas proper theories introduce theoretical terms, observational laws do not. But how can one make sense of this term-distinction?

The point of departure of the classical logical empiricists [Aliseda and Gillies, this volume] was a theory-free, hence theory-neutral, observational vocabulary. Starting from this postulate their explication of the distinction was obvious. Observational laws were by definition all those laws that could be formulated in this observational vocabulary. Proper theories on the other hand introduce new concepts not belonging to this observational vocabulary. Given their preference for the observational vocabulary the important question remained whether the new terms introduced by these theories could be reduced, in some way or other, to the observational vocabulary. However this may be, the existence of a theory-free observational vocabulary and the law-distinction were interwoven for the logical empiricists.

Gradually it became clear, even in empiricist circles, that the postulate of a neutral observational vocabulary was an unfortunate creation of the empiricist mind, a paradigm of wishful thinking not corresponding to anything in the empirical sciences. Looking backwards, the standard examples of observational laws, such as Galilei’s law of free fall or the Balmer series ordering the spectral lines of hydrogen, must have been dubious from the very beginning, for they are, at least \textit{prima facie}, not couched in a pure observational vocabulary. Non-empiricists were eager to embrace the doctrine that all observation was theory-laden. The most popular became the other extreme view, called meaning holism, which states that all terms occurring in a theory are laden with that theory, with the immediate consequence that an interesting distinction between a theory and the observational laws explained by it became impossible.

Empiricists like Nagel [1961], Hempel [1966; 1970] and Sneed [1971], started to elaborate the idea that certain terms occurring in a theory may be laden with that theory, whereas other terms may not. These latter terms may, however, nevertheless be laden with other theoretical connotations and with observational laws. In the relevant literature, however, these ‘theory-relative ideas’ have been presented or at least understood as just a reinterpretation of the \textit{two-level} distinction between an observational and a theoretical level. These two levels may enable accountability for part of the dynamics in science, the short-term dynamics, in particular the interaction between invention, evaluation and correction of

\(^1\)In the literature at least as much attention has been paid to the explication of the notion of ‘law of nature’. In particular, the question how to distinguish that notion from an accidental generalization is very difficult. We confine ourselves in this respect to a few references: [Nagel, 1961, Chapter 4; Bird, 1998, Chapter 1; Johansson, 2005]; and [Psillos, this volume, Section 12].
observational laws and proper theories. However, the picture hides the long-term dynamics. When a proper theory is accepted as (approximately) true, it usually enables the establishment of criteria for the determination of its theoretical terms. In this way it becomes an observation theory, and the corresponding theoretical level transforms into a higher observational level, enabling new observations and hence the establishment of new observational laws, requiring new, ‘deeper’ theories to explain them. Moreover, the acceptance of a theory enables experimental or technological applications of the theory, that is, applications presupposing that it is true. Of course, such applications will only be overall successful if the theory is in fact (approximately) true.

In this section it will be shown that the theory-relative ideas essentially lead to the suggested multi-level picture of knowledge and knowledge development. The two-level picture may then either concern just a fragment of the multi-level picture or it must be the result of a pragmatic contraction of essentially different observational levels. From the multi-level picture it becomes clear that the theory-relative move is a way of rejecting the idea of a neutral observational vocabulary that enables a new explication of the intuitive law-distinction that not only accounts for the short-term dynamics, but also for the indicated long-term dynamics.

The explication of the term- and law-distinction to be presented does not claim to do justice to the way in which some philosophers use the distinctions, but to the ways in which scientists use these distinctions. An impressive exposition of the far-reaching theory-laden character of what scientists call observations is given by Shapere [1982] under the revealing title “The concept of observation in science and philosophy” in which he uses several examples taken from astrophysics. Except when stated otherwise, we will also follow the scientific practice of already saying that a theory explains an observational law when it can (approximately) be derived from the theory. That is, we will say so whether or not one has good reasons to assume that the theory is true; it may even be known to be false. Hence, speaking of an explanation does not imply accepting it as a fully satisfactory explanation.

After the presentation, in Subsection 1.1., of some clear examples of observational laws and related proper theories, and a preliminary inventory of the characteristic differences, as potential conditions of adequacy, we will introduce in Subsection 1.2. the theory-relative distinction between theoretical and non-theoretical terms and use this distinction of terms for an explication of the law-distinction. The explication of the law-distinction will then make the postulate of a multi-level hierarchy of knowledge in terms of observational laws and proper theories highly plausible. The law-distinction will function as the construction principle for this hierarchy.

In Subsection 1.3. we will pay attention to the surprising fact that Popper pays so little attention to the law-distinction. He was not only one of the first proponents of the view that all observation is theory-laden, but by only assuming a theory-laden ‘empirical basis’, he did so without falling victim to the other extreme of...
meaning holism. It will be shown that from this ‘basis-relative’ perspective it is also possible to explicate the law-distinction. Given that this is not a difficult task and given Popper’s evident interest in the internal mechanisms of the development of knowledge, his neglect of the distinction is indeed surprising. It will appear to be instructive to dwell upon the good and the problematic reasons that may have been responsible for that neglect.

In Subsection 1.4. it will be shown that the perspective of Subsection 1.2. (and 1.3.) sheds also light on the idea of theory-laden observation. Three related notions can be clearly distinguished: theory-laden, theory-relevant and theory-guided observation.

In Subsection 1.5. we will first give an elementary account of the structure of theories, starting with the important distinction between epistemologically and ontologically stratified theories. In Section 2 we will present the sophisticated structuralist representation of (the structure of) theories. We will close Subsection 1.5. by briefly characterizing, mainly in terms of aspects of theories, the leading epistemological positions: epistemological relativism, along with observational, referential, constructive and essentialistic realism.

1.1 Examples and prima facie characteristics

We start the explication of the law-distinction by listing first a number of evident examples of both entities, and a number of *prima facie* characteristic differences that may serve as conditions of adequacy. Here, and later, we will speak of testing of a (complex) claim when we are only interested in its truth-value, and of evaluation of a claim when we are interested in its merits and failures. The first is usually the case with potential observational laws and the second with proper theories.

1.1.1 Examples of proper theories

In this section we will use ‘theory’, except when otherwise stated, to refer to a proper theory, a concept which is exemplified by the following theories represented here by a brief statement of their core ideas:

(a) Newton’s theory of gravitation. This theory states that all physical objects have a definite mass, that the sum of all forces exerted on an object equals the product of its mass and its acceleration, and that two objects exert an attractive force on each other proportional to their masses and inversely proportional to the square of their distance.

(b) The kinetic theory of gases. This theory postulates that gases consist of particles, called molecules, which exert forces on each other and which move in accordance with Newton’s laws of motion.

(c) Dalton’s theory of the atom (the example to be elaborated in Subsection 3.2.5.). This theory claims that all chemical substances are composed of
indivisible atoms. According to the theory these atoms can group together in certain ways to form molecules. The formation of molecules is associated with chemical reactions. Chemically pure substances are supposed to consist of one type of molecule.

(d) Bohr’s theory of the internal structure of the atom. According to this theory atoms are particles consisting of a nucleus and one or more electrons which circulate around the nucleus in fixed orbits. However, the electrons can jump from one orbit to another, absorbing or emitting electromagnetic radiation at the same time.

(e) (1) Mendel’s theory of genetics. According to Mendel the characteristics of (sexually reproducing) organisms are inherited by means of discrete genetic factors, called genes. For each gene there are different allelic forms in the game, each individual has a combination of two alleles, of the same or of a different form, and each parent transmits one of them to each of its offspring. That this is a 50-50% chance process amounts to the first law of Mendel’s theory, while the fact that the transmission of alleles related to different types of characteristics is independent is known as the second law of that theory.

(2) The theory of chromosomes. This theory states that in (eukaryotic) organisms the nucleus of each cell contains a number of pairs of so-called chromosomes, each consisting of two separate threads, called the chromatids. Each parent transmits by chance, in a very complex process, one chromatid to its descendant. The link with Mendel’s theory results of course from the fact that genes are materialized in a linear way in the chromosomes and that the alleles of a gene pair are supposed to be located on the corresponding positions of the two chromatids of a chromosome.

(3) The molecular theory of genetics. This theory tells that the material of the hereditary information consists of DNA-molecules. Moreover, the hereditary information is transformed by a special molecular mechanism to the offspring. The link with the previous theory results of course from the fact that molecular theory analyzes the chemical composition and working of the chromosomes.

(f) Festinger’s theory of cognitive dissonance. According to this theory the presence of cognitive dissonance, being psychologically uncomfortable, gives rise to pressures to reduce the dissonance and to achieve consonance. The strength of the pressures is a function of the magnitude of the existing dissonance.

(g) Utility theory or rational-choice theory. According to this theory people choose out of a set of alternative actions the action from which they expect the highest utility.
1.1.2 Examples of observational laws

The mentioned theories are said to be able to explain the following observational laws:

(a*) Galilei’s law of free fall stating that falling objects near the earth have constant acceleration.

(b*) The law that the velocity of sound is higher in gases with a lower density.

(c*) Proust’s law (or the law of definite proportions) according to which chemical compounds always decompose into component substances with constant weight ratios.

(d*) The Balmer series, which states that the wavelengths of light emitted by glowing hydrogen gas fit in a simple algebraic series.

(e*) Mendel’s interbreeding law on the fact that inherited characteristics manifest themselves after two generations in a certain statistical pattern.

(f*) The (quasi-)law stating that when people have made a decision there is active seeking out of information which is consistent with the action taken.

(g*) The macro-economic consumption function, which claims that total national consumption increases with increasing (average, and hence) national income.

1.1.3 Some characteristic differences

Let us now mention a couple of the characteristics of observational laws and proper theories, features that can help to consolidate the intuitive distinction of these two types of statements. We begin with an unimportant difference. To call a statement an observational law also means that it is well enough supported that it may be assumed to be (approximately) true. On the other hand, talking about a theory does not imply any veracity. Here we are essentially concerned with potential observational laws and theories apart from their truth-value, i.e., as hypotheses that may be true or false. Let us now turn our attention to relevant differences.

(i) Whereas an observational law is usually represented as one, possibly complex, statement, a theory is usually presented as a system, a coherent set, of statements (or as a variant of such a system). Of course, this does not exclude the possibility of an artificial representation of a theory as one conjunctive statement. With or without some extra definitions, even a reformulation in an elegant compact statement may be possible, in which case it is again tempting to speak of a law. The ideal gas law (see below) and the law of Archimedes (the upward force exerted on a solid body in a fluid is equal to the weight of the displaced fluid) are examples of this.
(ii) An observational law may specify what will happen under certain experimental conditions. Hence, it gives a partial characterization of what is not only conceptually, but also really possible in the context. The claim of a theory may be stronger: it may not only specify some necessary conditions for being really possible, it may claim to give a complete characterization of what is really possible in the context. But such a (relative) completeness claim is certainly not associated with every theory.

The first two differences do not only leave room for proper theories, but also for ‘observational theories’, i.e., coherent sets of (potential) observational laws for a certain context. Moreover, it may or may not be possible to summarize any theory in an elegant compact statement, and there may or may not be associated with it a completeness claim. Hence, the first two prima facie differences are not acceptable as strict conditions of adequacy.

(iii) Proper theories, however, not only use concepts that are used in the observational laws to be explained, but introduce also new concepts, called ‘the theoretical terms’ of the theory. For instance, Newton’s notions of mass and force do not occur in Galilei’s law; Dalton’s concepts of atom and molecule do not occur in Proust’s law; the notions of subjective utility and probability do not occur in the consumption function, etc.. (Of course, it may be that old terms are used, but then their old meaning is replaced by a new meaning provided by the theory.) On the other hand, observational laws do not introduce such new terms; for all non-logico-mathematical terms occurring in them there are independent application criteria in the form of experimental and argumentative procedures.

(iv) If an observational law can be explained by a theory, it can nevertheless be tested independently from that theory. This is of particular importance when some potential (corrections of) observational laws are predicted by the theory (and hence can be explained by it), and have still to be tested.

(v) The same observational laws can in principle be explained by different theories. It is for example conceivable that there would have been developed a new theory explaining the same laws as explained by Dalton’s theory, in which the notion of atom did not occur, although one or more rather different notions did occur. Hence, a theory can be rejected, without the consequence that the observational laws explained by the theory are also dragged down in its fall. When Bohr’s theory of the structure of the atom was rejected, this did not imply that the Balmer formula lost its descriptive adequacy.

From the classical logical empiricist point of view it was plausible to think that the fundamental difference between observational laws and theories, responsible for the above mentioned prima facie differences, is that observational laws are or at least can be expressed in pure observation terms, free from further assumptions. Although this assumption might be able to explain the differences, it should be
stressed that nothing that has been said so far implies that observational laws express regularities that can be (inductively) established on the basis of pure observation, i.e., observation not presupposing instruments or assumptions. On the contrary, it is not difficult to see that the testing of the observational laws mentioned presupposes all kinds of auxiliary assumptions.

Let us consider, as a tribute to Nagel, who used the same example, the testing of the innocently looking law (b*) about the velocity of sound in gases. To test this law we have to know how to produce and to register sound, and how to measure its velocity. Further we should know how to distinguish gases from substances in other aggregation phases, such as a liquid and a solid state, and how to measure the density of gases. All these identification and measurement procedures presumably presuppose the truth of certain theories. The measurement of the (mass-) density for instance requires the measurement of volumes and masses: both presuppose at least some general assumptions of stability and the like, and the first presupposes in principle a (naive or sophisticated) theory of space geometry, the second a theory of mechanics. Moreover, replication-measurements seldom lead to exactly the same results: to arrive at unique values on the basis of the test results presupposes general principles of dealing with ‘measurement mistakes’.

But if observational laws have no immediate relation to reality, what then is the fundamental difference between observational laws and proper theories? As suggested, the characteristic differences (iii) - (v) will serve as conditions of adequacy in the explication to follow.

1.2 Theory-relative explications

We will start with the theory-relative explication of the distinction between theoretical and non-theoretical terms, after which the explication of the law-distinction will be possible. This will naturally lead to the epistemological hierarchy of knowledge.

1.2.1 Theory-relative theoretical and non-theoretical terms

Let us consider in some detail the theory that is supposed to explain the law that the velocity of sound is higher in gases with a lower density (b*), viz., the kinetic theory of gases (b). In the context of this theory, sound is associated with wave movements jointly performed by the gas particles under certain conditions, and the velocity of sound then is identified with the velocity of these waves. The mass density of the gas is in this theory identified with the product of the number of gas particles per unit volume (the number density) and the mass of one particle. Theory (b) together with the mentioned auxiliary assumptions explains the law (b*) in the sense that the law is derivable from it.

Let us now look at the (non-logico-mathematical) terms of the theory, i.e., ‘gas’, ‘gas particle’, ‘sound’, ‘velocity of sound’, ‘wave movement performed by gas particles’, etc.. It is easy to see that some of these terms can be understood independently from the kinetic theory of gases, viz., ‘gas’, ‘mass density of a gas’,
velocity of sound in a gas’, etc. We know their meaning even if we do not yet know the kinetic theory. But also within the context of this theory these terms still have the same meaning: for example, it is still the case that we indicate with ‘gas’ a substance which is in certain respects different from liquid and solid substances. The same can be said about terms like ‘sound’ and ‘velocity of sound’: they have a clear meaning independent of the theory and they retain this meaning within the context of the theory. They are ‘antecedently understood’, to use Hempel’s phrase.

Let us now turn our attention to terms like ‘gas particle’, ‘wave movement performed by gas particles’, etc. These terms are not antecedently understood. On the contrary, what we have to understand by a gas particle is specified, or implicitly defined by, the theory itself, because it is the theory that introduces the term. By consequence, the correct use of the term presupposes the truth of the theory.

We will use the last point as the basic criterion for a general distinction between two kinds of (non-logico-mathematical) terms in relation to a statement $S$. We say that term $t$ is $S$-laden if the correct use of $t$ presupposes the truth of $S$, at least to some extent, and we say that $t$ is $S$-free if $t$ is not $S$-laden. We assume that this criterion can be made precise in such a way that it can always be applied unambiguously; see Subsection 2.4.2 for a structuralist specification. Note that we do not assume that $t$ occurs in $S$; in this way we leave room for the case that $t$ may be, in some way or other, indirectly laden with $S$. Be this as it may, it is also plausible to define that statement $S_1$ is (un-)laden with $S_2$ if $S_1$ does (not) contain terms that are laden with $S_2$.

Applying this definition to a theory $X$, conceived as a complex conjunction of statements, $X$-laden terms are also called theoretical terms with respect to $X$ or $X$-theoretical terms, and $X$-unladen terms are called antecedently understood or non-theoretical terms of $X$ or $X$-non-theoretical terms.

It is important to notice that in this way we do not make an absolute distinction between two kinds of terms in scientific language in general, but a theory-relative distinction: a term like ‘mass-density’ is non-theoretical with respect to the kinetic theory of gases. But, as noted before, the correct use of this term, defined as mass per volume unit, presupposes the truth of general assumptions and other theories, concerning the (macroscopic) notions of volume and mass. ‘Volume’ is theoretical with respect to Euclidean geometry; ‘mass’ is laden with Newtonian mechanics. With respect to these theories the term ‘mass-density’ is not antecedently understood but theoretical.

The foregoing definition of $X$-non-theoretical terms may even be liberalized in two respects. First, it may well be that the theory leads to a meaning enrichment in the sense that the theory may provide new criteria of application of the term to the already existing criteria. A new way of determining the term may be the result. Second, it may even lead to a proper meaning change in the sense that the old criteria of application are changed, but then only in such a way that the new criteria of application, though suggested by the theory, do not invoke it. In
the following the concept of X-non-theoretical terms is taken in this liberalized, sophisticated sense.

In the context of a particular theory X the terminology can be simplified by just speaking of theoretical and non-theoretical or even observation terms when the theory-relative qualifications are meant. The qualifications ‘non-theoretical’ or ‘observation(-al)’ may then of course not be misunderstood as implying ‘not laden with theories’.

1.2.2 Observational laws as improper theories

With the distinction between X-theoretical and X-non-theoretical terms we are close to a general explication of the intuitive distinction between observational laws and proper theories. The following formulation seems adequate at first sight: a theory is only a proper theory when it has at least some theoretical terms of its own, i.e., terms laden with the theory itself. An observational law, on the contrary, is an improper theory, a theory that has no theoretical terms of its own, i.e., no terms laden with the law itself. According to this characterization an observational law does not contain terms for which the correct use depends on the truth of the law.

It is easy to check that law (b*) satisfies this condition, and also that the other examples of observational laws satisfy it. However, there are also examples of laws which are, according to the proposed definition, not observational laws because they have theoretical terms of their own, whereas we are intuitively inclined to qualify them as observational law. A nice example is the ideal gas law \( PV = RT \) (\( P \): pressure; \( V \): volume; \( T \): empirical absolute temperature; \( R \): the ideal gas constant). Everyone calls it an observational law (to be sure, highly idealized), whereas it is at the same time generally known that \( T \) and \( R \) are laden with the law itself in one way or another. As a consequence, according to the above definition the law has to be qualified as a proper theory.

Closer inspection [Kuipers, 1982; 2001, Appendix 1 of Chapter 2] shows that the situation is as follows. Some observational laws can be formulated in the strict sense suggested above (hence, without \( R \) and \( T \) and also without other theoretical terms of their own) which are together sufficient to define \( R \) and \( T \) explicitly. Surprisingly enough, their conjunction is precisely equivalent to the, indeed very elegant, ideal gas law. Hence, although the terms \( R \) and \( T \) are, according to the theory-relative distinction between theoretical and non-theoretical terms, theoretical with respect to the ideal gas law, these terms can be explicitly defined on the basis of observational laws in the strict sense, and hence can be eliminated.

---

3The important restriction to asymptotic behavior has here been neglected. Contrary to Bickle’s suggestion [1998, 26], the asymptotic nature of the equivalence between temperature and mean molecular kinetic energy is unproblematic for the reduction of the ideal gas law. As shown in [Kuipers, 2001, Appendix 1 of Chapter 2], although temperature is asymptotically defined, the relevant identification is unrestrictedly that of equal thermal states with equal mean kinetic energy. The standard equivalence \( \langle u \rangle = (3/2) kT \), with \( \langle u \rangle \): mean kinetic energy, and
Many other scientific terms are introduced by explicit definition on the basis of observational laws in the strict sense. In particular, such laws may provide the existence and uniqueness conditions enabling the explicit definition, which is also the case in the above example. Another example of this kind is the notion of weight when based on the observational laws of the slide-balance, to be treated in Section 2.

Given the frequency of such definitions it is worthwhile to take them into account in the final definition. In this definition we will use ‘lawlike statement’ as a primitive term indicating general statements that may be considered for the qualification ‘observational law’ or ‘theoretical law’. Moreover, we will make a distinction between two kinds of inductive jumps that are made by accepting lawlike hypotheses as true. In case of observational induction or inductive generalization one essentially remains within the available observational vocabulary. In the case of theoretical induction, the acceptance of the relevant statements, among other things based on the observational laws they can explain, implies the conclusion that the new terms refer.\(^4\)

**Definition:**

**Observational laws, and its cognates**

An *observational hypothesis in the strict sense* is a lawlike statement not containing theoretical terms of its own, i.e., terms laden with the statement itself.

It is called an *observational law in the strict sense* when it is accepted as (approximately) true, a condition which requires observational induction.

An *observational hypothesis* is a lawlike statement containing at most theoretical terms of its own which can be eliminated with the aid of an explicit definition based on observational laws in the strict sense.

An observational hypothesis is called an *observational law* when it is accepted as (approximately) true, a condition which requires indirectly the observational inductions necessary for the defining observational laws in the strict sense and perhaps also directly some new observational inductions.

An *observational theory* is a coherent set of observational hypotheses. There may or may not be an associated (relative) completeness claim. That is, a claim to the effect that the conjunction of the observational hypotheses is the strongest true observational law for a certain domain that can be formulated with a given set of terms not laden with this

\(^k\): Boltzmann’s constant) results from this identity and the indicated observational laws and is usually, somewhat misleadingly, used in expositions of the reduction of the gas law.

\(^4\): For further refinements of the notion of induction, and abduction, see [Kuipers, 2004].
particular potential law in a non-eliminable way. If there is such a completeness claim we speak of a strong observational theory.

An observational theory is called an observational observation theory\(^5\) when it is accepted as (approximately) true, possibly, but not necessarily, as the strongest (approximately) true observational law, and used for observation.

**Proper theories, and its cognates**

A proper theory is a coherent set of lawlike statements, called theoretical hypotheses, their conjunction containing at least one non-eliminable theoretical term of its own; there may or may not be associated a completeness claim with it. If there is such a completeness claim we speak of a strong proper theory.

If the theory is accepted as (approximately) true, which requires theoretical induction, the constitutive statements are called theoretical laws, and the theory itself, when used for observation, a theoretical observation theory.

**Observational laws with respect to a theory**

A lawlike statement is an observational law with respect to a theory if it is an observational law that is not laden with that theory.

One might question whether the definition of a ‘proper theory’ cannot better be replaced by a corresponding definition of ‘theoretical hypothesis’. This would however be an unfortunate move, because of the fact that the theoretical hypotheses constituting a theory are usually interwoven in such a way that an isolated evaluation of, for example, the eliminability of terms would be unjustified.

It is also important to note the following. It frequently occurs that certain statements are considered as observational laws (to be) explained by a certain theory \(X\), whereas they are in fact clearly formulated in \(X\)-theoretical terms. For example, the results of Wilson chamber experiments, designed for the evaluation of theories in elementary particle physics, are usually couched in terms of orbits described by the particles postulated by the theory under evaluation. But the relevant aspects of these evaluation results can be formulated in terms of traces of water-drops. In such a case, if it is right, a reformulation is possible which avoids the use of \(X\)-theoretical terms. The resulting \(X\)-non-theoretical statements are, in that case, the genuine observational laws (to be) explained.

From their definitions it is directly clear that observational laws and proper theories satisfy the intuitive characteristic difference (iii), which was proposed as condition of adequacy in Section 1.1.3. However, they do not satisfy (i) and (ii), for an observational law may be a compact reformulation of an observational theory,

\(^5\)Though this terminology has a systematic background, it is certainly not very attractive; one might prefer to speak of a non-theoretical or experimental observation theory.
and such observational laws as well as proper theories may or may not be relatively complete, in contrast to what (i) and (ii), respectively suggested. The defined distinction also satisfies the conditions of adequacy (iv) and (v). An observational law can be tested independently from the theory which is supposed to explain the law, and it can remain well supported even when that theory is falsified, in which case the question then is: by what other theory can it be explained? This is easy to check in the example of the law about the velocity of sound in gases and the kinetic theory. If, however, a statement contains $X$-theoretical terms, i.e., when it is $X$-theoretical, it cannot be tested independently from $X$. Take for instance the statement that gases consist of particles. To test this statement we will have to know what to understand by gas particles, but it is precisely the kinetic theory that specifies this. We will have to presuppose this theory, or at least a part of it, in order to test the statement in question.

It may be true that a potential observational law can be tested independently of the theory that is supposed to explain it, but this fact makes testing not just the unproblematic affair suggested by common scientific parlance. For there are always underlying observational laws and (observational or theoretical) observation theories in the game. In most cases, however, the underlying laws and theories, with which the potential observational law in question is laden, are not in dispute; to put it differently, in most cases the underlying laws and theories belong to the background knowledge, i.e., they are assumed to be true. Against this background, assuming it as ‘underground’, one wants to know whether the potential observational law itself is true and whether the theory proposed to explain it does indeed imply the law.

This brings us to the hierarchy of knowledge.

### 1.2.3 The epistemological hierarchy of knowledge

Given that one statement may or may not be laden with another there are three possible relationships between two statements. They may be interwoven in the sense that the one is laden with the other and vice versa. Two statements are disconnected when neither of them is laden with the other. Finally, $S_1$ is an underlying statement of $S_2$ when $S_2$ is laden with $S_1$, but not vice versa. Of course, being an underlying statement of another is an asymmetric relation and it is safe to assume that it is also transitive, although exceptions to such a case are not inconceivable.

Conceiving observational laws as well as proper theories as (complex conjunctions of) statements, the relation of being an underlying statement of another leads to interesting cases of observational laws and proper theories underlying other observational laws and proper theories.

It is instructive also to consider the relation of an observational law being explained by a proper theory. Such a situation is of course an asymmetric relation. It should be noted that an explanation of an observational law by a proper theory need not be deductive. For example, such an explanation may be of a corrective type, in which case only a deductive explanation of some approximation (a cor-
rected version) of the law to be explained can be given. In other words, in this case the law itself can only be derived, and hence explained, in this approximate sense. Moreover, explanations of laws by proper theories always need auxiliary hypotheses, but they will not often be mentioned in the following.\(^6\)

If an observational law is explained by a proper theory then it has to be of course an observational law with respect to that theory in the sense defined above and hence the law may not be laden with the theory. The converse, however, is not excluded: an observational law \(L\) explained by proper theory \(X\) may or may not be an underlying law of that theory, depending on whether or not \(X\) uses terms presupposing \(L\). Moreover, if \(L\) is explained by \(X\) and if, in addition, \(X^*\) is an underlying proper theory of \(L\), then \(X^*\) will also be an underlying theory of \(X\). Note, moreover, that it is perfectly possible that two observational laws explained by the same proper theory share one or more underlying proper theories.

On the basis of the foregoing asymmetric relations there arises a hierarchy of crucial pieces of knowledge: a proper theory \(X\), the observational laws explained by \(X\), and their underlying proper theories and observational laws.

In Figure 1 we represent the suggested ordering, which we will call the epistemological hierarchy of the context.

The diagram indicates, for instance, that proper theory \(X\) explains observational law \(L\) and that \(X_1\) (or \(L_1\)) is an underlying proper theory (or observational law) of \(L\), and hence of \(X\). Of course, Figure 1 presents only a connected (and abstract) fragment of knowledge, which can in principle be extended to all sides. For example, if \(X\) has been accepted as (approximately) true, it can then be used as an observation theory with corresponding (new) observation terms. This use of it may lead, in combination with other accepted observation terms, to new observational laws. Furthermore, Figure 1 neither forces us to assume that there

\(^6\)See [Psillos, this volume, Section 13] and [Kuipers 2001, Chapter 3] for further details.
are fundamental observational laws and proper theories without underlying proper theories, neither must we assume that they do not exist.

It is important to note that Figure 1 presents an epistemological order which may not be interpreted ontologically in the sense of the lower the fragment the deeper the concerned level of reality. On the contrary, in particle physics and Mendelian genetics for instance, the ontological whole-part relation will roughly correspond to the upward direction. However, in cosmology for instance, the upward direction will sometimes correspond to the part-whole relation, e.g., from theories about heavenly bodies to theories about galaxies. Hence, there is not supposed to be any standard correspondence between the epistemological hierarchy of knowledge and an intuitive ontological hierarchy of the corresponding objects of knowledge.

The crucial question is of course whether local epistemological hierarchies frequently occur in scientific practice. Inspection teaches us that, for example, the observational law \((b^*)\) about the velocity of sound, explained by the kinetic theory \((b)\), does not indeed contain terms that are laden with the kinetic theory. Moreover, we have already noted that the law contains terms, such as ‘mass density’, that are laden with other theories, such as space geometry and mechanics. It is easy to see that these theories are laden neither with the law nor with the kinetic theory, hence they are indeed underlying theories of the law as well as of the kinetic theory.

Neither is it difficult to verify that the other examples of proper theories \((a), (c)-(g)\) and observational laws \((a^*), (c^*)-(g^*)\) explained by them are in accordance with the hierarchy.

The picture of the hierarchy has been restricted to the most essential elements. The following three features have been left out. First, an explanation of an observational law by a proper theory always needs auxiliary hypotheses, including observational laws and proper theories. Second, domains and subdomains have been omitted. Third and finally, besides observational laws explained by proper theories there are also other important forms of explanation, in particular, an observational theory explaining an observational law or theory, and, last but not least, a proper theory explaining another proper theory.\(^7\)

We confine ourselves to some examples of the last kind: Bohr’s theory of the atom \((d)\) explains (a corrected version of) Dalton’s theory \((c)\). Mendel’s theory \((e_1)\) is explained by the theory of chromosomes \((e_2)\), which is in its turn explained by the theory of molecular genetics \((e_3)\).

It is interesting to dwell further upon Mendel’s theory itself. It does not only explain observational laws, it also explains the core theory of population genetics, constituted by Hardy–Weinberg law, which states that, when there are no outside influences and no mutations, the gene ratio in a population remains constant over the generations. It is clear that this law is laden with Mendel’s theory, so it certainly is no observational law with respect to Mendel’s theory. On the other

\(^7\)In many cases, explanations of laws and theories are even called reduction. On closer inspection this is for various reasons, see [Kuipers, 2001, Chapter 3].
hand it predicts patterns of inheritance of outer characteristics that are, when approximately true, observational laws with respect to Mendel’s theory as well as population genetics.

The three elements not included in the above diagram would not alter the hierarchical nature of a more refined picture for the two relations involved: ‘to be underlying’ and ‘to be explained by’ remain asymmetric. We are therefore justified in using the simplified figure.

The underlying theories in the diagram essentially represent the proper theories and observational laws with which the terms occurring in the explicitly represented observational laws are laden. But in their turn, these proper theories and observational laws explain (other) observational laws, formulated in terms laden with other proper theories and observational laws, of still lower levels. By way of contraction we can collect together all the terms of all lower levels. Let us call this the contracted observational level, the observational level in short. Hence, determination of the terms of the observational level presupposes the (approximate) truth of all their underlying proper theories and observational laws, in consequence, the truth of all the observational laws explained by them. The combination of these laws and theories is called the background knowledge.

The acceptance of a (general) observational law, however it was generated, logically requires an ‘observational inductive jump’ from a finite number of singular observation statements, i.e., the data. The acceptance of a proper theory also requires a ‘theoretical inductive jump’ from the observational laws explained by it. Hence, although the background knowledge may be rather strong, it is not at all a foundation deductively based on data. It is a hierarchically ordered set of assumptions based on observational and theoretical inductions. We may speculate about its lowest level, if that exists at all. We might argue there that all terms are at least laden with some general assumptions, e.g., concerning durability and mutual relations, which guide their application. Apart from that, terms may be indirectly or directly applicable, that is, their application may or may not presuppose (observational or theoretical) inductions.

So far we have neglected a complication that arises due to the fact that one may accept a proper theory as true as far as its observational consequences are concerned, and that one may use it as an observation theory as far as its observation terms are concerned. Such occurrences may be called the empiricistic acceptance of a proper theory and the empiricistic use of a proper theory as an observation theory, respectively. Of course, the empiricistic use of a proper theory at least presupposes the empiricistic acceptance of it, and the inductive jump required for the empiricistic acceptance of a proper theory amounts to the observational induction or the theory’s observational consequences. In contrast to the empiricistic acceptance and use of a proper theory we may call the above straightforward definitions of the acceptance and use of a proper theory the realistic one. That is, the realistic acceptance of a proper theory amounts to the theoretical induction of all consequences of the theory, including its referential claims. And the realistic use

---

8For some more speculations, see [Kuipers 2000, Chapter 13].
of a proper theory as an observation theory amounts to its use for the application of all its terms, which of course presupposes its realistic acceptance.

The foregoing distinction between realist and empiricist attitudes is an initial indication of the different kinds of epistemological positions that are at present taken seriously. In the final subsection we will present a hierarchical survey of such positions.

In this subsection we have discussed the relative distinction between the theoretical and the observational level of a specific theory, and developed a theory-relative distinction of levels. However, these distinctions did not prevent us from giving a theory independent characterization of observational laws. This approach to the level distinction is primarily derived from Sneed [1971], but Nagel [1961] and Hempel [1966] also anticipated it. There is, however, another, completely compatible, approach to the level distinction and to the possible law-distinction, one which was also anticipated by Nagel and Hempel, and which is in addition easy to connect with Popper’s work.

1.3 The empirical basis

As mentioned, besides the theory-relative approach, there is another, completely compatible, approach to the level distinction and to the law-distinction, which is easy to connect with Popper’s work. We will first present this approach briefly and we will then speculate about Popper’s motives for neglecting the law-distinction. The next subsections do not presuppose any details of this subsection.

1.3.1 The basis-relative approach

We start with some of Popper’s core concepts. According to Popper [1934/1959] it is possible to reach provisional agreement in every scientific context about what belongs to the level of observation or, to use Popper’s favorite term, the (empirical) basis or the basic level. Popper has more than anyone else stressed the theory-laden, swampy character of this observation basis. It is however surprising that he did not make a number of plausible distinctions, let alone exploit them.

Let us call the (non-logico-mathematical) terms occurring at the basic level the \textit{basic terms}. As a matter of fact Popper reserved the term ‘basic statement’\footnote{In Subsection 2.2.3. we will present structuralist (non-statement) explications of some of the crucial terms of Popper, notably, ‘basic statement’, ‘counter-example’, and ‘empirical content’.} for a special type of statement that can be formulated in basic terms, viz., so-called singular existential statements, i.e., precisely those statements in basic terms about singular facts which can be in conflict with general statements.

The law-distinction can now be introduced as follows. Calling general statements which can be formulated completely in basic terms (general) observational hypotheses makes it also plausible to call such hypotheses observational laws when they are accepted for the time being after severe testing.
We are only concerned with a proper theory and hence with a theoretical level if it concerns a set of statements which, at least partly, breaks through the framework of basic terms. In other words, these statements should postulate new entities or attributes for which new terms have to be introduced that cannot be defined explicitly in terms of the available basic terms.

It is plausible to call the present approach to the level distinction the basis-relative approach. Notice that the corresponding characterization of observational laws is also basis-relative and hence not basis-independent, whereas in the theory-relative approach it was possible to give a theory-independent characterization of observational laws. It is nevertheless clear that the two approaches are essentially the same and that preferences will only depend on one’s further purposes.

1.3.2 Why neglect of the law-distinction?

Popper pays little attention to (the possibility of) the law-distinction, let alone to the importance of the distinction for the dynamics of science. We have to guess at Popper’s reasons for his lack of interest, because he is not explicit about it. This guessing may, however, be instructive. The only good reason we can think of is the impressive fact\(^\text{10}\) that strictly speaking the distinction is not necessary to characterize the logic of theory evaluation in the abstract terms of (singular) basic statements leading to falsification or confirmation (corroboration) of a theory. However, although this shortcut is possible, the distinction has to be introduced in a more realistic and sophisticated characterization of the evaluation of theories, and hence of the structure and development of science.

There is also a reason that has to be respected: Popper does not show any interest worth mentioning in the didactic of scientific textbooks. For someone who does have this interest and thinks that the law-distinction can be made relatively sharply it is clear that the distinction is not yet sufficiently exploited in textbooks.

Poor reasons\(^\text{11}\) for Popper’s lack of interest are also easy to conceive. In the first place, Popper was without doubt a victim of the misunderstanding that the law-distinction itself, or its importance, was related to the assumption of a theory-free observational vocabulary propagated by the classical logical empiricists. It is true that observational laws were considered by them as theory-free universal statements, or at least as reducible to such statements. In the light of the examples that were always mentioned as paradigms of observational laws, such as the laws of Galilei and Kepler, it gradually became clear that two interwoven, but distinguishable claims were involved. We have already shown above that the distinction can be based in a plausible way on a relative distinction of levels, and that Nagel already did this in 1961. The (rightly made) objections to a theory-free observational vocabulary do not provide a good reason for avoiding the distinction.

\(^{10}\)See [Kuipers, 2000, Chapters 5 and 6] or [Kuipers, 2001, Chapters 7 and 8].

\(^{11}\)From [Hark, ter, 2004] we may conclude that Popper tried to hide his inductive inclinations and psychological inspiration, without causing problematic features of his theory of science. Here we are dealing with a problematic aspect of that theory, the lack of the law-distinction, which partly originates from that same background in view of the three poor reasons.
A second poor reason has perhaps more of a psychological nature and is related to Popper’s oversimplified fight against induction. Popper is of course completely right in claiming that induction does not play and cannot play a role in the invention of proper theories. It is however also perfectly clear that observational laws, as far as they are not found by way of prediction by a theory, are frequently found by observational induction; i.e., they are thought of by way of inductive extrapolation. That is also the reason why they are frequently called inductive generalizations. So-called computational philosophy of science [Kuipers, 2001, Chapter 11; Aliseda and Gillies, this volume] exploits and elaborates the various methods of induction. Such methods do not alter the fact that inductively devised potential observational laws still have to be tested in the standard hypothetico-deductive way. A consequence of this is that the recognition of the importance of the law-distinction is almost impossible without simultaneously recognizing that as a matter of fact induction frequently guides the formulation of general observational hypotheses, but again this does not provide a good reason for avoiding the distinction.

The third and last poor reason concerns of course the under-estimation of the importance of the law-distinction for a realistic characterization of the structure and development of science. In general, one can say that Popper placed theories so central, that his attention to the role and nature of experiments became rather one-sided; they were seen from the point of view of theories. ‘The neglect of experiment’, the telling title of a book by Franklin [1986], was much severer in Popperian circles than in logical empiricist’s circles.12

Of course, Popper is well aware of the last point and when he is talking about explanation and prediction of facts (e.g., so-called ‘novel facts’), he is mostly referring to general facts, hence, observational laws. Many natural scientists also make this elision, but they do not claim, as Popper does, that evaluation of theories is straightforwardly conducted in terms of singular facts. Although such facts may play a crucial role, it is an indirect role, viz., in testing the observational laws predicted by the theory. And even Popper does not always stick to his doctrine. When the evaluation of theories is the main subject, basic statements by definition deal with singular facts, but when talking about basic statements in other contexts, it is often impossible to conclude otherwise than that such statements also include observational laws. This ambiguity, however, has a fortunate aspect, for what is more plausible than to call all statements that can be formulated in basic terms, basic statements?

From the foregoing we conclude that Popper, starting from his own premises, would have improved his analysis considerably by introducing and exploiting the law-distinction.

---

12In this respect the present author is strongly influenced by the logical empiricists, witnessed among others by the topic of the present section and the distinction between descriptive and explanatory research programs in the third section. Finally the fact that [Kuipers 2001, Chapters 3, 7 and 8] reflects that explanation and prediction primarily concern (potential) observational laws, with the explanation and prediction of singular facts as derivatives has an empiricist connotation.
1.4 Theory-ladenness of observation

The foregoing analysis of proper and improper theories can also be used to throw light on the so-called theory-ladenness of observation. Although it may be argued that all observation is in a sense theory-laden, we will see that this position does not imply that all observation is laden with any theory for which that observation is relevant, even if that observation was guided by that theory.

1.4.1 Theory-laden observations

The insight in modern philosophy of science that all statements about the world, however direct and unproblematic they may seem, rest on certain general assumptions, originates to a large extent with Karl Popper. For some other philosophers of science, such as Paul Feyerabend [1962; 1975] and, to a lesser degree, Thomas Kuhn [1962/1969; 1963], this insight has some rather negative implications for the possibility of theory evaluation. So it seems that Feyerabend’s ideas about theory-ladenness of terms boil down to the claim that all non-logico-mathematical terms are laden with all theories in which they occur. But this would inevitably lead to the consequence that, for instance, evaluating the theory of Newton with Galilei’s law of free fall cannot be more than a circular procedure. For under these circumstances the empirical establishment of the regularity expressed by Galilei’s law would already be laden with the theoretical principles of mechanics. Or, as it is also stated, the meaning of the terms occurring in the formulation of the law (distance, time) is determined, among other things, by the principles of mechanics. Some of Kuhn’s expositions seem to lead to the same conclusion.

The correct answer to this threatening impasse seems to be the following. We can concede to authors like Feyerabend and Kuhn that such statements as the distance covered by an object near the earth is proportional to the square of the elapsed time required are laden with theories; viz., theories concerning space and time measurement. But such statements therefore do not need to be laden with the specific principles of mechanics. In other words, the meaning of concepts such as position and time, needed for the formulation of such laws, may be fixed in and by theories existing independently from mechanics. In fact, it is in principle possible to reject mechanics, without changing our opinions about space and time measurement. For this reason, the proposed way of evaluating mechanics is not circular, although it is a conditional evaluation: assuming that the theories with which Galilei’s law is laden are true, Newtonian mechanics is supported by this law, at least in the sense that it can explain it. As is well known in the history of science, the condition had to be abandoned: Newton’s ideas about space and time had to be fundamentally revised, a need that was satisfied by Einstein’s theory of relativity. However, this does not exclude the possibility that the described condition could have been fulfilled, and that the suggested positive result of the conditional evaluation is still defensible as a first approximation of actual history.

The above mentioned discussions are frequently put in terms of the theory-ladenness of observations or facts, instead of statements. We shall reformulate the
purport of the foregoing in terms of observations. To begin with, it is plausible to make a sharp distinction between the unspecific statement that a certain observation is theory-laden and the specific statement that this observation is laden with theory $X$ or, in short, is $X$-laden.

We call an observation $X$-laden when it is, and has to be, formulated in an $X$-theoretical statement, that is, a statement essentially using an $X$-theoretical term in the sense of Subsection 1.2.1. Of course, $X$-laden observations cannot be used to evaluate $X$ without being confronted with the kind of circularity of which Feyerabend and Kuhn were thinking.

Fortunately, however, not all observations that might be relevant to $X$ are laden with $X$: there are also $X$-unladen observations, i.e., observations that are or can be phrased in terms of $X$-non-theoretical statements. Such observations can of course be used to evaluate $X$ without running into circularity problems. Though $X$-laden observations cannot be used to evaluate $X$ itself, they can perhaps be used to evaluate another theory, in which case it is necessary to presuppose $X$ as an unproblematic background theory, for example as an observation theory.

Let us now also consider the unspecific statement: “All observations are theory-laden”. The logical empiricists assumed in the beginning that there are observations that are not laden with general assumptions at all. According to them, the corresponding statements, which would hence be testable without assuming any theory, constituted the class of neutral, theory-free observational statements. The insight that this class is empty might be called Popper’s insight: all observations are theory-laden, or at least laden with general assumptions. This, however, does not imply the view that all observations that we can make in order to evaluate a certain theory $X$ are laden with $X$ itself. As mentioned, the latter view was held by Feyerabend and, to a lesser degree, by Kuhn. It has always been severely criticized by Popper, roughly along the suggested lines: for every empirical theory $X$ there are theory-laden but $X$-unladen observations, which can be used to evaluate $X$.

1.4.2 Theory-relevant and theory-guided observations

Now we want to discuss two questions that are usually also at stake in discussions about theory-laden observation. The first is that observations may or may not be relevant to or interesting for a certain theory. The second is that (relevant) observations may or may not have been governed or guided by a theory. In both cases we are primarily concerned with observations that are not laden with the theory in question. $X$-laden observations that are not relevant to $X$ in one way or another are difficult to conceive, and they are by definition guided in a certain sense by $X$, although it is not necessary that one always realizes this point. $X$-unladen observations on the contrary may or may not be relevant to $X$ ($X$-relevant), and if they are relevant to $X$, they may or may not be guided by $X$ ($X$-guided). The foregoing is summarized in Figure 2, which gives a classification of observations in relation to theory $X$. 
As far as relevance is concerned, the foregoing can also be formulated in terms of facts (in the sense of conceptualized facts or data), which may be individual facts or general facts, i.e., observational laws. In this perspective we see that $X$-laden facts are always relevant to $X$, but $X$-unladen facts may or may not be relevant to $X$.

The notion of relevance that is at stake in the present context is the idea that a certain ($X$-unladen) fact is relevant to theory $X$ if $X$ is not indifferent with respect to this fact. That is, if $X$ explains this fact, deductively or probabilistically, or if it contradicts it, in both cases with or without some relatively unproblematic auxiliary hypotheses.

There is also, however, a second sense of relevance which may be even more important because there need not be any consensus about it among different scientists. Facts with respect to which a certain theory is indifferent can nevertheless be considered as relevant to that theory in the sense that one may think that the true theory in question should not be indifferent with respect to these facts. However, the proponent of a certain theory can of course also be inclined to consider a fact as irrelevant when his theory is indifferent with respect to that fact.

Newton for instance did not consider it important that his theory neither explained nor contradicted Bode’s law, which gives a simple mathematical relation between the radii of the planetary orbits; the indifference of his theory with respect to this law was, according to Newton, no objection to his theory. Kepler, on the contrary, still insisted that the true theory about the solar system should be able to explain Bode’s law, and Kepler did have an explanation, which he based on Pythagorean ideas about numerical harmony. Following Newton, present day astronomers also think that Bode’s law is irrelevant, an accidental feature of actual orbits with a questionable status as a proper observational law. As a result, not only do we not have any explanation of this law but we also do not feel any need to have one. The phenomenon that a later theory does not give an explanation for a fact that had an explanation before but for which an explanation is no longer considered to be required is called Kuhn-loss.

If we want to evaluate a theory $X$ we aim at $X$-unladen observations that are relevant to $X$ in the first sense: we let ourselves then be guided by $X$ in helping
us to decide what to pay attention to. But also after the successful closure of the evaluation phase, when the theory has been accepted, at least for the time being, much research is guided by the theory. The periodic table (PT) of Mendeleev is not only a perfect example of a theory that was evaluated by PT-unladen, but PT-relevant and PT-guided predictions of chemical elements. It later also became an important means for predicting the possibility of artificial production of new elements\(^{13}\). All these cases concern observations that are guided by a theory but not laden with that theory.

Observations that are not guided by a theory are frequently called ‘accidental observations (or discoveries)’. Accidental observations can of course be perfectly relevant to a theory. A nice example is the Balmer series, which he discovered by trial and error on the basis of data provided by Angström. Hence, the discovery was not guided by a specific theory, but at most by some global Pythagorean ideas. Even so it was recognized as very relevant for the later developed theory of Bohr, for that theory was far from indifferent to the Balmer series: it could explain the series.

1.5 The structure of proper theories and the main epistemological positions

In a first analysis of the structure of theories we will emphasize the distinction between two main types of stratification, viz., epistemological and ontological. We will also pay some attention to non-empirical theories. Finally, we will present the main epistemological positions with respect to proper (empirical) theories and theoretical terms.

1.5.1 Epistemological and ontological stratification

Let us start by summarizing some of the main points made in this chapter so far. A proper theory \(X\) has been defined as an epistemologically stratified theory in the sense that it contains terms, and hence statements, that are laden with one or more of its principles: \(X\)-theoretical terms. The other terms of \(X\) are called \(X\)-non-theoretical. In contrast to proper theories, observational hypotheses are defined as improper theories, containing no theoretical terms of their own. A set of connected observational hypotheses is called an observational theory. It should be noted that being \(X\)-non-theoretical is a theory-relative, to be precise, an \(X\)-relative qualification of a term or a statement: they may well be laden with underlying theories. However, when the theory is clear from the context, a point that we will assume from now on, we will simply speak of theoretical and non-theoretical terms and statements, respectively.

\(^{13}\)To be sure, the kind of predictions mentioned is of a weak nature. Moreover, one may dispute whether PT is a proper theory or merely a classificatory observational law (see [Mahner and Bunge, 1997, 245-7]). However, we will indicate in Subsection 2.3.3. why PT initially was a proper theory, but transformed into an observational law in the light of quantum mechanics.
The main function of a proper theory is the explanation and prediction of observational laws relative to the theory, i.e., true general hypotheses containing no terms laden with them or the theory. For this function the distinction between observational laws and proper theories is of course crucial.

There is much more than this to say about the structure of proper and observational theories. Here we confine ourselves to some main points. In Section 2 we will present the so-called structuralist way of representing theories in detail.

Besides epistemological stratification there is ontological stratification: they frequently go together, but are essentially independent. A (proper or improper) theory is said to be ontologically stratified when there are two or more kinds of entities involved and when entities of one of these kinds are components of entities of the other kind. It is then plausible to speak of a lower, micro-level and a higher, macro-level. In this case some principles of the theory concern only the micro-entities, and their properties and relations, and are called micro- or internal principles, whereas others connect the different kinds of entities, and their properties and relations, and are called bridge principles. The example of the atomic theory (dealt with in Subsection 3.2.5.) provides a nice example of an ontologically as well as (along the same lines) epistemologically stratified theory. Of course, auxiliary hypotheses may also have an internal or a bridge character.

Another feature of some theories is that the principles of a theory, whether ontologically and/or epistemologically stratified or not, can frequently be differentiated into core or generic principles, claimed to be true for the whole domain concerned, and special principles, only claimed to be true for a certain subdomain. Of course, a similar distinction can be made for auxiliary hypotheses.

In the case of an epistemologically stratified theory it is plausible to define three types of statements: non-theoretical, purely theoretical and mixed (theoretical) statements. The division of theoretical statements in purely theoretical and mixed ones, however, seems only useful when the epistemological stratification reflects an ontological stratification, in which case the purely theoretical principles, i.e., the internal principles, constitute a clearly separable theory dealing only with the theoretical level. Compare the insightful distinction between the internal and the bridge principles of the atomic theory, in Subsection 3.2.5, with the principles of the (ontologically unstratified) theory of gravitation. In the latter example the distinction between pure principles (e.g., action = minus-reaction, the third law) and mixed principles (e.g., “\( f = ma \)”, the second law; and the special law of gravitation) plays no significant role.

In the case of epistemologically and/or ontologically stratified theories there is a natural distinction of two vocabularies: the complete vocabulary in which the theory is formulated, including theoretical and/or micro-terms, and the sub-vocabulary generated by the non-theoretical and/or macro-terms.

Of course, even if none of both stratifications apply, viz., when we are considering an ontologically unstratified observational theory, it may still be useful to make a distinction between the full theory and the corresponding vocabulary and a sub-theory and the corresponding sub-vocabulary. For an observational theory
may be designed to explain an observational sub-theory. Whatever kind of theory, our discussion more or less implicitly assumes that a theory can be formulated in terms of a finite number of principles. This feature can be conceived of as a very informal type of finite axiomatizability, which is a *conditio sine qua non* to talk about a theory at all. However, this condition should not be confused with the claim of finite axiomatizability in the sense of first or higher order logic. As a matter of fact, in Section 2 we will only illustrate the structuralist claim that it is possible and instructive for many theories to finitely axiomatize them in the set-theoretic sense of set-theoretic structures, defined by a finite number of axiom schemes, using as much mathematical language as necessary.

Explanation and prediction of observational laws have already frequently been mentioned as functions of theories. As additional functions, or at least additional forms of observational success, we should mention: unification, correction and enrichment. A theory may unify, by explanation, a number of *prima facie* rather heterogeneous observational laws. It may predict successfully a corrected version of an observational law, implying that the latter apparently was at most approximately true. Finally, it may predict observational laws concerning new observable phenomena. Of course, theories may also have theoretical success. One example is the conceptual unification of two previous theories into a new theory that is observationally equivalent to their conjunction. Another example is a theory providing a ‘deeper’ explanation of a proper theory.

1.5.2 Conceptual theories

Theories are up to now understood as empirical theories. Following Popper, we say that a theory is an empirical theory in the strict sense if it is, in combination with certain special or auxiliary hypotheses, falsifiable. And it is an empirical theory if it is intended to become an empirical theory in the strict sense, i.e., if one aims at special or auxiliary hypotheses that make the theory falsifiable. For instance, a generic theory, like Newton’s general theory of motion, may well be unfalsifiable as it stands, but become falsifiable together with appropriate special principles, such as the law of gravitation.

However, it also makes sense to leave room for theories that are not intended to be made falsifiable. In Subsection 2.5, we will indicate a number of kinds. Here we will restrict our attention to conceptual theories. A conceptual theory is intended to provide a perspective, a way of looking, at a certain domain without making a general empirical claim. Of course, conceptual theories may or may not be ontologically and/or epistemologically stratified.

The claims that are associated or made with a conceptual theory are either logico-analytic or restricted to individual intended applications. A typical logico-analytic claim is a theorem stating that the instances (models) satisfying the theory can be proven to have a certain explicitly defined property. A typical specific claim states that a certain intended application is (or is not) an instance (model) of the special theory. The very distinction advocated between observational laws and
proper theories is an example of a conceptual (meta-)theory for the domain of lawlike statements. This example makes clear at the same time that a conceptual theory may well be the result of concept explication. Of course, the claim that the result of concept explication, a conceptual meta-theory, roughly captures an intuitive concept or distinction is a (quasi-)empirical meta-claim. However, the main point is that, although it may always be possible to formulate a falsifiable general claim with a conceptual theory, the (meta-)claim that all theories are observational theories is usually not intended.

As already suggested, generic theories may well be unfalsifiable as such. They cannot only be made falsifiable, they can also be used as purely conceptual theories.

1.5.3 Epistemological positions

Returning to empirical theories, the core of the ongoing instrumentalism-realism debate concerns the nature of proper theories, or rather the attitude one should have towards them. Here we will briefly sketch the most important epistemological positions in that debate, viz., instrumentalism, constructive empiricism, referential realism and theory realism. In the introductory chapter of [Kuipers, 2000], they are more extensively introduced and ordered according to the ways in which they answer a number of leading questions, where every next question presupposes an affirmative answer to the foregoing one. Moreover, the questions are considered from four perspectives on theories. On the one hand, theories supposedly deal primarily with ‘the actual world’ or primarily with ‘the nomic world’, that is, with what is possible in the natural world. On the other hand, one may primarily be interested in whether theories are true or false, or whether they approach ‘the truth’ regarding the world of interest. It should be stressed that ‘the truth’ is always to be understood in a domain-and-vocabulary relative way. Hence, no language independent metaphysical or essentialist notion of ‘THE TRUTH’ is assumed.

The survey of positions and the analysis are restricted to the investigation of the natural world and hence to the natural sciences. Several complications arise if one wants to take the social and cultural world into account. However, the survey of epistemological positions in the natural sciences may well function as a point of departure for discussing epistemological positions in the social sciences and the humanities.\(^{15}\)

As we have seen, proper theories arise from the two-level distinction between observation and theoretical terms, as opposed to observational laws and observational theories, which only use, by definition, observation terms. Recall that the resulting two-level distinction between observational laws and proper theories

\(^{14}\)It is plausible to conclude this section with a brief treatment of epistemological positions as they arise from the above treatment of the law- and level-distinction. However, Ladyman deals in his chapter extensively with epistemological positions, including ‘structural realism’.

\(^{15}\)The complications are mainly due to the fact that the social and cultural world is constructed by humans in a sense not applicable to the natural world. It is a topic of increasing interest, e.g., Bhaskar [1979], Giddens [1984], Searle [1995], Tuomela [1995], Balzer and Tuomela [1997], to mention a few.
gives rise to the short-term dynamics in the development of scientific knowledge. Moreover, the long-term dynamics is generated by the transformation of proper theories into observation theories, by accepting them as true. This gives rise to a multi-level distinction according to which proper theories may not only explain or predict a lower level observational law, but also be presupposed by a higher level one. This description of the long-term dynamics typically has a theory realist flavor. However, the other positions have their own way of describing such dynamics. In the following brief survey of questions and answers we restrict ourselves to (the ingredients for) the short-term dynamics as seen from the different positions.

**Question 0:** Does a natural world that is independent of human beings exist?

**Question 1:** Can we claim to possess true claims to knowledge about the natural world?

**Question 2:** Can we claim to possess true claims to knowledge about the natural world beyond what is observable?

**Question 3:** Can we claim to possess true claims to knowledge about the natural world beyond (what is observable and) reference claims concerning theoretical terms?

**Question 4:** Does there exist a correct or ideal conceptualization of the natural world?

In the following elucidation, we always presuppose an affirmative answer to the foregoing question. **Question 0,** about the existence of an independent natural world, is not an epistemological question, but a preliminary ontological question. The negative answer leads to ontological idealism, and the positive one to ontological realism. A negative answer to the first epistemological question, **Question 1,** about the possibility of true claims about the natural world, leads to the position of epistemological relativism or skepticism. It has two forms: experiential skepticism, that is, skepticism with respect to claims about sensory and introspective experiences, and inductive skepticism, that is, skepticism merely with respect to inductive extrapolations in the sense of inductive predictions and inductive generalizations. The positive answer to **Question 1** leads to epistemological objectivism or epistemological realism.

**Question 2,** about the possibility of more than observational knowledge, brings us to the heart of the distinction between observation and theoretical terms. A negative answer assumes that the notion of observability is relatively fixed. It indicates observational realism or just empiricism, of which there are two versions. According to instrumentalism, advocated for instance by Schlick [1938] and Toulmin [1953], talking about the reference of theoretical terms does not make sense, let alone talking about true or false (proper) theories. The only function of proper theories is to provide good derivation instruments; that is, they need to enable the derivation of as many true observational consequences as possible and as few false
observational consequences as possible. Hence, the ultimate aim of the instrumentalist is the best derivation instrument, if any. According to the second type of empiricism, called \textit{(constructive) empiricism} by its inventor and main proponent Van Fraassen [1980; 1989], it may make sense in principle to say that theoretical terms have referential value and that proper theories can be true or false. The problem is that we will never know if such is the case beyond reasonable doubt. Hence, what counts is whether such theories are empirically adequate or inadequate or, to use our favorite terminology, whether they are observationally true or false.

A positive answer to \textit{Question 2} amounts to so-called \textit{scientific realism}, according to which proper theories, or at least theoretical terms, have to be taken seriously. Since the books by Hacking [1983] and Cartwright [1983], there is a weaker version of realism than the traditional one, which amounts to a negative answer to \textit{Question 3} on the possibility of more than (observational and) referential knowledge. Primarily thinking of the referentiality of entity terms, they call their position \textit{entity realism}. However, it seems highly plausible to extrapolate that position to attribute referentiality, in some plausible sense, to many types of terms, and speak of \textit{referential realism}.

The positive answer to \textit{Question 3} brings us to so-called \textit{theoretical} or \textit{theory realism}, in some version or another advocated by, for instance, Peirce [1934], Popper [1963], and Niiniluoto [1987; 1999]. Theory realism adds to referential realism that theories are claimed to be true and that we have from time to time good reasons to further assume that they are true, that is, to carry out a theoretical induction.

A positive answer to the last \textit{Question 4}, about the existence of a correct or ideal conceptualization, brings us to a position that is not purely epistemologically built on the positive answer to the preliminary, ontological \textit{Question 0} (i.e., ontological realism). It amounts to an extreme kind of metaphysical realism, which we like to call \textit{essentialistic realism}. According to that view, for instance, there must be natural kinds, not only in some pragmatic or nominal sense, but also in the sense of categories in which entities in the natural world \textit{perfectly} fit. Philosophers of science like Boyd [1984] and Harré [1986] seem to come close to this view.

The negative answer to \textit{Question 4} gives rise to what we call \textit{constructive realism}. It combines theory realism with the view that vocabularies are constructed by a human mind guided by previous results. Of course, one set of terms may be more appropriate than another, in the sense that it produces, perhaps in cooperation with other related vocabularies, more and/or more interesting truths about the domain than the other set of terms does. The fruitfulness of alternative vocabularies will usually be comparable, at least in a practical sense, despite the possibility of fundamental incommensurability. There is however no reason to

\footnote{A kind of antipode of referential realism, and hence of entity realism, arises by denying that referential theoretical claims have truth-values, but lawlike theoretical claims (structural relations) have. This position is known as ‘structural realism’, see [Niiniluoto, this volume; Ladyman, this volume].}
assume that the improvement of vocabularies will ever become impossible.

We summarize the preceding survey in Figure 3.

<table>
<thead>
<tr>
<th>Q0</th>
<th>independent natural world?</th>
<th>⇒</th>
<th>ontological idealism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>yes (\Downarrow) ontological realism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>true claims about the natural world?</td>
<td>⇒</td>
<td>epistemological relativism</td>
</tr>
<tr>
<td></td>
<td>no (\Downarrow) epistemological realism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2</td>
<td>true claims about the natural world beyond the observable?</td>
<td>⇒</td>
<td>empiricism (observational realism)</td>
</tr>
<tr>
<td></td>
<td>no (\Downarrow) instrumentalist</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>yes (\Downarrow) constructive empiricism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>beyond reference?</td>
<td>⇒</td>
<td>referential realism</td>
</tr>
<tr>
<td></td>
<td>no (\Downarrow) entity realism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4</td>
<td>ideal conceptualization?</td>
<td>⇒</td>
<td>constructive realism</td>
</tr>
<tr>
<td></td>
<td>no (\Downarrow) essentialist realism</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The main epistemological positions

The four perspectives, indicated at the beginning of this survey, imply that all (non-relativistic) epistemological positions have an ‘actual world version’ and a ‘nomic world version’. Moreover, they may be restricted to ‘true-or-false’ claims, or emphasize ‘truth approximation claims’. In both cases it is plausible to distinguish between observational, referential, and theoretical claims and corresponding inductions. Instrumentalists, in parallel, speak of theories as ‘reliable-or-unreliable’ derivation instruments or as ‘approaching the best derivation instrument’.

All four perspectives occur, in particular in their realist versions. Standard or traditional realism focuses on ‘true/false’ claims about the actual world. Giere [1985], who introduced the term ‘constructive realism’, focuses on the nomic world, but does not take truth approximation into account. Peirce, Popper and Niniluoto, however, do take truth approximation into account. Moreover, whereas Peirce and Niniluoto primarily focus on the actual version, Popper and Giere seem to have primarily the nomic version in mind, without excluding the actual version. In our view, the nomic version of constructive realism is best suited to scientific practice.
The epistemological positions of instrumentalism, constructive empiricism, referential realism, and constructive realism have been further characterized in [Kuipers, 2000]. Moreover, with the emphasis on their nomic interpretation, they have been compared in the light of the results of the analysis of confirmation, empirical progress and truth approximation in the rest of that book. The (of course, contestable) conclusions reached in that study are encapsulated in the following summary.

There are good reasons for the instrumentalist to become a constructive empiricist; in his turn, in order to give deeper explanations of success differences, the constructive empiricist is forced to become a referential realist; in his turn, there are good reasons for the referential realist to become a theory realist. The theory realist has good reasons to indulge in constructive realism, since there is no reason to assume that there are essences in the world. As a result, the way leads to constructive realism and amounts to a pragmatic argument for this position, where the good reasons mainly deal with the short-term and the long-term dynamics generated by the nature of, and the relations between, confirmation, empirical progress and truth approximation.

Besides these epistemological conclusions, there are some general methodological lessons to be drawn. There appear to be good reasons for all positions not to use the falsificationist but the instrumentalist or ‘evaluation(ist)’ methodology. That is, the selection of theories should exclusively be guided by empirical success, even if the better theory has already been falsified. This common methodology, directed at the separate and comparative evaluation of theories, is extensively presented in [Kuipers, 2000, Chapters 5 and 6; Kuipers, 2001, Chapters 7 and 8].

According to the evaluation methodology, the role of falsifications has to be strongly relativized. This does not at all imply that we dispute Popper’s claim that falsifiable theories are characteristic for empirical science; on the contrary, only falsifiable theories can obtain empirical success. Moreover, instead of denouncing the hypothetico-deductive method, the evaluation methodology amounts to a sophisticated application of that method. As suggested, the evaluation methodology may also be called the instrumentalist methodology, because the suggested methodology is usually associated with the instrumentalist epistemological position. The reason is, of course, that it is quite natural for instrumentalists not to consider a theory to be seriously disqualified by mere falsification. However, since we will argue that the instrumentalist methodology is also very useful for the other positions, we want to terminologically separate it from the instrumentalist epistemological position, by calling the former the evaluation methodology, and the latter ‘instrumentalism’.

We close this subsection with a warning. The suggested hierarchy of the heuristics corresponding to the epistemological positions is, of course, not to be taken in any dogmatic sense. That is, when one is unable to successfully use the constructive realist heuristic, one should not stick to it, but try weaker heuristics: hence first the referential realist, then the empiricist, and finally the instrumentalist heuristic. For, as with other kinds of heuristics, although not everything goes
always, *pace* (the suggestion of) Feyerabend’s slogan “anything goes”, everything goes sometimes. Moreover, after using a weaker heuristic, a stronger heuristic may become applicable at a later stage: “reculer pour mieux sauter”.

2 INTERMEZZO. THE STRUCTURALIST APPROACH TO THEORIES

*Introduction*

This section gives a systematic introduction to the structuralist reconstruction of empirical theories. Although it bridges in a sense the first and the third section, it is an optional intermezzo for readers interested in semi-formal approaches in the philosophy of science (See also [Aliseda and Gillies, this volume]). In Subsection 1.4 we have already presented a first exploration of the nature and structure of empirical theories. Recall that a theory usually has a core of principles and a belt of auxiliary hypotheses. A theory is said to be ontologically stratified when there are two or more kinds of entities involved, where entities of one of these kinds are components of entities of another kind. It is then plausible to speak of a lower micro-level and a higher macro-level. In this case some principles will only concern the micro-entities, and their properties and relations, and are called internal principles, whereas some other principles will connect the different kinds of entities, and their properties and relations, and are called bridge principles. A similar distinction holds for auxiliary hypotheses.

Besides ontological stratification there is epistemological stratification: these frequently go together, but are essentially independent. A proper theory \( X \) was defined as an epistemologically stratified theory in the sense that it contains terms, and hence statements, that are laden with one or more of its principles, that are \( X \)-laden or \( X \)-theoretical, for short. The other terms of \( X \) are called \( X \)-unladen or \( X \)-non-theoretical. In contrast to proper theories, observational hypotheses were defined as improper theories, containing no theoretical terms of their own. A set of connected observational hypotheses was called an observational theory. It should be recalled that being \( X \)-non-theoretical is a theory-relative qualification of a term or a statement: such a term or statement may well be laden with underlying theories.

The main function of a proper theory \( X \) is the explanation and prediction of \( X \)-unladen, observational laws, i.e., true general hypotheses containing neither terms laden with these laws themselves nor terms laden with \( X \). For this purpose the distinction between observational hypotheses and proper theories is of course crucial.

Recall, finally, that the principles of a theory, whether ontologically and/or epistemologically stratified or not, can frequently be distinguished in main or generic principles, claimed to be true for the whole domain concerned, and special principles, only claimed to be true for a certain subdomain.

\(^{17}\)This section profited a lot from Wolfgang Balzer’s detailed criticism.
In this section we will analyze the structure of (improper and proper) theories in more detail. In Subsection 2.1, we will discuss the attractive features of the structuralist approach in general. Starting with the simple example of a slide balance we will first present, in Subsection 2.2., the structuralist representation without making the distinction between theoretical and non-theoretical terms. Reconsidering the slide balance in the light of attempts to measure quantities in a non-circular way, we will introduce, in Subsection 2.3., the structuralist representation with the distinction between theoretical and non-theoretical terms. We will also give the basic outline of this kind of representation for three examples: classical particle mechanics, the periodic table, and psychoanalytic theory. In Subsection 2.4. we will continue with some further refinements of the structuralist approach, viz., the distinction between absolute and relative empirical content, the possibilities of determining the intended applications, relations between theories, theory-nets, and constraints. We will conclude by briefly considering which parts of the structuralist approach are also useful for non-empirical theories, such as metaphysical, mathematical, conceptual and normative theories.

In the previous section, \(X\), \(Y\), and \(Z\) were used as variables for theories, whereas in the structuralist presentation \(M\), \(M'\), \(M^*\) and the like are usual for this purpose, a practice which we also adopt in this section.

### 2.1 Why the structuralist approach?

There are two main ways of viewing the structure of empirical theories. The statement approach conceives theories primarily as sets of statements in a formalized language. In case of an axiomatized theory all these statements are logical consequences of a subset of so-called axioms. This ‘logico-linguistic’ approach has long been considered as the only and obvious approach, e.g., by Carnap and Popper. Kyburg [1968] illustrates the approach in great detail.

In the semantic approach theories are primarily conceived as sets of ‘logico-mathematical’ models. One version, the state-space version, goes back to Beth and is favored for instance by Suppe and Van Fraassen. Our favorite version, the set-theoretic or structuralist approach was introduced by Suppes and refined by Sneed, Stegmüller, Balzer and Moulines. Its basic idea is that theories frequently specify classes of set-theoretic structures satisfying certain conditions. A set-theoretic structure is an ordered set of one or more domain- or base-sets, and one or more properties, relations or functions defined on them, which satisfy certain conditions.

A biological family e.g., can be represented as a structure \(\langle A, C \rangle\) with \(A\) as the set of members of the family and \(C\) as a ternary relation on \(A\). That is, \(C\) is a subset of \(A \times A \times A\), such that \(C(x, y, z)\) states that \(z\) is a child of \(x\) and \(y\). If there are precisely two members \(x\) and \(y\) in \(A\) such that for all other \(z\) in \(A\), \(C(x, y, z)\) is true, then this structure can be used to represent a proper two-generation biological family.

According to the structuralist view an axiomatized theory defines such a class of structures and the conditions imposed on the components of the structures are the
axioms of the theory. The link with reality is made by the claim, associated with the theory, that the set of set-theoretic representations of the so-called intended applications forms a subset of the class of structures of the theory.

Unfortunately, there has been much debate about what the proper approach to theories is\(^\text{18}\), whereas it is easy to see that the two approaches are not at all incompatible. At least for so-called first-order statement theories, i.e., theories formulated as a set of statements of a so-called first-order language this is evident. For the set of models of such a theory, i.e., the structures for which the statements of the theory are true, is precisely a set of structures that might also have been introduced directly in the structuralist way. If we do not restrict ourselves to first-order languages, both approaches are still essentially intertranslatable. Hence, the choice is a pragmatic question.

The main advantage of the structuralist approach is that it is much more a bottom-up approach than the statement approach. It invites us, as it were, to represent and analyze a theory as close to the actual presentations in textbooks as is formally possible. As in scientific practice, all kinds of useful mathematics may be used for that purpose.\(^\text{19}\) It does not mean that structuralist reconstruction of theories is an easy task. However, the statement approach is certainly more difficult for specific reconstructions. It is primarily useful in talking about theories in general and in studying logical, in particular model-theoretic, questions about the relation between sentences and their models.\(^\text{20}\) These questions and their answers become very complicated as soon as substantial mathematics is involved, e.g., real numbers. Happily enough, not all interesting theoretical questions need logical treatment. For example, as we have demonstrated in Kuipers [2000, Chapter 10], theoretical questions concerning, for instance, idealization and concretization as a truth approximation strategy can be treated relatively easily in structuralist terms.

Another advantage of the structuralist approach is the ‘systemic’ perspective that takes the world to consist of many systems. Though the statement view could take up this perspective, it has not done so. However this may be, it is plausible that in this respect the structuralist representation is also much simpler than a statement version, if such a version would be made explicit. A third advantage

\(^{18}\text{See, for example, [Mahner and Bunge, 1997, Subsection 9.3.2.]. In my view, they rightly oppose against some naive aspects of standard structuralist reconstructions of theories. In particular, when structuralists neglect the interpretation of terms and the empirical claims (statements) stating that the so-called intended applications, see below, can be represented as models of the core of the relevant theory. However, they evidently also prefer a ‘logico-mathematical’ rather than a ‘logico-linguistic’ axiomatization of theories, witness Bunge’s [1967, Chapter 3] axiomatization of classical particle mechanics.}\)

\(^{19}\text{This consideration is rather similar to Giere’s [1999] point of view. Enlarging the scope from set-theoretic structures to models of one kind or another, he argues that a ‘representational’ view of models “is much more adequate to the needs of empirical science” than the standard or ‘instantial’ view of models. The latter corresponds to the statement view in the sense that models may or may not be instances making a (set of) statement(s) true.}\)

\(^{20}\text{In fact, there is a great distance between (abstract) model theory and models in the empirical sciences, see [Mahner and Bunge, 1997, Section 3.5].}\)
is that syntactical features, like, for example, the type of a function, can be expressed in a realistic way, talking about functions, and not about building strings of symbols.

Given our preference to be as useful as possible for actual scientific research, we will restrict our attention to the structuralist approach. The best textbooks presenting the structuralist view on theories are Diederich [1981], Balzer [1982], Stégmüller [1973; 1986] and Balzer, Moulines, Sneed [1987]. We will present briefly a number of examples and the main general features, referring to extensive expositions when possible. We will not go into technical details which are not of primary importance for actual practice. Our main goal is to present by way of examples and general exposition the kind of entities that one may be looking for in theory formation and the ways in which standard questions about these entities can be explicated.

2.2 The epistemologically unstratified approach to theories

Starting with the simple example of a slide balance we will first present the structuralist representation of theories and the corresponding terminology without making the epistemological distinction between theoretical and non-theoretical terms.

2.2.1 Example: The slide balance

Consider the slide balance as represented in Figure 4.

On either side there can be placed a finite number of objects of various weights at all possible distances from the turning-point $S$. The balance is assumed to be completely symmetric, the equal arms are as long as necessary, and the objects are point masses, i.e., dimensionless particles. Our domain of interest consists of the equilibrium states, i.e., all possible distributions of objects resulting in equilibrium.

A plausible way to represent the equilibrium states, the intended applications, is as follows. We start by characterizing a possible or potential equilibrium state by a structure of the form $(P, Pl, d, w)$. Here $P$ is the finite set of particles involved and $Pl$ the subset of $P$ such that $Pl$ and $P-Pl$ represent the particles to the left and to the right of $S$, respectively. For every particle $p$ in $P$, $d(p)$ indicates the distance of $p$ from $S$ and $w(p)$ the weight of $p$. Technically speaking, $d$ and $w$ are positive real-valued functions on $P$. Let us call the set of structures $(P, Pl, d, w)$ satisfying
all formal conditions the set of potential equilibrium models of our theory about the equilibrium states of the slide balance, indicated by $SBp$.

Accordingly, our conceptual claim is that the equilibrium states can be represented as members of $SBp$, i.e., there is a subset $E$ of $SBp$ representing the nomic, that is, the nomically possible, equilibrium states: the $SBp$-set of intended applications.

The ultimate purpose of theory formation now is to try to characterize $E$ explicitly by one or more additional conditions. As is well known, the adequate condition in the present case is specified by the so-called law of the balance: the sum of distance times weight of the objects on the left should be equal to that sum involving the objects on the right. Let us call the subset of $SBp$ of members satisfying this condition the set of equilibrium models of our theory, indicated by $SB$. The proper empirical claim of the law of the balance can now be formulated as “$E=SB$”. Assuming some idealizations, e.g., that the objects can be conceived as point-masses, this claim is (generally supposed to be) true. As we will see in other cases the relevant claim need not be as strong as in the present case. The claim might just have been that $E$ is a subset of $SB$.

It will be helpful for later examples to add a more formal presentation of the naïve theory of the slide balance.\footnote{Set theoretic symbols that are used in this and other schemes: $\in$: element of, $\subseteq$: subset of, $\subset$: proper subset of, $\cap$: intersection, $\cup$: union, $\mathbb{N}(+)$: set of (positive) natural numbers, $\mathbb{R}(+)$: set of (positive) real numbers.}

\begin{center}
| The naïve theory of the slide balance $(SBp, SB, D, E)$ contains $\rightarrow$ $(P, Pl, d, w)$ iff |
|---|---|---|---|
| 1. $P$ is a finite set and $Pl$ is a subset of $P$ | particles | $P$ | $Pl$ |
| 2. $d : P \rightarrow \mathbb{R}^+$ | the distance of $p$ from $S$ | $d(p)$ | $d$ |
| 3. $w : P \rightarrow \mathbb{R}^+$ | the weight of $p$ | $w(p)$ | $w$ |
| 4. $\sum_{p \in Pl} d(p) \cdot w(p) = \sum_{p \in P \setminus Pl} d(p) \cdot w(p)$ | the law of the balance | $\Sigma$ | $\Sigma$ |

Concepts and claims

$SBp - SB$ empirical content (to be explained)

$E \subseteq SBp$ conceptual claim: all intended domain of applications $D$ can be represented, by $E$, as potential models, the intended applications

$E \subseteq SB$ (naïve weak) empirical claim: all intended applications are equilibrium models

$E = SB$ (naïve) strong empirical claim: … and vice versa

Later we will see that the empirical claims are still naïve in the sense that it turns out to be impossible to test them in a non-circular way. But first we will present the general structuralistic set-up for unstratified theories.
2.2.2 Unstratified theories

Let there be a given domain $D$ of natural phenomena (states, situations, systems) to be investigated. $D$ is supposed to be circumscribed by some informal description and may be called the intended domain of applications. Although $D$ is a set, its elements are not yet mutually well distinguished. For this reason we do not yet speak of the domain of intended applications.

In order to characterize the phenomena of $D$, a set $Mp$ of conceptual possibilities or potential models is construed. Technically speaking, $Mp$ is a set of structures of a certain type, a so-called similarity type. In practice $Mp$ will be the conceptual frame of a research program (see Section 3) for $D$.

The confrontation of $D$ with $Mp$, i.e., $D$ seen through $Mp$, is assumed to generate a unique, time-independent subset $Mp(D) = def I$ of all $Mp$-representations of the members of $D$, to be called the $Mp$-set of intended applications. Apart from time-independence, this assumption is a conceptual claim. Of course, since nomic impossibilities can, by definition, not be realized, $I$ will be a subset of the ($Mp$-) set of nomic possibilities, but it may be a proper subset, i.e., a more specific set of intended applications satisfying certain additional (more or less precise, but relatively observational) conditions. Assuming that the set of nomic possibilities is a proper subset of $Mp$, i.e., not everything that is conceivable is nomically possible, $I$ is also a proper subset of $Mp$. In certain cases $I$ may be a one-element set, in particular when we want to describe ‘the actual world’ in a certain context, that is, a realized (hence nomic) possibility, e.g., the description of conditions and results of a particular experiment. When dealing with truth approximation [Kuipers, 2000], the attention is focussed on the special case that $I$ is the set of nomic possibilities.

A specific theory about $D$ is concentrated around an explicitly defined subset $M$ of $Mp$, the models of the theory. More specifically, a specific unstratified theory is any combination of the form $UT = \langle Mp, M, D, I \rangle$ with, beside the conceptual claims that $M$ and $I$ are both subsets of $Mp$, the (weak) empirical claim that $I$ is a subset of $M$. Sometimes the strong empirical claim is made that $I$ is equal to $M$, but here we take the weak claim as standard. It is plausible to call $UT$ true when its claim is true, and false otherwise.

The general set-up of the structure of epistemologically unstratified theories will now be presented in a scheme. Such a theory is a meta-structure of the following form:

$\langle Mp, M \rangle$ is sometimes called the theoretical core of the theory, and $\langle D, I \rangle$ may be called the application target of the theory.

The unstratified set-up of theories seems to be rather adequate for observational theories, recall, a combination of one or more observational hypotheses, which contain by definition only terms that are understood independently of the theory concerned.
\( \langle M_p, M, D, I \rangle \) is an epistemologically unstratified theory iff

- \( M_p \): potential models: a set of structures of a certain type
- \( M \subseteq M_p \): models: the potential models that satisfy all axioms
- \( M_p - M \): empirical content (to be explained)
- \( D \): the intended domain of applications
- \( I \subseteq M_p \): intended applications, resulting from the conceptual claim that \( D \) can be represented as a set of members of \( M_p \), i.e. \( "I = M_p(D)" \)

\( I \subseteq M \) (weak) empirical claim

\( I = M \) (strong) empirical claim

### 2.2.3 Basic terminology

Before we go over to stratified theories, we would like to present some useful basic terminology, which can largely be seen as a structuralist explication of Popperian ‘statement terminology’ [Popper, 1934/1959]. We will neglect all necessary provisos, in particular in regard to the complications arising from underlying theories. To use Lakatos’s term [Lakatos, 1978], we explicate naive falsificationism, first unstratified, later stratified.

When the claim of theory \( UT = \langle M_p, M, D, I \rangle \) is false \( I - M \) is by definition non-empty, in which case it is plausible to call its members instantial mistakes or (empirical) counter-examples of \( UT \). Note that being a counter-example in this sense does not imply that it has been realized already and registered as such. The set of counter-examples \( I - M \) is by definition a subset of \( M_p - M \). Hence, \( I - M \) can, whatever \( I \) is, only be non-empty when \( M_p - M \) is non-empty. In other words, the members of \( M_p - M \) may be called the potential counter-examples of the theory and, as has already been stated, the set \( M_p - M \) itself the empirical content of \( UT \). From the present point of view, Popper had similar things in mind with his notions of ‘potential falsifier’ and ‘empirical content’.

Other plausible explications of Popperian terminology (which will however not be used in the sequel) are for instance: \( UT \) is falsifiable (or empirical) if and only if \( M_p - M \) is non-empty, and \( UT^* \) is better falsifiable than \( UT \) when \( M_p - M \) is a proper subset of \( M_p - M^* \). The latter condition is equivalent to: \( M^* \) is a proper subset of \( M \). In its turn, this is equivalent to stating that the claim of \( UT^* \) implies that of \( UT \), and not conversely, that is, \( UT^* \) is stronger than \( UT \).

The well-known verification/falsification asymmetry also arises naturally in the present set-up. To verify theory \( UT \) it would be necessary to show that all members of \( I \), that is, all \( M_p \)-representations of \( D \), belong to \( M \). In interesting cases, this demonstration will always be an infinite task, even in the case that \( I \) is finite, for the task is only finite when \( D \) is finite. To falsify \( UT \), however, it is ‘only’ necessary to show that there is at least one member of \( I \) not belonging to \( M \). Hence, if a theory is true, verification will nevertheless not be obtainable if \( D \) is infinite. On the other hand, when a theory is false, falsification is attainable in principle,
viz., by realizing one counter-example. If an attempt to falsify fails in such a way that the experiment provides an (empirical) example of UT, i.e., a member of \(M\), this is called confirmation (or corroboration) of UT.

In the present set-up Popper’s distinction between universal and existential statements gets an adapted interpretation. Here it becomes the distinction between the general claim of the theory \((I \subseteq M)\) that all intended applications are models of the theory, and the negation of this claim, the existential claim that at least one intended application is not a model \((I - M\) is non-empty).

A basic statement (see Subsection 1.3.1 for Popper’s specific idea of a basic statement) becomes a claim to the effect that a certain intended application \(x\) in \(I\) belongs to a certain subset \(F\) of \(M_p\), defined by a certain condition being imposed on potential models, i.e., \(x \in I \cap F\). An accepted basic statement presupposes of course that the relevant intended application has been realized.

The basic statement \(x \in I \cap F\) is in conflict with theory UT if it can be demonstrated on conceptual grounds that \(F \cap M\) is empty. Such basic statements may be seen as a more direct explication of Popper’s idea of ‘potential falsifiers’, compared to ‘potential counter-examples’. However, it is easy to show that the suggested statement concept of potential falsifier has become essentially redundant. It is readily verified that the discovery of a true potential falsifier of UT, i.e., the discovery of a \(x\) in \(M_p\) for which \(x \in I \cap F\) is true, implies that \(I - M\) is non-empty and hence that \(x\) is a counter-example of UT. Conversely, the existence of counter-examples of UT is easily seen to imply that there must be true potential falsifiers. As a consequence, empirically demonstrating the existence of a counter-example, i.e., realizing a potential counter-example, goes hand in hand with demonstrating that there is a true potential falsifier. Hence, the statement concept of potential falsifier is not needed in the face of the concept of a (potential) counter-example.

2.3 The stratified approach to theories

Starting by reconsidering the slide balance in the light of attempts to measure the relevant quantities in a non-circular way, we will introduce the structuralist representation of (prima facie proper) theories and the corresponding terminology with the epistemological distinction between theoretical and non-theoretical terms. We will give the basic outline of this kind of representation for three examples: classical particle mechanics, the periodic table, and psychoanalytic theory.

2.3.1 The slide balance reconsidered

The problem with the slide balance is that it might be impossible to test the claim in a non-circular way without leading to an infinite regression. For, to test the claim, it is necessary to measure the distances and the weights. Whereas distance measuring does not require something like a slide balance, weight measuring may not only actually be done by using a slide balance, there might even be no other possibility. If the weight of a particle is measured by a slide balance the law of the balance is obviously presupposed. Hence, assuming that the weight of a particle
can only be measured by a slide balance, the concept of weight is SB-theoretical and leads to the so-called problem of theoretical terms. A test of the claim of the theory would presuppose that the weights of the particles have been measured before with the same or another slide balance. Hence, we get either circular testing or an infinite regress, if we stick to “I ⊆ M” as the empirical claim of the theory. There is, however, a way-out of this dilemma, by restricting the empirical claim to SB-non-theoretical terms. Later we will see that for two reasons the situation in the present example is not as dramatic as suggested, but this did not exclude the fact that the example could, by way of a thought experiment, be transformed into an instructive example of genuine theoretical terms.

In order to formulate a new empirical claim we introduce the set of potential partial equilibrium models \( SBpp \), being the structures of \( SBp \) without the SB-theoretical weight component and the corresponding ‘status-condition’, viz., clause 3). By consequence, there is a restriction or projection function \( \pi \) from \( SBp \) onto \( SBpp \) projecting every potential model on the potential partial model arising from deleting \( w \) and clause 3). Hence, for \( x = \langle P, Pl, d, w \rangle \in SBp \), the projection of \( x \), \( \pi(x) \), is equal to \( \langle P, Pl, d \rangle \in SBpp \). For an arbitrary subset \( X \) of \( SBp \), \( \pi X \), the projection of \( X \), is defined as the subset of \( SBpp \) containing precisely the projections of the members of \( X \).

For stratified theories we assume that the set of intended applications \( E \) no longer represents the equilibrium states seen through \( SBp \), but seen through \( SBpp \). Hence, the corresponding conceptual claim that \( E \) is a subset of \( SBpp \) is not laden with our theory about \( SB \).

Now it is also easy to see that the claim that \( E \) is a subset of \( \pi SB \) is not laden with the weight term, in the sense that it does not presuppose that the weights of the particles have been empirically determined. Hence, this revised empirical claim can be tested in a non-circular way.

It is plausible to call the members of \( E-\pi SB \), if any, counter-examples of the theory. It is clear that they have to come from \( SBpp-\pi SB \). Hence, it is now plausible to call this set the empirical content and its members potential counter-examples. Note that the empirical content reduces to the empirical content of the unstratified theory (\( SBp-SB \)) when \( SBpp \) and \( SBp \) are identical and \( \pi \) is, by consequence, the identity-function.

Unfortunately, the new claim is not only non-circular, it is also vacuous, for the empirical content is empty. The claim says in fact that all intended applications can be extended to models of the theory. To be precise, the claim is that every \( \langle P, Pl, d \rangle \in E \) can be supplied with a positive real-valued function \( w \) on \( P \) such that \( \langle P, Pl, d, w \rangle \) is in \( SB \). But it is easy to check that this is possible for every member of \( SBpp \). In other words the empirical content \( SBpp-\pi SB \) is empty.

However, the situation changes when we take so-called constraints into consideration: in the present case we have to require also that the weights assigned to the same particle, occurring in different applications should be the same. In contrast to the distance of the objects from the turning point \( S \), our concept of weight is such that the weight of particles is constant in different applications. The formal
treatment of constraints, however, will be postponed to Subsection 2.4.4.

Before we return to the general exposition we will summarize the formal features
of the theory of the slide balance, leaving out the plausible specification of the
projection function $\pi$:

### The refined theory of the slide balance \( \langle SBp, SBpp, SB, \pi, D, E \rangle \)

<table>
<thead>
<tr>
<th>SBpp</th>
<th>contains $\rightarrow$ $\uparrow$ contains $\rightarrow$ $\uparrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>( SBpp )</td>
<td>( (P, P_l, d, w) ) iff ( (P, P_l, d) = \pi(P, P_l, d, w) ) iff</td>
</tr>
<tr>
<td>1. ( P ) is a finite set and ( P_l ) is a subset of ( P )</td>
<td>particles the particles on the left of ( S )</td>
</tr>
<tr>
<td>2. ( d : P \rightarrow \mathbb{R}^+ ) ( d(p) ): the distance of ( p ) from ( S )</td>
<td></td>
</tr>
</tbody>
</table>

| SBp | \( 3 \ w : P \rightarrow \mathbb{R}^+ \) \( w(p) \): the weight of \( p \) |

| SB | \( 4 \sum_{p \in P_l} d(p).w(p) = \sum_{p \in P - P_l} d(p).w(p) \) \textit{the law of the balance} |

### Concepts and claims

- **SBpp - $\pi$SB** \textit{empirical content} without \( w \)-constraint empty, with \( w \)-constraint non-empty
- **$E \subseteq SBpp$** \textit{conceptual claim}: all intended domain of applications \( D \) can be represented, by \( E \), as potential partial models, the intended applications
- **$E \subseteq \pi$SB** \textit{(weak) empirical claim}: all intended applications can be extended to models
- **$E = \pi$SB** \textit{strong empirical claim}: \ldots and vice versa

By way of digression, it is interesting to note that, assuming the weight-constraint, the \( SB \)-theory explains the following observational, i.e., \( SB \)-unladen, \textit{factor slide law}: if, starting from an equilibrium, the distances of all objects are multiplied by the same factor, there is again equilibrium. For it follows trivially from the law of the balance.

As a matter of fact, in the present case it is not difficult to formulate an observational law such that the notion of weight can be explicitly defined, apart from a proportionality constant, on its basis. The law referred to states the following: given a unit object at a unit distance at one side of \( S \), every other object \( p \) has a ‘unique equilibrium distance’ \( d_u(p) \) at the other side. The weight \( w(p) \) is then defined as \( 1/d_u(p) \), hence, such that in the relevant cases the law of the balance is satisfied by definition. Consequently, for these cases the law cannot be tested in a non-circular way. But there is no regress, let alone infinite regress. For, given the definition, the rest of the law of the balance is a straightforward empirical claim that can be directly tested.

As a consequence, the theory of the slide balance does not, on closer inspection, give rise to the problem of theoretical terms, when certain observational laws are taken into consideration. Of course, this does not affect the instructiveness of the \( SB \)-theory as an almost proper theory. Moreover, it illustrates an interesting way
in which a seemingly proper theory may on closer inspection be a sophisticatedly formulated observational theory, in the present case: the conjunction of the ‘unique equilibrium distance law’, the weight-definition on its basis, and the law of the balance.

There is still one other reason why the problem of theoretical terms is not so dramatic in the case of the slide balance: there are other ways of measuring the weight of objects than by using a slide balance. But let us now turn to the general set up of stratified theories, designed for proper theories.

2.3.2 Stratified theories

The general set-up of the structure of epistemologically stratified theories can now directly be presented in a scheme. Such a theory is a meta-structure of the following form:

\[
\langle M_p, M_{pp}, M, \pi, D, I \rangle
\]

is an epistemologically stratified theory iff

- \( M_p \) potential models: a set of structures of a certain type
- \( M_{pp} \) potential partial models: the substructures of \( M_p \) restricted to non-theoretical components
- \( M \subseteq M_p \) models: the potential models that satisfy all axioms
- \( \pi: M_p \rightarrow M_{pp} \) the projection function (from \( M_p \) onto \( M_{pp} \))
- \( \pi X = \{ \pi(x) | x \in X \} \), for \( X \subseteq M_p \), implying \( \pi X \subseteq M_{pp} \)
- \( \pi M \) projected models
- \( M_{pp} - \pi M \) empirical content
- \( D \) the intended domain of applications
- \( I \subseteq M_{pp} \) intended applications (non-theoretical), resulting from the conceptual claim that \( D \) can be represented as a set of members of \( M_{pp} \), i.e. “\( I = M_{pp}(D) \)”
- \( I \subseteq \pi M \) (weak) empirical claim
- \( I = \pi M \) strong empirical claim

Now it is plausible to call \( \langle M_p, M_{pp}, M, \pi \rangle \) the theoretical core of the theory and \( \langle D, I \rangle \) remains the application target. Figure 5 illustrates the refined empirical claim: the shaded area, representing \( I-\pi M \), should be empty. To be precise, \( I-\pi M \) should be empty on conceptual grounds, that is, the conceptual characterization of \( I \) and \( \pi M \) should not leave room for conceptual possibilities in \( I-\pi M \) (let alone for actual intended applications).

2.3.3 Examples

In this subsection we will give the theoretical core of the structuralist reconstruction of three well-known theories, viz., Newton’s classical (gravitational) particle mechanics, Mendeleev’s and the refined theory of the periodic table of chemical elements, and Freud’s psycho-analytic theory. The presentation will always start
with the representation in a table followed by a brief elucidation. For details of the theory and the reconstruction, the reader is referred to the original or other publications of the reconstructions. The theories (more precisely, the theory cores) will be named by their basic class of models.

From the fact that Freud’s theory can be reconstructed in the structuralistic way it follows that this way of reconstruction is, like the statement approach, applicable to qualitative, non-mathematical theories. From the other examples, it is evident that the present approach is also well suited for quantitative theories, a kind of theory for which the statement approach leads to all kinds of complications.

In a sense it is a trivial claim that every empirical theory can be reconstructed in structuralist fashion. Hence, there should be additional reasons to do so in particular cases. A general reason frequently is the desire to get a better insight into the theory; besides that, one may be interested in particular questions, such as whether the theory has empirical content, whether it is an observational or a proper theory, what its precise relation is to another theory, etc. The examples to be presented are supplied with some comments to illustrate both reasons of reconstruction. But the main function of getting acquainted with the structuralist approach in general and by way of examples is of course the heuristic role it may play in the construction of new theories.

After the presentation of the three examples we will continue in the next subsection with general matters, such as the distinction between absolute and relative empirical content, the possibilities of determining of the intended applications, relations between theories, theory-nets, and constraints.

Classical particle mechanics

As is well known, Newton’s theory of gravitation is based on the generic theory of particle motion, i.e., classical particle mechanics (CPM). The core of this theory is formed by three interrelated so-called laws of motion: the first law: the law of inertia, the second law: \( F = m.a \), and third law: action is minus reaction. This general or generic theory can be specialized by adding the special law of
**Classical particle mechanics for one dimension** (with gravitation as specialization)

\[ CPM = (CPM_p, CPM_{pp}, \pi, CPM, GCPM) \]

<table>
<thead>
<tr>
<th>contains (\uparrow)</th>
<th>({P, T, s, m, f}) iff</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\uparrow) contains</td>
<td>({P, T, s} = \pi(P, T, s, m, f)) iff</td>
</tr>
<tr>
<td>1) (P) is a finite set</td>
<td>particles</td>
</tr>
<tr>
<td>2) (T) is real interval</td>
<td>time-interval</td>
</tr>
<tr>
<td>3) (s : P \times T \to \mathbb{R})</td>
<td>position</td>
</tr>
<tr>
<td>giving rise to 1(^{st}) and 2(^{nd}) time derivatives:</td>
<td></td>
</tr>
<tr>
<td>(v : P \times T \to \mathbb{R})</td>
<td>velocity</td>
</tr>
<tr>
<td>(a : P \times T \to \mathbb{R})</td>
<td>acceleration</td>
</tr>
<tr>
<td>4) (m : P \to \mathbb{R})</td>
<td>mass</td>
</tr>
<tr>
<td>5) (f : P \times T \times P \to \mathbb{R})</td>
<td>force</td>
</tr>
<tr>
<td>(f(p, t, q))</td>
<td>force from (q) on (p) at (t)</td>
</tr>
<tr>
<td>6) second law (implying the first law in this formulation):</td>
<td></td>
</tr>
<tr>
<td>for all (p) in (P) and (t) in (T) (\Sigma_{q \in P} f(p, t, q) = m(p) \cdot a(p, t))</td>
<td></td>
</tr>
<tr>
<td>7) third law (action = - reaction):</td>
<td></td>
</tr>
<tr>
<td>for all (p) and (q) in (P) and all (t) in (T) (f(p, t, q) = - f(q, t, p))</td>
<td></td>
</tr>
<tr>
<td><strong>CPM</strong></td>
<td></td>
</tr>
<tr>
<td><strong>CPM_{pp}</strong></td>
<td></td>
</tr>
<tr>
<td><strong>GCPM</strong></td>
<td></td>
</tr>
<tr>
<td><strong>πCPM</strong> and <strong>πGCPM</strong> provide the projected models of <strong>CPM</strong> and <strong>GCPM</strong>, respectively. <strong>CPM_{pp}-πCPM</strong> and <strong>CPM_{pp}-πGCPM</strong> constitute the empirical content of <strong>CPM</strong> and <strong>GCPM</strong>, respectively. Note that the former is a subset of the latter, just as it should be, for <strong>GCPM</strong> is stronger than <strong>CPM</strong>.</td>
<td></td>
</tr>
<tr>
<td>As long as the identity constraint for mass is not taken into consideration <strong>GCPM</strong> has no empirical content, let alone <strong>CPM</strong>. With the mass constraint <strong>CPM</strong> still lacks empirical content, but <strong>GCPM</strong> gets it.</td>
<td></td>
</tr>
<tr>
<td>The intended domain of applications of <strong>CPM</strong> concerns in the first place that of <strong>GCPM</strong>, for instance, planetary orbits, falling stones, paths of projectiles, etc.,</td>
<td></td>
</tr>
</tbody>
</table>
but also movement of objects by spring or electric forces. Moreover, it contains compound applications, i.e., applications in which two or more force types operate, e.g., three in the case of an electrically charged ball on an isolated vertical spring on a charged table.

Periodic table

<table>
<thead>
<tr>
<th>Periodic table of chemical elements (naive and refined)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPT/SPT = (PTpp, PTp, π, NPT/SPT)</td>
</tr>
<tr>
<td>contains → (E, m, ≈, z) iff</td>
</tr>
<tr>
<td>↑ contains → (E, m, ≈) = π(E, m, ≈) iff</td>
</tr>
<tr>
<td>1) E: a finite set</td>
</tr>
<tr>
<td>2) m: E → R</td>
</tr>
<tr>
<td>3) ≈: equivalence relation on E</td>
</tr>
<tr>
<td>PTpp</td>
</tr>
<tr>
<td>4) z: E → R</td>
</tr>
<tr>
<td>NPT</td>
</tr>
<tr>
<td>6N) naive periodic law</td>
</tr>
<tr>
<td>e ≈ e' iff</td>
</tr>
<tr>
<td>RPT</td>
</tr>
<tr>
<td>6R) refined periodic law, elegant, but complicated: core:</td>
</tr>
<tr>
<td>if e ≈ e' and if there is no element with z-number between</td>
</tr>
<tr>
<td>z(e) and z(e')</td>
</tr>
<tr>
<td>then</td>
</tr>
<tr>
<td>i.e., 2 or 8 or 18 or 32 etc.</td>
</tr>
</tbody>
</table>

For a detailed exposition the reader is referred to Hettema and Kuipers [1988; 2000]. The following remarks highlight some crucial points.

As is well known, Mendeleev developed the periodic table on the basis of the observation that the chemical elements can be classified in groups of elements with chemically similar behavior. Moreover, he noted that the ordering of the elements by increasing atomic mass roughly leads to a matrix in which the groups appear as columns. To explain the system in this matrix he introduced the concept of atomic number and formulated the (naive) periodic law (NPT), which was later refined by others (RPT).

In the present example the intended domain of application concerns the chemical elements taken together, such that the conceptual claim states that this domain can be represented by just one potential projected model, say (E*, m*, ≈*).

---

22 For a critical discussion of the main historical and reductive claims in [Hettema and Kuipers, 1988] see [Scerri, 1997]. For a continued discussion, see [Hettema and Kuipers, 2000; Scerri, 2005; Kuipers, 2005].
Mendeleev’s empirical claim was that this pp-model belongs to $\pi NPT$ and the modern empirical claim localizes it in $\pi RPT$. Or, equivalently, there is $z*$ such that $\langle E^*, m^*, \approx^*, z^* \rangle$ belongs to $NPT$ and $RPT$, respectively.

It is not difficult to verify that both theories have empirical content. In fact both claims are false. To fulfill the claims as much as possible, we must allow counter-examples to the three technical conditions imposed by clause 5). They amount to, using plausible names: 5a) missing elements, which may be discovered later, and some have been, 5b) order disturbers, having greater mass and lower atomic number than others or vice versa, and 5c) isotopes, i.e., different elements with the same atomic number.

Note that the notion of a counter-example is used here on a lower level than in the general set-up. This is possible because there is only one overall intended application, viz., $\langle E^*, m^*, \approx^* \rangle$. If that does not fit into $\pi NPT$ or $\pi RPT$, this failure must be due to lower level counter-examples, i.e., specific elements. There may be systematic or just local counter-examples. In this sense $NPT$ has both types of counter-examples, whereas $RPT$ has only local counter-examples.

The history of $PT$ provides marvelous examples of all four combinations of theory (un)laden and theory (un)guided observation, as described in Subsection 1.4.2. A successful search for missing links, for instance, means theory guided but theory unladen observation.

The quantum mechanical theory of the atom provides a reductive explanation, see [Kuipers, 2001, Chapter 3] for $RPT$, by means of identification of $z$ with the number of electrons of the atom concerned. In view of the fact that this number can be measured in $RPT$-independent ways, $RPT$ is in fact an observational theory. Of course, Mendeleev’s $NPT$ was a proper theory, with $z$ as a theoretical term. $RPT$ was developed hand in hand with atomic theory, in which process it transformed from a proper theory into an observational theory.

\textit{Psychoanalytic theory}

Presenting the structure of Freud’s theory does of course not mean that we uncritically subscribe to that theory. One may even denounce that theory as totally out of date, and still be interested in its structure. Compare interest in the structure of the phlogiston theory. For a detailed exposition the reader is referred to Balzer [1982], Stegmüller [1986] or, for the most refined one, to Balzer and Marcou [1989]. The general psychoanalytic theory $PA$ is intended for all human beings, they are all supposed to repress negative experiences ($PA$-11) and to satisfy the main axiom $PA$-10 that all unconscious impulses are sooner or later realized. Note that $PA$-11 does not use theoretical terms, so it makes sense, as indicated, to introduce (non-theoretical or observational or) partial models ($PA$part) as potential partial models satisfying this non-theoretical but substantial axiom. In the next subsection we will generalize this idea and investigate its consequences.

The psychoanalytic theory of neurosis $PAN$ is a specialization to people with a neurosis generating experience, as implicitly defined by $PA$-13. It will be illumi-
### Psychoanalytic theory

\[ PA = \langle P_{App}, P_{Ap}, \pi, P_{Apart}, PA \rangle \]

$\uparrow$ contains $\rightarrow$

<table>
<thead>
<tr>
<th>Contains $\rightarrow$</th>
<th>$\uparrow$ Contains $\rightarrow$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\langle T, E, L, \leq, ASS, B, N, A, U, REAL \rangle$ iff</td>
<td></td>
</tr>
<tr>
<td>$\langle T, E, L, \leq, ASS, B, N \rangle = \pi(\ldots A, U, REAL) \iff$</td>
<td></td>
</tr>
<tr>
<td>1) $T$ is an interval of real numbers; variable $t, t^*$, etc.</td>
<td></td>
</tr>
<tr>
<td>2) $E$ is a non-empty set</td>
<td></td>
</tr>
<tr>
<td>3) $L$ is a proper subset of $E$</td>
<td></td>
</tr>
<tr>
<td>4) $\leq$ is a weak linear ordering on $T$ not later than $\leq$ by definition $\leq$ and $\neq$ earlier than</td>
<td></td>
</tr>
<tr>
<td>5) $B(t)$ is a non-empty subset of $E$ consciousness at time $t$</td>
<td></td>
</tr>
<tr>
<td>6) $N(t)$ is a subset of $B(t)$ and $L$ negative experiences at $t$</td>
<td></td>
</tr>
<tr>
<td>7) $ASS$ is a relation on $E$ associated experiences</td>
<td></td>
</tr>
<tr>
<td>8) $A$ is a non-empty set; $A \cap E = \emptyset$ unconscious impulses</td>
<td></td>
</tr>
<tr>
<td>9) $U(t)$ is a non-empty subset of $A$ unconsciousness at $t$</td>
<td></td>
</tr>
<tr>
<td>10) $REAL$ is ternary relation on $E \times A \times T$: $REAL(e, a, t)$</td>
<td></td>
</tr>
<tr>
<td>- if $REAL(e, a, t)$ then $e$ in $B(t)$</td>
<td></td>
</tr>
<tr>
<td>and $a$ in $U(t)$</td>
<td></td>
</tr>
<tr>
<td>- not for all $t, e$ in $B(t)$ and $a$ in $U(t)$: $REAL(e, a, t)$</td>
<td></td>
</tr>
<tr>
<td>- if $REAL(e, a, t)$ and $REAL(e', a, t')$</td>
<td></td>
</tr>
<tr>
<td>then $ASS(e, e')$</td>
<td></td>
</tr>
<tr>
<td>11) $P_{App}$ repression axiom: repression of negative experiences, incl. associated ones:</td>
<td></td>
</tr>
<tr>
<td>if $e$ in $N(t)$ and $ASS(e, e')$ and $t &lt; t^<em>$ then $e'$ not in $B(t^</em>)$</td>
<td></td>
</tr>
<tr>
<td>12) $P_{App}$ + main axiom: every unconscious impulse is realized sooner or later: for all $t$ and for all $a$ in $U(t)$</td>
<td></td>
</tr>
<tr>
<td>there are $e$ in $E$ and $t^<em>$ such that $t &lt; t^</em>$ and $REAL(e, a, t^*)$</td>
<td></td>
</tr>
<tr>
<td>13) $P_{App}$ + neurosis axiom: having a neurosis generating experience for an impulse: there are $t_0, e_0$ in $E$, $a_0$ in $A$ such that</td>
<td></td>
</tr>
<tr>
<td>$REAL(e_0, a_0, t_0)$ and $e_0$ in $N(t_0)$</td>
<td></td>
</tr>
</tbody>
</table>
nating to define some additional notions:

‘$e_0$ is repressed after $t_0$’ iff 
\[ e_0 \text{ is in } B(t_0) \text{ and for all } t > t_0 \text{ } e_0 \text{ is not in } B(t) \]

‘being neurotic with respect to $a_0$ after $t_0$’ iff 
\[ \text{for all } t > t_0 \text{ there is no } e \text{ in } E \text{ such that } \text{REAL}(a_0, e, t) \]

Now it is easy to prove the following

*Theorem:* if one has had at $t_0$ a neurosis generating experience $e_0$ with respect to impulse $a_0$, $e_0$ is repressed after $t_0$ and one is neurotic with respect to $a_0$ after $t_0$.

Note that ‘being neurotic with respect to $a_0$ after $t_0$’ is formally almost in conflict with the main axiom, but the point is that the realization of $a_0$ required by the main axiom has already taken place at $t_0$, in particular as a result of a negative experience.

A serious problem in the present formulation is what happens when $a_0$ recurs in the unconsciousness, because the main axiom requires repeated realization. But here we will not deal with the necessary refinements for this and other reasons. We will just mention one other example of a further refinement of the theory of neurosis, which can be obtained by integrating it with another specialization of the general theory, viz., the theory of sublimation.

We conclude this section with references to some further examples: classical and relativistic collision mechanics [Balzer, Sneed and Moulines, 1987]; Lagrangian mechanics [Balzer, Sneed and Moulines, 1987]; special relativity theory [Balzer, 1982]; old quantum theory ([Hettema and Kuipers, 1995], see also [Kuipers, 2000, Chapter 11]); simple equilibrium thermodynamics [Balzer, Sneed and Moulines, 1987]; Daltonian stoichiometry [Balzer, Sneed and Moulines, 1987]; modern genetics [Balzer and Dawe, 1986; 1997]; Jeffrey’s theory of decisions [Stegmüller, 1986]; the Arrow-Debreu theory of individual and collective demand [Janssen and Kuipers, 1989], capital structure theory ([Cools, Hamminga, Kuipers, 1994], see also [Kuipers, 2000, Chapter 11]); folk psychology and connectionism [Bickle, 1993; 1998]; Jakobson’s theory of literature has also been reconstructed [Stegmüller, 1986]. Several psychological theories are reconstructed in [Westmeyer 1989, 1992]. Finally, [Balzer, Sneed and Moulines, 2000] contains a representative sample of those above, as well as other ones.

Of course, structuralist representation of a theory may not be necessary for the purposes at hand. However, for detailed questions, such as “Does a certain theory have empirical content?” such a representation is almost unavoidable. In the next subsection we will introduce a number refinements of the structuralist approach that enable us to answer such refined questions.
2.4 Refinements

Now we will continue with some further refinements of the structuralist approach, viz., the distinction between absolute and relative empirical content, the possibilities of determination of the intended applications, relations between theories, theory-nets, and constraints.

2.4.1 Absolute and relative empirical content

It is always possible to divide the axioms into, on the one hand analytic (A) and synthetic or substantial (S) axioms and, on the other, non-theoretical (N) and theoretical (T) ones. As a result, there are four types of axioms: NA, TA, NS, and TS. We do not mean to suggest that the two distinctions are unproblematic. We have discussed extensively in Section 1 and the previous subsections how the N/T-distinction can be made. The A/S-distinction is at least as notorious, and it is undoubtedly partly a matter of conventional decision where the boundary is drawn. See Niiniluoto [1999, Chapter 5] and Sober [2000] for lucid accounts, against Quine’s well-known challenges, of the tenability of this distinction. However, here, and with respect to the N/T-distinction, it is advisable in case of doubt to choose the cautious classifications, i.e., S and T, respectively.

The following survey of sets and names of their elements will now speak for itself.

<table>
<thead>
<tr>
<th>Types of models in relation to types of axioms</th>
<th>A: analytic, or S: synthetic; N: non-theoretical, or T: theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Mpp}$</td>
<td>potential partial models</td>
</tr>
<tr>
<td>$\text{Mp}$</td>
<td>potential models</td>
</tr>
<tr>
<td>$\text{Mpart}$</td>
<td>partial models</td>
</tr>
<tr>
<td>$\text{M}$</td>
<td>models</td>
</tr>
</tbody>
</table>

It is clear that $\langle \text{Mpp}, \text{Mpart} \rangle$ is the (theoretical) core of a partial theory, i.e., an unstratified, hence observational theory, constituting a substantial part of the full theory. The empirical content of the full theory was defined as $\text{Mpp-}\pi\text{M}$, let us call it more specifically the (total or) absolute empirical content (AEC). The empirical content of the partial theory, the partial empirical content (PEC), is of course $\text{Mpp-Mpart}$. Given the trivial fact that $\pi\text{M}$ is a subset of $\text{Mpart}$, the partial empirical content is automatically a subset of the absolute empirical content. The interesting question is whether the full theory has something to add to the partial theory, i.e., whether the (extra or) relative empirical content (REC), defined as $\text{Mpart-}\pi\text{M}$, is non-empty. Figure 6 depicts the three kinds of content.

It is easy to check that the absolute empirical content ($\text{Mpp-}\pi\text{M}$) is the union of the partial empirical content ($\text{Mpp-Mpart}$) and the relative empirical content ($\text{Mpart-}\pi\text{M}$). In consequence, if a theory has relative (and/or partial) empirical content it has absolute empirical content.
Conversely, however, a theory may have absolute empirical content without having relative empirical content, in which case the absolute empirical content coincides with the partial empirical content.

Balzer [1982] claimed that the general psychoanalytic theory has partial, and hence absolute, but no relative empirical content. Stegmüller [1986], however, is able to prove that it also has relative empirical content. Stegmüller then continues with the interesting observations that for this proof it is not necessary to take constraints and/or special laws into consideration and that, as we already noted, classical particle mechanics CPM has only (relative) empirical content when constraints and special laws are taken into consideration. Hence, according to the relative content criterion for empirical impact Freud’s theory is in a sense even superior to that of Newton.

But we would like to add that the CPM example makes it clear that a generic theory (including constraints) need not have relative empirical content in order to be useful. The important research question is whether a generic theory can be supplemented with special laws (leading to specializations, see below) which have relative empirical content (Cf. [Bunge, 1977]).

It is interesting to note that the status of utility theory (Subsection 1.1) is to some extent comparable to that of classical particle mechanics. The generic versions of both theories have no empirical content. However, although it is quite clear that the latter theory can be specialized so as to have empirical content, this is not beyond dispute for the former.

2.4.2 Intended applications reconsidered

In this subsection we will begin by making a few general remarks about the set of intended applications $I$, then we will formulate three different ways of determining $I$, and conclude by elaborating, to a certain extent, the problem of theoretical terms.

$I$ was introduced as ‘$D$ seen through $Mp$’ and later revised as ‘$D$ seen through $Mpp$’, i.e., $I$ represents the intended domain of applications with the conceptual means of $Mpp$. We will restrict our formulations to the refined case, when not otherwise stated, for it includes the extreme case that $Mp=Mpp$.

It is evident that $I$ is $Mpp$-dependent, and $Mpp$ is manmade. Hence we subscribe
to a fundamental form of conceptual relativity. But this need not imply an extreme form of relativism: empirical claims are objectively true or false, for their truth or falsehood depends on nature, assuming that they have empirical content.

In its turn, the objective character of empirical claims does not imply that $D$, $Mpp$ (and hence $I$) and $Mp$ are fixed beforehand, and that the task remains to formulate a subset $M$ of $Mp$ leading to a true empirical claim. As a matter of fact, in practice, the determination of $D$, $Mpp$, $Mp$ and $M$ is a complicated dialectical interaction process, guided by the desire to formulate informative and true empirical claims. Unfortunately, it seems difficult to discern general patterns in this interaction process, without making some important idealizations.

However, if we assume, by idealization, that $Mpp$ is fixed, the determination of ($D$ and hence) $I$ can be governed by at least three different principles.

If we are interested in all relevant nomic possibilities, $I$ coincides with the set of nomic possibilities at the $Mpp$-level. Let $To$ indicate this subset of $Mpp$. Although we may not have an explicit characterization of $To$, there is a clear empirical criterion for membership: $x$ in $Mpp$ belongs to $I=To$ iff $x$ can be realized. If we are interested in a well-defined subset of the set of $To$, i.e., nomic possibilities satisfying some explicit condition, membership determination is not fundamentally different. In both cases we will speak of the empirical determination of $I$.

In this case, the obvious target of theory development is an explicit characterization of $I$, i.e., a set of models $M$ is sought for which the strong claim holds: $I=\pi M$, such that $M$ may be called the true ($Mp$-)theory about $I$, or simply, the (conceptually relative) truth. In [Kuipers, 2000, Chapter 7 and 9, respectively], the formal structure of truth approximation by unstratified and stratified theories is studied in detail. Here truth approximation is restricted to revising theories, leaving $D$, $Mp$, and hence $I$, fixed. Of course, when $I$ is restricted to a partially well-defined subset of $To$, this also enables keeping the vocabulary $Mp$ and the theory $M$ constant, and revising (the partial definition of) $I$, and hence truth approximation by revision of the domain of intended applications (Kuipers, forthcoming b). Instead of speaking of the ‘empirical determination’ of $I$, we may then speak of the ‘empirical specification’ of $M$.

In [Kuipers, 2000, Chapter 7] the notion of a nomic possibility is presented as an absolute qualification. However, there may well be cases where it makes good sense to distinguish levels of nomic possibility, e.g., as suggested by the following sequence of ‘lower’ to ‘higher’ levels: the physical, chemical, biological, psychological, cultural-socio-economical level. Being nomically possible at a higher level then implies being nomically possible at a lower level, but not the converse. Another example concerns the idea of nomically possible states of an artifact, assuming that it remains intact, which means a severe restriction to its physically possible states, including broken ones. Such refinements can also easily be built into the empirical determination of $I$, as long as the boundaries between the different levels of nomic possibility may be assumed to be sharp.

In many cases, however, the interest is directed to a proper subset of $To$, of which the membership is not sharply defined. One important way in which $I$ can
then have been circumscribed is by so-called paradigmatic determination.

**Definition:** \( I \) is paradigmatically determined if there are \( PAR \) and \( SIM \) such that

1. \( I \) is a subset of \( To \) the intended applications
2. \( PAR \) is a finite subset of \( I \) the paradigmatic examples
3. \( SIM \) is a binary relation on \( Mpp \) a similarity relation
4. for all \( x \) in \( I-PAR \) there is \( y \) in \( PAR \) such that \( SIM(x,y) \)

![Figure 7. Paradigmatic determination of \( I \)](image)

Figure 7 depicts the relations between the various sets in the case of paradigmatic determination of \( I \). The elements of \( PAR \) may, for instance, be determined by the founding father of the theory and corresponds to one of the meanings Kuhn [1962] had in mind with the term ‘paradigms’ and which he later called ‘exemplars’. Of course, the main source of vagueness is the notion of similarity, for it will not as a rule be possible to define this notion sharply, at least not at the beginning of the research process. As a matter of fact, the relevant definition of similarity is to be discovered by trying to undertake successful excursions out of \( PAR \) and a given \( M \), roughly in the same way as described in [Kuipers, 2006] for the first way of determination of \( I \). Of course, each tentative excursion entails tentative sufficient conditions of similarity, and each definite conclusion not only requires clear-cut sufficient conditions but also a definite \( M \), together enabling a definite empirical claim.

In both cases of determination, assuming that the theory has (at least absolute) empirical content, the empirical claim of a stratified theory will not be trivial. As is easy to verify, the empirical claim becomes trivial in the third way of determination of \( I \), so-called auto-determination: for \( x \) in \( To \), \( x \) belongs to \( I \) iff \( x \) belongs to \( \pi M \), i.e., is the projection of a model of a stratified theory. In the case of auto-determination, the theory in question is typically not something to be tested, but it will have been designed for other purposes.
In the case of an unstratified theory, the set of intended applications is of course a subset of the set of nomic possibilities on the $M_p$-level, which is then in its turn a subset of $M_p$. For the further determination of $I$ there are again the same three possibilities of empirical, paradigmatic and auto-determination. Where there are proper theoretical terms involved, all three forms of determination result in problems.

Let us briefly restate and elaborate the background of (epistemological) stratification of a theory in $T$-theoretical and $T$-non-theoretical terms. Let there be an unstratified theory $UT = \langle M_p, M, D, I \rangle$ and assume that $UT$ has non-empty empirical content $M_p-M$. The term $t$ occurring as a component in $M_p$ is said to be $T$-theoretical iff every known method of measuring $t$ in a specific intended application results in a model of $UT$. It is otherwise $T$-non-theoretical. Let $UT$ contain at least one $T$-theoretical term and let us first assume that $I$ is supposed to be empirically or paradigmatically determined, in which case the empirical claim “$I$ is a subset of $M$” is non-trivial. However, testing this claim is impossible, for it leads demonstrably either to circularity or to an infinite regress. In the case of auto-determination, the problem is that determination of the membership of $M$ leads to circularity or infinite regress.

The remedy for these problems is the epistemological stratification of the theory in terms of a partial theory containing only and all $T$-non-theoretical terms. Assuming that the stratified theory has non-empty (absolute) empirical content $M_{pp}-\pi M$, the empirical claim or auto-determination is non-trivial, depending on whether $I$ has or has not been fixed in advance.

The indicated definition of $T$-theoriticity is a pragmatic one, due to the “every known method of measuring”-clause and goes back to Sneed [1971]. Given the fact that the class of known methods can only increase, the definition is perfectly compatible with the advice to classify a term as $T$-theoretical in case of doubt. However, it is tempting to look for an intrinsic definition of theoriticity. Gähde [1983] has put forward an intrinsic definition. This proposal is not only highly technical and restricted to quantitative terms, it has also been criticized for other reasons (Cf. [Schurz, 1990]). However, Balzer [1996] has rebutted this criticism and has, moreover, proposed an essentially simpler formal criterion than Gähde’s is.

2.4.3 Links between theories, and theory-nets

Theories that are roughly about the same domain are frequently related. At each moment they may constitute a network of theories, i.e., an ordered set of theories that are directly or indirectly related. Such a network depicts the synchronic situation, the succession of networks indicates the diachronic development.

Let us first define the main relations between theories, also called (intertheoretical) links. We will presuppose that all theories considered are stratified, hence the theories are of the form $ST = \langle M_p, M_{pp}, M, \pi, D, I \rangle$. It is easy to derive from the definitions what links result if the stratification is assumed to disappear ($M_p=M_{pp}$ and $\pi$ becomes the identity function).
We have already indicated some specializations of theories, in the case of the theories of Newton and Freud.

\( ST^* \) is specialization of \( ST \) iff

1. \( Mp^* = Mp \) and \( Mpp^* = Mpp \) and \( \pi^* = \pi \)
2. \( M^* \) is a subset of \( M \) and \( D^* \) is a subset of \( D \) (and hence \( I^* \) is a subset of \( I \))
   and at least one of the subsets is proper.

Specialization is one of the main research activities within a research program starting from some basic generic theory.

A new theory may add new (non-)theoretical components. It may be a genuine superposition, such that the old theory remains completely intact. Balzer [1982] describes the example of the classical kinematical theory built on the classical space-time theory. When at least one new theoretical component is introduced this type of link is called (conservative) theoretization and can be defined as follows:

\( ST^* \) is a theoretization of \( ST \) iff

1. all (non-)theoretical components of \( Mp \) remain (non-)theoretical components of \( Mp^* \)
2. \( Mp^* \) adds to \( Mp \) one or more new (non-)theoretical components,
   at least one theoretical one, all of which are stripped off by a function \( f \) from \( Mp^* \) onto \( Mp \)
3. for all \( x^* \) in \( M^* \), \( f(x^*) \) belongs to \( M \)

Of course, it is possible to define non-conservative links between theories when new components are added.

The third important link between theories is that of reduction.

\( ST \) is reducible to \( ST^* \) iff there is a relation \( r \) on \( Mp \times Mp^* \) such that

1. for all \( x \) in \( M \) there is \( x^* \) in \( M^* \) such that \( r(x, x^*) \)
2. if \( r(x, x^*) \) and \( x^* \) in \( M^* \) then \( x \) in \( M \)
3. for all \( y \) in \( I \) there is \( y^* \) in \( I^* \) such that \( r_\pi(y, y^*) \),
   where \( r_\pi \) indicates the projection of \( r \) on \( Mpp \times Mpp^* \)

Roughly speaking, the definition captures the explanation of one theory or law by another when the models and intended applications can be formally related in a way which is typically possible when one or more of the three basic types of reduction distinguished in [Kuipers, 2001, Chapter 3] apply.

The last type of link between theories to be mentioned is that of (idealization, or conversely) concretization, of which a precise definition has been given in [Kuipers, 2000, Subsection 10.4., see also Chapter 11]. There it is shown that concretization
plays, for example, a crucial role in the truth approximation analysis of the transition of the theory of ideal gases to that of Van der Waals, and various transitions in the old quantum theory and capital structure theory. An informal explication of ‘idealization & concretization’ will be given in Subsection 3.3.2.

It is not difficult to check that all defined links generate partial orderings and we assume that all links to be considered are of this type. Let two theories be called directly related when one such link applies and let them be called related when there is a chain of directly related theories between them. Of course, ‘being related’ is again a partial ordering. A *theory-net* is defined as a set of theories related in this way such that there is a *basic theory* $T_b$, in the sense that all other theories are directly or indirectly related to this theory. Figure 8 depicts such a theory-net.

![Figure 8. A theory-net](image)

For the succession of theory-nets it is plausible to distinguish two basic types, leaving room for mixed cases. A transition to a new net may be *conservative* in the sense that the new net retains all theories of the old one, but one or more theories to the old net are added in some way. A transition is called *corrective* if one or more theories in the old net are replaced by new theories that are considered to be improvements.

### 2.4.4 Constraints

We have already referred several times to so-called constraints. Whereas laws and axioms in the normal sense lay down restrictions on individual potential models, constraints impose restrictions on sets of potential models. A particular type of constraint is a so-called identity-constraint, guaranteeing that a function assigns in different potential models, with some common base-sets, the same value to the same individuals. The weight-function in the case of the slide balance as well as the mass-function in the case of classical particle mechanics are cases in point.

A constraint can be formally defined in a very general way.

*Definition: C is a constraint on the set $S$ iff*

1. $C$ is a set of subsets of $S$
2. the union of the sets in $C$ exhausts $S$ ($UC=S$)
3. if $X$ is in $C$ and $Y$ is a subset of $X$ then $Y$ is in $C$
(subset-preservation)

It is easy to prove that all singleton sets $\{x\}$, for $x$ in $S$, belong to $C$, hence a constraint does not exclude any individual potential model.

Let us now first concentrate on the typical role of a constraint $C$ on $Mp$ in a stratified theory $ST = \langle Mp, Mpp, M, C, \pi, D, I \rangle$. The standard empirical claim was “$I$ is subset of $\pi M$”, which could be paraphrased by saying that all members of $I$ can be extended with theoretical components to genuine models, i.e., there is a subset $X$ of $M$ such that $\pi X = I$. Taking the constraint into consideration this claim is strengthened: there is a subset $X$ of $M$ belonging to $C$ such that $\pi X = I$. Hence, now both $M$ and $C$ restrict the degrees of freedom for the supplementation of theoretical components. In the corresponding versions of the strong claim the clause “$X$ is a subset of $M$” is simply replaced by “$X = M$”. It is clear that a stratified theory may even be a pure constraint-theory, in the sense that $C$ is non-trivial and $M$ is trivial, i.e., $M = Mp$.

In [Kuipers, 2000] we deal with truth approximation by theories by assuming that theories are sets of structures, with or without the distinction between theoretical and non-theoretical terms. It is not too difficult to check that, when the truth is a constraint-theory, similar (basic and refined) definitions may be given of the claim that one constraint-theory may be closer to the truth than another. Of course, it is then also possible to deal with truth approximation by theories consisting of a ‘normal’ and a constraint part. However, we will leave the elaboration of the suggested possibilities to the reader.

Taking a constraint into account, the following reformulation of the standard claim is instructive. Let $A(ST)$, the application space of $ST$, be defined as the set of projections of all subsets of $M$ satisfying $C$ (formally: $A(ST) = \pi(P(M) \times C)$). The standard claim now comes down to: $I$ is in $A(ST)$.

It is also now plausible to define the (absolute) empirical content $AEC(ST)$ as $P(Mpp) - A(ST)$, i.e., the subsets of $Mpp$ which are excluded by $M$ and $C$. Note that $AEC(ST)$ reduces to $P(Mpp) - \pi(P(M))$ when $C$ is trivial, i.e., when $C = P(M)$, and to $P(Mp) - P(M)$ when $\pi$ is trivial, i.e., when $\pi$ is the identity function. It is easy to verify that $AEC(ST)$ is empty in these respective cases iff the originally defined empirical contents $Mpp - \pi M$ and $Mp - M$ are empty. Hence, the suggested new definitions of empirical content reproduce the original ones on the level of sets of potential (partial) models.

Similar relations hold for the plausible definition of the relative empirical content $REC(ST)$: $P(Mpart) - A(ST)$. And again it follows almost trivially that non-empty $REC(ST)$ implies non-empty $AEC(ST)$, but not the converse.

Constraints also make sense in other cases, e.g., in partial theories and in unstratified theories. For the first case, let $Cpart$ be a constraint on $Mpp$. Like $Mpart$, $Cpart$ represents empirical restrictions. We may say that $Mpart$ captures the standard observational laws, whereas $Cpart$ captures constraint observational laws. Of course, the associated claim states that $I$ is a subset of $Mpart$ belonging
to $C_{\text{part}}$. It is important to note that many empirical laws are constraint laws, or mixtures of standard and constraint laws.

If the indicated partial theory with constraint is isolated from the full theory, it is clear that we have an unstratified theory with constraint. In general, an unstratified theory with constraint is of course of the form $(Mp, M, C, D, I)$, where $C$ is a constraint on $Mp$.

## 2.5 Non-empirical Theories

In this section we have dealt with empirical theories, but let us finally consider the question of which structuralist concepts are also useful for non-empirical theories. Following Popper, non-empirical theories are by definition theories which are not (intended to be) falsifiable. One may distinguish at least four types of non-empirical theories:

- **Metaphysical theories** are supposed to make claims about reality without assuming any particular conceptualization or, equivalently, they make claims generalizing over conceivable conceptualizations of reality.
- **Mathematical and logical theories**, some of which deal with defined abstract objects, i.e., mental constructs, e.g., the theory of groups, other ones with ‘concrete’ mathematical objects, e.g., the theory of natural numbers.
- **Conceptual theories** concern ways of looking (perspectives) at a certain domain.
- **Normative theories** deal with what is (supposed to be) ethically, legally, aesthetically (in)admissible.

It is evident that almost all technical ingredients presented for empirical theories are also useful for non-empirical theories. In fact, Suppes [1957] invented the structuralist representation of empirical theories by transferring, as far as possible, the standard way of presenting mathematical theories, such as the theory of groups, to empirical theories. The crucial difference is that non-empirical theories do not make general empirical claims. The claims which are associated with them typically are either conceptual (logical, mathematical etc.) or restricted to individual intended applications. A typical claim around a mathematical theory is a mathematical theorem to the effect that the models of the theory can be proven to have a certain explicitly defined property. A typical claim of a specific conceptual theory is that a certain intended application is (or is not) a model of that special theory. Of course, generic theories, i.e., theories with vacuous empirical claims, are conceptual theories.

The ‘structuralist theory of the structures of empirical theories’ is a perfect example of a theory that is primarily intended as a conceptual theory (although one may strengthen it to a genuine empirical theory by adding a substantial claim). As
a consequence, the foregoing exposition not only provides an elaborate example of a conceptual theory, it can also convince the reader of the usefulness of conceptual theories.

3 RESEARCH PROGRAMS AND RESEARCH STRATEGIES

Introduction

One of the most important insights of the philosophy of science since about 1960 is the awareness that the development of science should not be described in terms of the development of specific hypotheses and theories, but in more encompassing terms. The two main proponents of this insight are Kuhn and Lakatos. Kuhn first preferred the term ‘paradigm’ and later ‘disciplinary matrix’ [Kuhn, 1962/1969]. Lakatos [1970; 1978], basically aiming to capture the development of sequences of related theories, introduced the notion of a ‘research program’. Half a dozen other terms are used to denote roughly the same concept, although their details may differ.

We prefer Lakatos’s term ‘research program’ for its literal meaning: program of research. Although our favorite conception of research program will be somewhat weaker than that of Lakatos, we will use the same term. However, it should be stressed in advance that nobody means program in the detailed sense of a well-ordered sequence of things to do. At most a program of research in some global sense is meant. From now on, ‘program’ always means a research program.

Subsection 3.1. distinguishes four types of research programs. Subsection 3.2. presents a necessarily incomplete summary of current insights in the structure and development of research programs and some other global cognitive units. Finally, a number of strategic lessons are suggested, which have been freely derived from both the insights in global cognitive units and the distinction of four types of research programs. Subsection 3.3. deals with program internal strategies and Subsection 3.4. with interaction strategies. In the concluding remarks, we will reconsider the dynamics of descriptive and explanatory programs in the light of the law-distinction treated in Section 1.

3.1 Four types of research programs

Four ideal types of research programs will be described, followed by a survey of the main similarities and differences.

3.1.1 Descriptive, explanatory, design and explicative programs

The four types of programs to be distinguished are the following: descriptive, explanatory, design and explicative programs. They form ideal types. In consequence, mixtures are the rule, rather than the exception. However, it is often possible to describe a mixed program as a cooperative enterprise of two or more
programs of an ideal type, with one of them being in some way dominant. To put it still more cautiously, the first characterizations may well be read as descriptions of four types of research, which are, in practice, part of complex undertakings. However, in these four types of research it is then often possible, at least analytically, to identify the underlying programs of an ideal type mentioned above.

Programs of the first three types are usually considered to belong to the empirical sciences. Programs of the fourth type, explicative programs, are not only characteristic for constructive analytic philosophy, but also occur elsewhere, viz., in mathematics and the empirical sciences.

Our claim is that these four types of research and research programs reflect the core of two divisions of labor. One figuratively between the cognitive products of scientific inquiry themselves, viz., descriptions, explanations, products, and concepts. The other literally between their producers, roughly speaking, experimentalists, theoreticians, engineers, and philosophers/mathematicians. Moreover, it is claimed that these related divisions of labor on the level of products and producers, and their interaction, play a crucial role in the dynamics of science, which can be even more fully exploited by understanding their nature in more detail.

Descriptive programs are meant to describe a certain domain of phenomena, primarily in terms of individual facts (individual programs) or primarily in terms of general observable facts (general or inductive programs). Descriptive programs form a certain kind of observation program and may be fundamentally based on experiments, in which case it is plausible to speak of experimental programs. A famous example is Boyle’s search for a relation between the pressure and volume of a gas, followed by Charles, Gay-Lussac and others with their quest for the relation with temperature. To mention just one other historical example for the moment, the famous investigation by Durkheim of what he called the social facts about suicide was typically a descriptive program.

Descriptive research takes place by more or less selective (experimentation and successive) observation, and the resulting facts are couched in so-called observation terms. These observation terms are not given by the natural world, but form the specific glasses through which the researcher in that program is looking. At the start of a descriptive program there usually is only some core vocabulary. For the rest it is not altogether clear which further observation terms are to be considered as relevant and precisely how certain observation terms are to be interpreted. Additional terms are only selected and shaped in the course of the development of the program. In line with Section 1, it should also be stressed that, at least as a rule, observation and hence observation terms are, and remain, laden by theoretical presuppositions, which are considered to belong to the so-called unproblematic background knowledge.

Explanatory programs have another aim. Individual and general explanatory programs are directed at the explanation and further prediction of the observable individual and general facts in a certain domain of phenomena. Hence, an explanatory program has a (quasi-) deductive nature and is always built on an underlying descriptive program. For this reason explanatory programs are frequently devel-
oped along with underlying descriptive programs, in which case the two types of program can be distinguished only analytically. The kinetic theory of gases on the one hand and the anomy theory of Durkheim on the other provide paradigm cases of explanatory programs built on the previously mentioned descriptive programs. The primary objective of the kinetic program was the explanation and detailed prediction of the precise relation between pressure, volume and temperature by applying Newton’s laws to collisions of molecules between each other and with the wall. To illustrate this fact, we confine ourselves to one representative of the many researchers conducting this type of research: van der Waals. Similarly, Durkheim tried to explain the social facts about suicide with his anomy theory.

Other examples of explanatory programs are Newtonian mechanics, the transformational generative grammar of Chomsky, and the theory of rational choice or general utility theory, the latter providing the foundation of, among other things, neo-classical economics and so-called explanatory sociology.

It is important to be aware of the fact that several explanatory programs may arise on the basis of the same descriptive program. They may be competitive, but need not be.

The most important tools used by explanatory programs are theoretical terms, denoting fundamentally new concepts. The distinction between theoretical and observation terms has already been introduced and studied in the previous sections. In the present context of explanatory programs it is important that theoretical terms have not yet been firmly established as observation terms, neither inside nor outside the program. Of course, the terms as such may have been used before to refer to a related concept. Examples of theoretical concepts are the concept of force in Newtonian mechanics, Chomsky’s concept of deep structure, and the concept of utility in utility theory.

The new terms may refer to theoretical properties, relations and functions, as suggested by the examples, but also to newly postulated entities, such as atoms and genes. If an explanatory program introduces theoretical terms, it may also be called a theoretical program. If it does not, which certainly is possible, it belongs to the explanatory subtype of observational programs, to be distinguished from the descriptive subtype.

For most of the empirical sciences the above characterization of descriptive and explanatory programs makes sense and is useful. Although analogous programs occur in the historical sciences, especially programs about individual facts, the characterizations above of descriptive or explanatory programs are not suitable for historical research. In particular, general historical programs are rare, probably due to the fact that general historical facts, i.e., empirical laws and theories, are rare. Unfortunately, it seems that neither type of program in either their individual or general forms have yet been elaborated for the historical sciences.

There remain two further points to make about descriptive and explanatory programs of the general kind. The first may also be called inductive and the second deductive, because induction dominates the first and deduction the second type of program. Moreover, whereas descriptive programs are always observational,
explanatory programs may or may not be theoretical.

In the current philosophy of the empirical sciences the main attention is paid to description, explanation and prediction. However, an important part of the empirical sciences is not primarily concerned with any of these three tasks. Design or constructive research programs involve the design and actual construction of certain products. Some examples are: programs directed at the production of new medical drugs, the improvement of breeding methods of plants, the design of training programs for certain types of handicaps, the design of so-called expert systems, and the construction of new materials. As the examples illustrate, the products of design programs need not be products in a strict sense but may also be processes, or their improvement. The product targeted by a design program has to satisfy certain previously chosen demands; these demands are of course derived from the intended use of the product being developed.

The examples also illustrate that design programs do not only occur in what are traditionally called the technical or technological sciences but also in other areas of scientific research. This is the reason for not choosing the term ‘techn(ological) research programs’, for that might be too narrowly interpreted.

Since design programs often use knowledge obtained in descriptive and explanatory programs, the design process will only be considered to belong to scientific research if it is not fully based on existing knowledge and techniques. That is, new theories have to be developed or new experiments have to be performed in order for a design program to be scientific in nature.

For philosophy and mathematics the fourth type of program, the explicative research program, is the most important type. Such programs are directed at concept explication, i.e., the construction of a simple, precise and useful concept that is, in addition, similar to a given informal concept (cf. [Carnap, 1963, 118]). For example, the concepts of ‘logical consequence’ and ‘probability’ have given rise to very successful explicative programs in the borderland between philosophy and mathematics. One of the main explicative programs dealt with in [Kuipers, 2000] is intended to explicate the intuitive idea of ‘truthlikeness’. In Section 1 we have dealt with the explication of the intuitive conceptual distinction between observational laws and proper theories.

The strategy of concept explication is the following. From the intuitive concept to be explicated and, when relevant, empirical findings one tries to derive conditions of adequacy that the explicated concept will have to satisfy, and evident examples and counter-examples that the explicated concept has to include or exclude.

Explication may go further than the explication of intuitive concepts, it may also aim at the explication of intuitive judgments, i.e., intuitions, including their justification, demystification or even undermining. A main example in [Kuipers, 2000] concerns the intuition about the functionality of choosing empirically more successful theories in order to enhance truth approximation. The strategy of ‘intuition explication’ is a plausible extension of that involving concept explication.
3.1.2 Similarities and differences

Although the four types of programs distinguished are different in many respects, they also have an important similarity. In all cases we can identify an internal goal, viz., the true description, the true theory, the intended product and the intended concept.

The fully correct observation and registration of the totality of facts observable by the glasses of a descriptive program is called the true description of the domain. All other descriptions of the domain in terms of the program are either incomplete or (partially) false. This true description constitutes the internal goal of the descriptive program. It is important to note that the true description not only depends on reality but also on the program in which the choice of the observation terms delimiting its viewpoint, co-determines what will and what will not be observed. Hence, the true description is a program relative but nonetheless informative characterization of reality. If the program concentrates on individual facts, i.e., object, place and time specific (conceptualized) facts, we will speak of the true individual description, if it concentrates on general facts, i.e., generalizations of (conceptualized) individual facts, we will speak of the true general description. In the latter case, the true description corresponds to the true theory within the (observational) vocabulary of the program.

In the case of a (general) explanatory program there is, usually, supposed to be a unique theory. That is, there is assumed to be a theory in terms of the observation and eventual theoretical terms of the program, which not only explains, and predicts as far as relevant, all observable facts of the domain. It also uses only those theoretical terms in a substantial way that refer to something in reality. This theory, in fact the strongest true hypothesis, will be called the true theory of the domain, constituting the internal goal of the explanatory program. Like the true description, the true theory is determined by the specific combination of program and reality, hence it is program relative. If the vocabulary is observational it corresponds to the true general description.

The intended product, i.e., a product that satisfies the demands put forward, forms of course the internal goal of a design program, the analogue of the sought-after true description or true theory. Finally, in the case of explicative programs the intended concept, i.e., a concept satisfying the conditions of adequacy, constitutes the analogue of the internal goal of the previously considered programs.

Despite the fact that, similar to descriptive and explanatory programs, design programs always have internal goals, such goals differ greatly from those involved in description and explanation. In descriptive and explanatory programs internal goals are only indirectly characterized, and all the efforts are directed at the explicit characterization of the true description or the true theory. In design and explicative programs, the internal goal, the intended product and the intended concept are all explicitly characterized from the beginning, at least to a certain extent.

23 Of course, not every aspect of this true description may have our interest. It may or may not be possible to restrict the observational vocabulary, and hence the program, to some sub-vocabulary of what we find really interesting.
Another, related difference is the degree of uniqueness of the internal goal. As mentioned above, the true description is in principle uniquely determined jointly by the program and reality. Hence, it cannot change in the course of the program without either changing the program or the domain. The same holds *mutatis mutandis* for the true theory. In the case of design programs, on the contrary, the intended product need not be determined uniquely at all, for there is, as a rule, the possibility of functional equivalents, i.e., different products serving the same purpose, in which case it is also said that the purpose is ‘multiply realizable’. Moreover, the desired product has to be determined in more detail in the course of its development, in which the strategic considerations of feasibility, affordability and salability play an important role. As an aside, it should be remarked that the intended product could also be over-determined by the sum of all demands.

Most of the differences mentioned also apply *mutatis mutandis* to explicative programs. This is no accident, for they form a kind of abstract design program, viz., of concepts.

3.2 Structure and development of research programs

We will discuss the structure of programs mainly in terms of five possible components. Then we will discuss the development of programs in terms of an internal and external phase. A brief presentation of the atomic theory as a developing research program will illustrate most of the components and phases. The section concludes with a global survey.

3.2.1 Five components of research programs

So far the descriptions of the four types of programs may well be used to indicate just four types of goal-directed research. However, when we start to discuss the structure of programs it will become clear how programs acquire more identity than defined by their internal goals. We will discern an ordered set of five possible components of a program, viz., domain, problem, idea, heuristic, model. Since each component is supposed to presuppose the foregoing ones, this leads to five qualitative degrees of strength of programs.

A possibly disenchanting reformulation of the similarity between the four types of programs is that they are all directed at the solution of a certain problem, viz., to attain the internal goal of the program. This orientation seems the least one may expect of scientific research, i.e., that it is directed at the solution of a certain problem. Programs satisfying this minimal requirement might be called *programs with a problem*. In the practice of research policy, however, one even speaks of research programs when there is only a more or less well-defined domain of research, without a clear problem, for lack of an internal goal, in which case one might speak of *programs with (only) a domain*.

\[24\text{E.g., the intended concept is usually not uniquely determined, such that there are degrees of freedom in explication, leaving room for diverging explications. For this reason one may even prefer to speak of concept modeling instead of concept explication [Brink, 1989].}\]
From the descriptions of the four types of programs it now follows that a research program is minimally conceived as a program with a problem. However, the prototypical meaning we want to advocate for the term research program is that of a program based on an idea, i.e., a program with not only a domain and a problem but also a fundamental idea governing the search for the solution to the problem. It could be called Popper’s requirement, because more than anyone else he has stressed the equal importance of problems and ideas in scientific research. Of course, such a fundamental, leading idea is usually a complex idea, i.e., a set of coherent ideas. It will at least include the choice of a core vocabulary, and usually it includes one or more principles using that vocabulary. The idea should be strong so that it can provide secure footing for a research venture that should be able to withstand some critical blows. In other words, it should be possible to protect the fundamental idea somewhat dogmatically against falsification or similar threats. The standard way to do this is by trying to blame auxiliary hypotheses, but there are several other defense strategies.

According to Lakatos, the leading idea constitutes the hard core of a program. However, the notion of hard core has a double face. Lakatos’s primary meaning is that a program is only one and the same program as long as the hard core remains the same. However, it frequently occurs that one feels obliged sooner or later to adjust the fundamental idea of a program, in which case one should strictly speak of a new program. But it seems more adequate to leave room for a semi-hard core of the program, a core that may be adjusted, when no other escape seems possible. We would like to stress another meaning component of the notion of a hard core. An idea, before or after a possible change, may be hard in the sense that it is supposed to be valid for the whole domain. It does not leave room for exceptions. Speaking of a ‘core idea’ may indicate this quality. Incidentally, one way to retain the goal of no exceptions, in the face of persistent threats is to adjust the core idea, another is to adjust the domain. In sum, we conceive the fundamental idea on which a program can be based as a core idea of a semi-hard character.

To be sure, Lakatos only speaks of a genuine research program when there is, in addition to a hard core, also a so-called positive heuristic providing suggestions for protecting auxiliary hypotheses and their adjustment. Hence, a program, in Lakatos’s sense, is a program with a hard core and a positive heuristic. That is, a program governed by two ideas, the first one directly bearing on the solution of the problem, and the second one concerning the way in which the first idea can be defended against attacks.²⁵

Zandvoort [1984] has convincingly shown that the impressive examples of Lakatosian programs frequently are programs in which the positive heuristic is provided by an analogy or model, where he refers in particular to the notion of analogy as discussed by Nagel [1961]. Such programs with a hard core and a model as a positive heuristic are maximally equipped to provide internal guidelines for

²⁵Lakatos’s concept of negative heuristic coincides with the intention of keeping fixed the hard core as long as possible. Hence, this notion does not add something to the notion of the hard core (in the sense of Lakatos).
research.

Research programs with a core idea and a stable positive heuristic, whether or not in the form of a genuine hard core and a model frequently occur in all forms of empirical science, not only in the natural sciences, but also in the social sciences and the humanities. However, many other programs have a semi-hard core idea about the way in which the goal has to be attained, without having a strong idea about the way in which that fundamental idea has to be protected. In other words, although they have a core idea, they don’t have a stable positive heuristic. To put it differently, the historical claim in the beginning of this section can be stated more precisely as follows: the global history of science can best be described in terms of rising, winning and falling programs based on a core idea. For this reason we will henceforth assume the prototypical meaning of the term ‘(research) program’ to be a program based on a semi-hard core idea, but not necessarily equipped with a stable positive heuristic.26

### 3.2.2 Examples of (the core ideas of) research programs

We now present a list of examples of ideas forming the core of equally many well-known programs, starting with explanatory programs. These are the idea in kinetic theory that gases consist of molecules which move and collide according to Newton’s law; Mendel, starting from the problem whether regularities can be found in the apparent chaos of the manifold variations of observable features in all the hybridizations known to plant breeders at that time, postulated the idea that the male and female hereditary material is not a unit of which one wins out, but that it can be factorized so that part of the male and part of the female hereditary material determines the observable features of the organism; the idea in general utility theory that choices are governed by maximizing expected utility; Chomsky’s original idea that the grammatical sentences of languages can be generated by the application of a limited number of transformation rules on an equally limited number of deep structures; and, finally, the central idea in classical computationalism, or symbolism (Newell and Simon), according to which human behavior should be (described and) explained in terms of problem solving.

The last example is a nice borderline case between explanatory and descriptive programs. An example of a purely descriptive program is network analysis, which is based on the idea that schemas using connecting arrows can result in very informative descriptions. Another example is fractal geometry, initiated by Mandelbrot, based on the idea that shapes in nature on different scales may nevertheless be congruent. A coastline of 10 meters looks like a coastline of 10 kilometers. So-called discourse-analysis provides a similar example. Behaviorism can also be viewed as a broad, descriptive program, with the core idea that one should restrict the scientific attention to the description of (patterns in) observable behavior. Finally, the goal of the Human Genome Project, essentially completed in 2000, was the true description of the (almost unique) composition of the 23 human chromo-

---

26Unfortunately, the term ‘idea program’ is linguistically not very attractive.
somes as pairwise sequences of the four bases C, T, A, and G, that is, the typical vocabulary of DNA. To be sure, although the sequences are almost unique, the individual variations provide an almost perfect means of identification.

The core idea of descriptive programs is frequently formed by a methodological searchlight principle, for instance the principle of causality, functionality or intentionality or by a description or representation principle, as in the cases of network analysis and fractal geometry. Such leading principles usually open, for different domains, the possibility of a specific program directed at that domain. Such a representation principle may be guided, or is at least made available, by accepting a theory as an observation theory, that is, a theory that has become accepted as (approximately) true. This important phenomenon has already been explained in more detail in the first section. A typical example is the functional (descriptive) genomics program, building upon the ‘structural’ descriptive (Human) Genome Project, identifying the function of fragments of the chromosomes. That is, each fragment may or may not play a crucial role in (‘code for’) the generation of certain characteristics of organisms. This research is guided by the principle of functionality, rooted in the theory of evolution, and according to which features of organisms have functions. Even more than on the macro-level, there are many exceptions to this principle on the present micro-genetic level. As a matter of fact, most DNA-fragments seem without function.

In the case of a design program the leading idea is frequently called the lead. It is the core idea about the way in which the intended product should be construed and possibly with what material. Some examples are the following, starting with an example of a purely technical, non-scientific, nature. The idea of a bicycle chain was developed at the end of the nineteenth century, and enabled the, still time consuming, design of a riding bike with two wheels of equal size, i.e., the bicycle, which was a very attractive feature of the walkbike of much earlier date (cf. [Bijker, 1995]). The idea of nuclear fission resulted in the development of nuclear power stations, and, to be honest, atomic bombs. The development of power stations based on nuclear fusion is still one of the main challenges of applied physics. One of the main starting points of computer science, the development of the standard Von Neumann architecture of digital computers, began with Turing’s idea of a universal computing machine, the so-called Universal Turing Machine. Within the Von Neumann architecture the idea of so-called production systems, containing and generating complex ‘if, do then’ (production) rules, has turned out to be very successful, in particular for creative computing tasks. According to the central idea of the technological version of connectionism, learning mechanisms can be produced by connections between knots that are strengthened or weakened according to whether the previous response was or was not adequate. One of the leading ideas in cancer research, due to Judah Folkman (see [Boehm, et al., 1997]), is to try to stop the formation of blood vessels leading to the tumor or, alternatively, to try to block their functioning. To mention an example of quite a different nature, the idea of conversation groups made up of people with similar personal problems was only recently introduced in all kinds of therapeutic contexts, with
Laws, Theories, and Research Programs

varying but increasing degrees of success. Their development was partly due to a systematic search for the best specific conditions in which to conduct therapy. For example, certain group therapies for breast cancer patients, an idea of Spiegel, Bloom, Kraemer and Gottheil [1989], see also [Spiegel, 1993], seem to improve their immune system. Finally, in drug research [Vos, 1991, 62], the lead can be identified with the set of wished for properties, the wished for profile, and some idea about how to realize it. More specifically, the ‘lead compound’ comprises a chemical compound with certain operational characteristics, the operational profile, together with the wished for profile. Only if there is already known to be an interesting overlap between the two profiles, it is a serious lead compound, and the challenge is to reduce the differences.

A nice example of an explicative program is the famous idea of Rawls, according to which the determination of the concept of a just society can best be undertaken by way of a thought experiment. In this scenario the future members of the just society to be construed do not know the place they are going to occupy in that society; that information is hidden behind ‘the veil of ignorance’. Another example is the core idea of logical model-theory that the intuitive concept of ‘logical consequence’ should be explicated in terms of the models of the relevant language and according to the principle: the conclusion should be true in at least all models in which the premises are true.

In the last few decades several research programs have been described in detail and, of course, in the terminology preferred by the respective authors. To give one example, Von Eckardt [1995], whose aim is to characterize cognitive science, describes the ‘research framework’, as a combination of domain-specifying assumptions (domain), basic research questions (problem(s)), and substantive assumptions (core idea(s)). In these terms she has given a lucid description of the framework for cognitive science, to be precise, as far as it is focused on adult, normal, typical cognition. In her approach, symbolism and connectionism appear as two different specifications of the ‘computational system (substantive) assumption’.

3.2.3 Additional considerations

A significant problem arises when any attempt is made to identify a respectable or even strong core idea. A plausible procedure to conclude to the existence of a research program based on an idea when there have appeared in the blindly refereed international literature several publications, from one or more authors, in which the idea is exposed, discussed and elaborated. In principle, the existence of international publications coalescing around one idea should be an adequate criterion since science is an international activity in the sense that national borders should not play an important role, in particular when the distribution of strong research ideas is concerned. However, presence in the international literature is certainly not an infallible criterion; it is neither a necessary nor a sufficient one. Referees and journals are necessarily selective, nor are they immune to trends
and fashions. Hence, occasionally it may happen that bad ideas are promoted and that good ideas repressed. Moreover, there may well be strong ideas for which there is not very much interest in the discipline itself, but outside that discipline there may be considerable interest from other disciplines or scientific externals, i.e., from society and technology. The last case may particularly apply to ideas in design research, in which case it is not plausible to expect international scientific publications, as they may be prevented by the need for secrecy. Again, strong external interest is not a safe criterion, but the combination of international publications and lasting external interest is the best criterion we can think of for the identification of valuable research programs.

Although our concept of research program resembles Lakatos’s concept of research program the most — it is a weakened version — this does not mean that ideas about structure and development of research programs can only be derived from the writings of Lakatos. As already mentioned, other authors have distinguished related cognitive units, and described their structural and dynamic features. Kuhn [1962/1969] speaks first about ‘paradigms’ and later about ‘disciplinary matrices’, Fleck [1935/1979] introduced the notion of ‘styles of thought’, and Laudan [1977] deals with ‘research traditions’. We have also seen that Von Eckardt [1995] has more recently dealt with the notion of ‘research frameworks’. Additionally, ‘theory nets’ are distinguished in the structuralist approach to scientific theories that has been presented in the previous section. To conclude this incomplete list, Hamminga [1983] also uses the term ‘research program’, but gives it a detailed meaning tailored to economic research programs.

In the next subsection we will deal with the main dynamic features of research programs, with emphasis on explanatory programs. We conclude this subsection by mentioning one other structural feature, derived from the structuralist approach. The domain of a research program can frequently be divided into a number of subdomains. In such cases it is possible to make a distinction between the core idea associated with the core vocabulary. That is, to distinguish general or generic principles that are supposed to be valid for the whole domain from special concepts and principles that are only supposed to be at stake for a subdomain. Think of Newton’s general laws of motion and the special force laws. In many such cases the division into subdomains is such that it makes a lot of sense to speak of sub-programs, as the crucial idea for a special principle pertaining to a particular subdomain may well constitute a genuine research program in itself.

3.2.4 Phases of developing research programs

Our treatment of the dynamics of programs will mainly concentrate on explanatory programs, and close with a few remarks on the validity of the findings for other types of programs. One might prefer to read first the next subsection, dealing with the development of the atomic theory, and then return to this subsection.27

27 For a detailed analysis of design programs the reader is referred to Chapter 10 of [Kuipers, 2001] and of explicative programs [Kuipers, forthcoming].
We begin by elaborating a previously mentioned relativization of the term ‘program’. A program is never fully mapped out in advance. At each moment only a few principal features are established. They enable researchers to look forward no more than a little bit, and depending on the results of their efforts, the program is adjusted and mapped out a bit further. This is the main reason why responsible bureaucratic middle- and long-term planning of research is impossible.

A program can pass through several phases. In cases of successful programs it is frequently possible to make a global distinction between an internal and an external phase.

In the **internal phase** the elaboration and evaluation of the core idea are central. When a program persists for some period of time it is usually possible to divide the internal phase into two subphases, viz., a heuristic and a test or, as we prefer to call it, evaluation phase. In the **heuristic phase** the new idea breaks through and the first auxiliary strategies are invented to protect the idea. This phase may or may not take place against the background of a so-called Kuhnian crisis of another program, for which seemingly unsolvable problems, called anomalies, have accumulated.

Gradually there comes a transition to the **evaluation phase**. The idea is elaborated for a small number of contexts or subdomains into specific theories, and these are evaluated. The core idea now constitutes the so-called core theory or generic theory, common to all specific theories. Evaluating a specific theory implies as a rule, that for the particular subcontext, a sequence of specific theories is developed, each containing auxiliary hypotheses, each resulting ideally in increasing success, and each including a decreasing number of (types of) counter-examples. Usually such a sequence satisfies the pattern of idealization and concretization: the consecutive theories take into account factors neglected by the foregoing ones.

If this way of branched evaluation is not overall successful, the program is not necessarily deemed useless and made to disappear forever into ‘the museum of knowledge’, exposing, for example, abandoned research programs. A failing program cannot only continue to inspire new research questions, it may also be the case that in a later stage someone succeeds in giving a successful turn to the program.

When the evaluation proceeds successfully, this usually leads to the more or less general acceptance of the core theory of the program and it has become clear for which domain and in what sense and to what extent the core theory can be assumed to be true. It should be stressed that many, if not most, programs in the empirical sciences, not to mention philosophy, do not reach this point. But if this stage is attained, the researchers in that program are left with two options. The first possibility is to look for another program presenting a new challenge. The second possibility is to try to direct the program for the benefit of questions

---

28 The branched portrait of evaluation of a research program is mostly a product of the structuralists; the idea of a sequence of improving theories comes from Lakatos, following Popper; the pattern of idealization and concretization originates with Nowak [1974; 1980], and is further developed by others, e.g., Krajewski [1977].
that are *prima facie* independent of the program. The program then enters the *external or application phase*. The so-called Starnberg school [Schäfer, Böhme and Burgess, 1983] calls this finalization, and means by it in particular the application of the core theory to technological or social problems. This is seldom a matter of simple application of the theory. It usually requires highly specialized theory development, and may even lead to the start of a new scientific discipline, e.g., aerodynamics, in the case of the technological goal of airplanes and the like. We will nevertheless simply speak of application, in this case more in particular about the external application of science. Another form of application arises from the fact that an accepted theory may be usable as observation (or measurement) theory.

Zandvoort [1986; 1988; 1995] has convincingly established that research programs in the natural sciences, which have successfully passed the internal phase, are not always directly applied to problems external to science. They are at least as frequently applied to science internal problems. Hence, the terms ‘internal’ and ‘external phase’ of a program should be strictly interpreted as program relative: internal and external to the program. As a rule, the science internal application of a successfully established program means that the program is directed at the solution of specific problems generated by other programs, possibly but not necessarily design programs. It may also be used for observations relevant to other programs, requiring the acceptance of the core theory as an observation theory. Zandvoort has shown that Popper, Kuhn, Lakatos and others have unjustly neglected this type of cooperation between programs: it constitutes the main part of successful interdisciplinary research within the natural sciences. Among others, Zandvoort’s findings make it clear why it is not only difficult to show the practical, i.e., science external, relevance of natural science research programs that are still in the internal phase, even the practical relevance of programs in the application phase may well be only indirectly so. In the next section we will return to this and other types of cooperation between programs.

So far we have not paid any attention to the question of what is precisely meant when a program is successful or makes progress. To be sure, there are many sorts of success; not all of them need to be sufficient or even relevant for real progress. Success criteria for progress should of course be derived from what scientists themselves count as progress, as for instance expressed by Nobel prizes and other prestigious scientific recognition. According to Popper and Lakatos the factual criterion for progress used for scientific recognition is not just increasing explanatory success\(^29\), predictive success is also required. That is, it is not enough that a program succeeds in explaining new facts, from time to time it should predict, and hence also explain, new facts. To put it differently, there should not only be postdictive but also predictive explanatory success. Although this criterion turns out to be logically too strong as an indicator of truth approximation\(^30\), it

---

\(^{29}\)The notion of ‘explanatory success’ should be taken here in the liberal sense of derivational success, that is, facts that can be derived from the theory.

\(^{30}\)See [Kuipers, 2000, Chapter 6]. For the purpose of truth approximation, however, it is, in addition to increasing explanatory success, (almost) necessary that there are no new (types of)
should be conceded that explanatory and predictive success is in practice the employed criterion, at least as far as the internal phase of (explanatory) programs is concerned. For the external phase predictive success does not appear to be necessary, although this does not mean that only explanatory success is sufficient for progress. For the external phase another supplementary criterion to explanatory success is obvious: external success. From time to time the program should successfully solve external problems to which it is directed. Indeed, Nobel Prize motivations frequently report, in addition to (new) explanatory success, either predictive or external success.

An important case study undertaken by Zandvoort [1986] concerns the main theory transitions in the nuclear magnetic resonance (NMR-) program, which originates from nuclear physics and is based on quantum mechanics. He had to conclude that in almost all cases theory transition concerned theory accommodation on the basis of newly discovered facts. On closer inspection it also became clear that nobody doubted the possibility that the NMR-program could explain the new facts by some further articulation and hence that such doubts could not be the reason why the program was prolonged. In fact, the program was continued because it simultaneously solved important problems that were very relevant to other programs, in particular in chemistry and biology.

In this short overview of scientific progress we came across the following basic types of success in science: truth approximation, explanatory success, predictive success, external success. At least the following other types of success should also be mentioned: scientific recognition, textbook treatment, financial support, popular-scientific publicity, and institutional power. As is frequently claimed, and regretted, only the first two types are highly correlated with the basic types.

Although the foregoing is generally applicable to all explanatory programs, some distinctions may nevertheless only be made analytically. For instance, the application phase may well have been started when the internal phase has not yet come to an end. However this may be, the proposed distinctions may also be applied in an adapted form to other types of programs. For instance, for descriptive, design and explicative programs similar phases can be distinguished, at least analytically. Moreover, the exposition about progress also applies for example to explicative programs. For such programs the analogue of (intended) increasing explanatory success, is intended increasing explicative success, i.e., succeeding in explicating the informal concept in a more satisfactory way, as determined by the conditions of adequacy, along with the evident examples and counter-examples that have counter-examples, a condition which Popper and Lakatos just presuppose.

For practical (and ethical) reasons, predictive success may not always be possible. A paradigm case is the string theory program, in which, at least so far, theoretical predictions cannot be tested because “we will never be able to produce energies anywhere near [the required] value” [Atkinson, 2005, 100]. Atkinson considers for this reason the need of introducing a fifth fundamental kind of research program, that “aims at unification, that is the bringing together of apparently different explanations into one coherent logical or mathematical framework” (p. 102). See also [Psillos, this volume].
been put forward. However, it is also considered to be very important that the proposed explication turns out to give rise to unintended explications, that is, to satisfactory explications of related concepts and intuitions. This type of success is the analogue of the extra, i.e., predictive or external, success of explanatory programs. Again the question is whether this form of success is formally defensible as a necessary condition for progress, but the fact remains that in practice this type of explicative success plays an important role. For descriptive and design programs it is less clear whether there are similar criteria of ‘more than explicitly intended success’. In [Kuipers, 2001, Chapter 10] the plausible claim is developed that a transition in a design program is successful when a modified prototype satisfies more of the desired properties than the foregoing prototype, which evidently is the design analogue of explanatory success. Finally, in [Kuipers, 2000] some refined ideas are introduced in Chapter 5 concerning the successes and problems of theories which are particularly relevant for explanatory programs. Moreover, Chapter 7 introduces refined notions of successes and problems of descriptions; they are relevant for explanatory as well as descriptive programs.

3.2.5 The atomic theory as a developing explanatory program

The atomic theory and its development may well serve as an example of a successful research program. The following portrait is a simplification of the reconstructions given by Holton [1973] and Zandvoort [1989].

Dalton [1766-1844] introduced the theory of the atom in order to explain certain laws of chemical reactions and to possibly predict some other ones. Hence, it is an explanatory program. Its domain consists of reactions between chemical substances. Along with the development of that program the distinction between pure substances and mixed substances (mixtures) and the division of pure substances into elements and compounds emerged as observational categories. The following exposition will presuppose the idealization that these terms were without problems available to Dalton. The development of the program can be described in three phases.

Phase 1: Dalton’s primary problem was to explain certain relatively well established (observational) laws of reaction:

- \( LR1 \) (Lavoisier): the total weight of the substances before and after a reaction remains the same.
- \( LR2 \) (Proust: the law of definite proportions): compounds always decompose into components with constant weight ratios.

The core ideas of the program initiated by Dalton can be summarized into four principles, with the notions of (types of) atoms and molecules as the program specific theoretical vocabulary. The first two are internal principles, only dealing with postulated, hence theoretical micro-entities; the remaining two are bridge principles, in fact identity postulates, relating the theoretical terms to the observation terms.
1: *Atoms* are indivisible, unchangeable, hence indestructible, small material particles of a certain type.

2: Atoms are grouped into *molecules* of a certain type, and they may regroup into other types of molecules.

B1: - Pure substances consist of one type of molecule; in the case of elements these molecules consist of one type of atom; in the case of compounds they consist of more than one type of atom.
- Mixed substances consist of more than one type of molecule.

B2: Chemical reactions amount to systematic regrouping of the molecules of a substance.

Let us indicate the core idea consisting of I1 & I2 & B1 & B2 by C, and the specific theory at stage i by Ti, consisting of C + Hi, where Hi indicates the auxiliary hypotheses at stage i. Hence, T1 = C, as there are not any substantial auxiliary hypotheses at the start. It is not difficult to check that T1 explains the two target observational laws RL1 and RL2.

In agreement with both Popper’s and Lakatos’s views, Dalton felt obliged to obtain a more impressive result and predicted a third law, viz.,

**RL3** (the law of multiple proportions): when two different elements unite into two different compounds, the different proportions bear a simple relation to one another.

One successful test tuple of compounds consists of carbonic oxide and carbonic acid, both composed of carbon and oxygen. In the first case the proportion, in terms of weights, of oxygen to carbon is about 4:3, in the second case about 8:3. Hence, the ratio of these proportions is 1:2. For this prediction Dalton needed a strong rule of simplicity concerning the possible composition of molecules of the same type of atoms:

**A-s** (internal simplicity assumption): if a certain type of molecule exists then all the conceivable more simple types of molecules composed of the same type of atoms exist as well.

In combination with the assumption that the number of existing compounds of two elements is rather limited (an auxiliary hypothesis that will be neglected in the remainder of this section), RL3 can easily be derived. Although A-s is certainly false according to our present knowledge, the derived predictions came true and RL3 became accepted. Hence, according to the progress standards proposed by Popper and Lakatos, the transition from T1 to T2 = C + A-s is progressive: intended (or postdictive) explanatory success is supplemented with predictive explanatory success.
Phase 2: However, in the meantime a severe anomaly was arising: the law of combining volumes, independently established by Gay-Lussac. First an example: two liters of hydrogen gas and one liter of oxygen gas result into two liters of water vapor. In general:

\[ RL4 \] pure gases combine with simple integer numbers of volume units, into an integer number of volume units of the compound gas, not necessarily equal to the sum total of volume units of the component gases.

It is easy to see that it is not possible to derive \[ RL4 \] from \[ T2 \], nor its negation, for the simple reason that \[ T2 \] does not say anything about volumes of atoms and molecules. However, Dalton and some of his followers actually favored an auxiliary (bridge) assumption \( A-g \) about the nature of gases:

\[ A-g \text{ (gas assumption): gases consist of non-moving, contiguous gas particles, in their turn consisting of a molecule and a caloric mantle.} \]

It is quite obvious that the resulting \( T3 = C + A-s + A-g \), constituting a specific theory for the subdomain of gases, predicts the negation of \( RL4 \). Indeed, Dalton was confronted with a very big explanatory problem. However, he himself was inclined not to blame his theory, but to put the truth of \( RL4 \) into question. To be sure, the transition from \( T2 \) to \( T3 \) cannot be called progressive.

Avogadro (1776-1856) took the explanatory problem more seriously. He not only proposed a modified version of the rule of simplicity \( A-s^* \), but also a totally different and, at first sight, very surprising auxiliary hypothesis about gases:

\[ A-g^* \text{ (gas assumption proposed by Avogadro; Avogadro's hypothesis): equal volumes of different gases, at equal pressure and temperature, contain equal numbers of molecules.} \]

Note first that \( A-g^* \) typically is an extra bridge principle. Even without specifying \( A-s^* \), it is plausible that Avogadro could show that the resulting specific theory, \( T4 = C + A-s^* + A-g^* \), is indeed able to explain, in addition to \( RL1, 2, \) and 3, Gay-Lussac’s law of combining volumes, \( RL4 \). Moreover, it enabled him to produce, using examples of weight and volume ratios in the line of \( RL3 \) and 4, molecular composition formulas and molecular reaction equations. However impressive these results were considered to be, they did not lead to the general acceptance of Avogadro’s specific version of the atomic theory. Why? Referring once again to an insight developed by both Popper and Lakatos, the crucial point in this case is that, although \( T4 \) is explanatorily more successful than \( T2 \) and \( T3 \), it has, at this point, not yet achieved (specific) predictive success.

Phase 3: It took about half a century before Cannizaro (1826-1910) was able to derive a new prediction, using an additional auxiliary (bridge) hypothesis concerning dissociation:
A-d (dissociation assumption): large molecules of a certain composition will fall apart in a gaseous state.

From the resulting specific theory $T_5 = C + A-s^* + A-g^* + A-d$, using Avogadro’s findings on substances that consist of large molecules, the following prediction could be derived:

$RL_5$: substances so and so will dissociate in a gaseous state.

These predictions turned out to be largely correct, hence the transition from $T_2$, or $T_3$, or $T_4$ to $T_5$ is progressive in the sense of predictive success.

As a matter of fact, largely due to Cannizaro’s results, Avogadro’s hypothesis and the whole specific theory became generally accepted, i.e., its internal phase came to an end. It nicely illustrates that a research program can be very successful without a stable overall positive heuristic, which is difficult to discern in the case of the atomic program. At most there is something like a positive heuristic for partial use only, viz., to modify the rule of simplicity when problems arise involving molecular formulas. To be sure, $C$ seems to have been really a hard core, in the sense of Lakatos. However, for the sub-program restricted to gases, the transition from $A-g$ to $A-g^*$ may well be construed as a fundamental change of the core idea.

As is well-known, the atomic theory has been very successfully applied in other areas of scientific research, and has turned out to be indirectly of great practical use in chemical technology. Moreover, by accepting it, determination criteria for atoms and (the composition of) molecules where also established, and it not only became a background theory, but an observation theory.

This concludes our treatment of the example for the moment, but we will come back to it a number of times.

3.2.6 A pictorial summary

We conclude this section with a pictorial summary of its main content.

The concept of research tradition developed by Laudan [1977] can be interpreted as an even more global conceptual unit than that of a research program. As a matter of fact, a research tradition can be seen as the metaphysical and methodological core of a number of research programs. Behaviorism is a good example; it generated several research programs in psychology and biology. If we include this in a picture we get a double branched description of the state of affairs at a certain moment in a certain tradition, bringing together ideas from Kuhn, Lakatos, Sneed, Laudan, the Starnbergers and Zandvoort. In Figure 9 $RP_1/2/3$ denote research programs belonging to a certain research tradition, $CT_2$ the core theory of $RP_2$, $T_2.i,j$ a specific theory of $RP_2$, to be precise, the $j$-th attempt for the $i$-th subdomain. Moreover, $RP^*$ denotes a research program which may or, more likely, may not belong to the same research tradition, and $DP$ denotes a design program for some science external product. A thin arrow denotes ‘gives rise to’, a thick arrow denotes a transition with explanatory and predictive-or-external
success, and, finally, an dashed arrow indicates ‘tries to contribute to the solution of a problem of’ a certain (research or design) program.

3.3 Program internal research strategies

The above exposition of the structure and development of research programs raises at least two questions:

- how are new specific theories developed within a research program?
- in what ways can research programs interact?

These questions essentially concern research strategies for the (further) development of research programs. On the one hand there are program-bound strategies, that is, strategies within a single program aiming at improving the last specific theory. On the other hand there are strategies directed at interaction between programs for the benefit of at least one of them. Both types of strategies will be discussed in the indicated order. As far as the interactive strategies are concerned no detailed reconstructions of such strategies have been developed, and it is doubtful whether they ever will be. However, knowledge of their global nature is essentially sufficient for application in new contexts. The same is true for the program-bound strategies.

In the present subsection we start with a discussion of the importance of working within research programs, the program strategy itself. Then some more specific
strategies will be discussed, research guided by ‘idealization and concretization’ and by an ‘interesting theorem’. We will conclude with some remarks about descriptive and explanatory research programs in the light of truth approximation.

Here, as well as with respect to the interaction of programs in Subsection 3.4., we expose the global lessons to be drawn from the work of Popper, Kuhn, Lakatos, Sneed, Hamminga, the Starnbergers, and Zandvoort. Moreover, we integrate insights taken from Nowak [1980], Krajewski [1977], Darden and Maull [1977], and Bechtel [1988a/b]. See also Bechtel and Hamilton in this volume. It is more or less a strategy oriented synthesis of their insights as far as they are compatible. The reader should understand that the assertive tone concerning strategies should not be taken too seriously. As with all heuristics, you may consider a strategy, you may even try it out, but you cannot blame it.

3.3.1 The program strategy

From the long-term development of the sciences it has become clear that scientific research can be aptly characterized in terms of research programs. This does not alter the fact that the historiography of science can frequently be accused of concentrating too much on the success stories, on the successful research programs. Historiography should also pay much attention to programs that have lost the competition. The results of these programs are of course not added to the body of knowledge, but they rightly deserve a decent place in ‘the museum of knowledge’. Unfortunately, the only programs arriving in the museum of knowledge are those that were winning programs until superseded by newer programs. But deposition in the museum of knowledge should for instance also take place when two competing programs started more or less at the same time, with one of them having to give up sooner or later in favor of the other. Hence, the claim is, more precisely, that the main lines of the history of science can well be described in terms of rising, winning and falling research programs. If that history is written with any attention being given to mutual interaction, it will not only become apparent that programs frequently compete\textsuperscript{32}, but also that they may fruitfully cooperate.

The lessons from these observations seem to be the following. Anyone who wants to undertake frontier research,\textsuperscript{33} will in general also aim at getting the results of his research sooner or later incorporated in the international knowledge base, or at least in the museum of knowledge. Taking all things into consideration, and whether one likes it or not, in order to achieve this goal, participation in one or more of the internationally recognized research programs is virtually unavoidable.

Many university researchers consider this point to be obvious. However, there are nevertheless quite a few researchers who think otherwise. One seldom meets

\textsuperscript{32}For a brief history of the successive and competing research programs in high-energy physics, see [Cushing 1982]; for the history of the NMR-(nuclear magnetic resonance) program, engaged in asymmetric reductive cooperation with chemical and biological research programs, see [Zandvoort 1986].

\textsuperscript{33}For example, my home university, the University of Groningen, advertises with the slogan “to work at the frontiers of knowing”.
them in the natural sciences, but regularly in the human sciences and, for sure, in philosophy. As a consequence, the feature of having in a certain domain of inquiry only a few (interacting) research programs is badly developed in the human sciences. In the natural sciences this fruitful characteristic certainly has been partly instigated by the high costs of experimental research.

A frequent objection to program participation is the claim that it inhibits creativity. But the converse seems to be the case. Given the enormous potential of competitors one needs stronger creative talents to deliver a substantial contribution. Moreover, some critics are of the opinion that it should be possible to develop a new program. Hence, in the human sciences many researchers start their own shop, complete with a new publication medium. It is strange that such initiatives are usually taken rather seriously. It is interesting to compare this with a researcher who announces that he is going to make an important new discovery and who will be regarded rather skeptically. But the invention of new ideas that can lay the foundation of new research programs is as rare and difficult as making pioneering empirical discoveries. In fact it concerns pioneering theoretical discoveries. For both types of scientific achievements Nobel prizes are awarded. Moreover, the inventors of ideas for new programs are frequently steered by a fresh look at the severe problems they and others met when conducting research in existing programs.

In the social sciences and the humanities there even seems to be an abundance of research ideas. One therapy against this condition is the foregoing type of plea for program-bound research, the other is a plea, which is to follow in Subsection 3.4., for stimulating interaction between research programs, by cooperation and/or competition.

3.3.2 Program development guided by idealization and concretization

We will now sketch two strategies for the internal development of a research program, particularly for the succession of improving specific theories. To be sure, these strategies can also be used without assuming the boundaries of a research program, the only claim is that they are frequently used within a program. We start with idealization and concretization. Idealization is frequently applied in empirical scientific practice as an unavoidable step in theory formation. This is certainly true in the natural sciences; in the human sciences and also in philosophy the necessity of explicit idealization is not yet generally accepted.

Surprisingly enough, on closer inspection Marx developed his ideas in Das Kapital rather systematically according to the method of idealization and successive concretization. Nowak [1974; 1980] has pointed out that this procedure was used by Marx; in particular he shows how Part I and Part III in their succession can be seen as illustrations of what Marx used to call ‘rising from the abstract to the concrete’. Another Polish philosopher, Krajewski [1977], freely following Nowak, has also contributed importantly to the growing awareness of the systematic role of what he calls ‘idealization and factualization’.
Although idealization-and-concretization (from now on, I&C) also occurs in qualitative theorizing, it is primarily explicated for quantitative theorizing, in particular the succession of specific theories within a research program. The general idea is that it is frequently possible to make an ordering in the degree of importance or relevance of all the factors that influence the value of a certain quantity $G$, which may even lead to a division of primary and secondary factors. Starting from such an ordering of factors $f_0, f_1, ..., (fm)$, in the $n-$th stage of concretization factors $f_0$ up to $fn$ have been accounted for, while the remaining factors are still neglected, leading to the typical I&C-formulation of the $n$-th specific theory:

$$\text{if } f_0 \neq 0, f_1 \neq 0, ..., fn \neq 0 \text{ and } f(n+1)=0, f(n+2)=0, ....$$

$$\text{then } G = Gn(f_0, f_1, ..., fn)$$

In the 0-th stage there is maximal idealization and when all factors have been concretized, maximal concretization has been achieved. Note that, although any given functional representation of a factor is allotted the value 0 on a formally arbitrary basis, the neglect of a certain factor is empirically speaking usually not arbitrary, in which case the functional representation can be chosen in accordance with this.

The transition of the ideal gas law to the Law of Van der Waals is a paradigm case. This transition can be represented in a stepwise decomposition, of which the crucial formulas include:

$$(0) \quad P = RT/V$$

$$(1) \quad P = RT/V - a/V^2 \text{ (or, alternatively, } P = RT/(V - b))$$

$$(2) \quad P = RT/(V - b) - a/V^2 \text{ (or, standard form: } (P + a/V^2)(V - b) = RT)$$

where $P$, $V$, $T$ and $R$ indicate pressure, volume, temperature and the ideal gas constant, respectively, and $a$ and $b$ refer respectively to specific gas constants related to mutual attraction between the molecules and the volume of the molecules.

The book series Poznan Studies for the Philosophy of the Sciences and the Humanities (since 1990 with a subseries on idealization, e.g. [Nowakowa, 1994]) includes many examples of I&C taken, in particular, from physics, biology, economics and sociology.

I&C can be used to structure theories in their research stage as well as in textbooks. Although it seems very plausible to do so, to say the least, it is very surprising that it seldom is explicitly done. However, in general expositions about what one has been doing or how one should do it, there is frequently reference to I&C. A specific reason for the relative neglect of I&C in the social sciences may be the great social pressure to avoid very strong idealizations: fear of being accused of distorting reality too much seems to be very rampant.

The above mentioned paradigm example raises a very interesting question concerning explanations: is it possible to (re)construct the explanation of a concretized
law as a concretization of the explanation of the (more) idealized law? [Kuipers, 2000, Chapter 10] deals with this question in some detail.

Another, at least as important, question is whether and in what sense the I&C-strategy is functional for truth approximation in the empirical sciences. A detailed positive answer is given in [Kuipers, 2000, Subsection 10.4.] and in [Kuipers, forthcoming] that answer is given as a paradigm illustration of concept explication by the I&C-strategy, that is, ‘conceptual I&C’ as a conceptual twin of ‘empirical I&C’. In general, as already mentioned, the I&C-heuristic can also be used in the qualitative theorizing that often occurs in mathematics and philosophy. The ordered textbook presentation of first propositional logic and then predicate logic provides a famous example. In Section 2 the I&C-heuristic has been used to present the structuralist approach to theories.

3.3.3 Program development guided by interesting theorems

Hamminga [1983] made a related strategy of theory development in economics explicit. The time when economists thought that economics could and should in principle be done along naive Popperian lines has passed, but the question remains how economists in fact do their job.

Hamminga studied the development of the theory of international trade in the period 1930-1970 and reached the following diagnosis. Economists direct their attention to theorems that they find interesting and they try to prove their validity for an increasing number of conceivable cases. Probably they have the following motive in the back of their mind: the desire to increase the plausibility that the theorem also holds in the actual world (or the nomic world, see Subsection 1.5.3.). Apart from this motive, the world does not play a clear role: it is all and only mathematics, or so it seems. Nevertheless, or precisely because of this, one can find a large amount of systematics in the details of what theoretical economists do.

To begin with, it is possible to systematize the specific claims of the research program to an even greater extent than Lakatos could have imagined: under such and such conditions it is possible to prove the interesting theorem (IT), or in a formal schematization:

\[ V_{lmn}(C_1 \cdots C_i; C_{i+1} \cdots C_j; C_{j+1} \cdots C_k \rightarrow IT) \]

The division of conditions here is as follows. \( V_{lmn} \) indicates the field conditions that describe the domain of the claim; in the example, \( l, m, \) and \( n \) indicate, respectively, the number of countries, goods and production factors taken into consideration. \( C_1, \ldots, C_i \) indicate the generic or basic principles, i.e., the core ideas of the general research program with which the problem area is attacked. In the case of international trade this basic program is that of neo-classical economics, of which the core consists of utility theory. \( C_{i+1}, \ldots, C_j \) indicate the special or specific principles for the particular subdomain of international trade, e.g., that the production functions are the same in all countries, while the endowments of production
factors may vary greatly. Finally, $C_{j+1}, \ldots, C_k$ indicate the technical conditions of a mathematical nature, e.g., that the production functions are continuous. An example of an interesting theorem is that of factor-price-equalization: the price of a certain factor becomes equal in the dynamic equilibrium with international trade.

Theory development or, more precisely, results that are considered to be important theoretical achievements consist of both new specific claims and their proof. In the latter, field and/or technical conditions have been liberated in one or more of the following ways:

- field extension: increasing the number of countries, goods or factors (2 is for each the point of departure)
- weakening technical conditions
- substituting more plausible conditions for technical conditions
- introducing alternate technical conditions

In [Kuipers, 2000, Subsection 10.4] such developments have been shown to be formally similar to concretization. They are therefore functional for truth approximation.

The picture that Hamminga draws seems representative of neo-classical economics; reconstructions of the theory of the market [Janssen and Kuipers, 1989] and of capital structure theory ([Kuipers, 2000, Subsection 11.2], based on [Cools, Hamminga, Kuipers, 1994]) confirm this diagnosis. Of course, Hamminga does not provide a complete picture of the whole of the science of economics; in particular applied econometric models are overlooked. His views do however characterize an important part of the discipline, viz., so-called theoretical economics. Moreover, his work suggests the question of what the systematics is, if anything, in theory development in those areas of economics where the picture drawn seems inadequate.

The sketched diagnosis of the mathematical nature of economics may be an important underlying motive for the striking ambivalence of economists about the question of whether economics is an empirical social science or not. Moreover, the diagnosis illustrates that the cognitive aims of the social sciences in general and of economics in particular appear to be less evident than philosophers of science use to assume on the basis of an analogy to the natural sciences.

To be sure, the described strategy is certainly not restricted to economics. For instance, in population biology similar strategies are used with respect to the Law of Hardy-Weinberg [Lastowski, 1977]. In mathematics and philosophy the ‘theorem-strategy’ is also frequently used. Aiming at soundness and completeness proofs for increasingly complex logical systems is well known. Another example is the ‘success theorem’ in [Kuipers, 2000, Chapters 7-10]. It amounts to the claim that a theory closer to the truth than another will also be more successful. It can be proved under more and more realistic conditions.
3.3.4 Descriptive and explanatory research programs in the light of truth approximation

As a matter of fact, truth approximation analysis in general ([Kuipers, 2000], see also [Niiniluoto, this volume]) enables a richer perspective on the nature of descriptive and explanatory research programs. Recall that such programs presuppose, by definition, at least a domain, a problem, and a core idea, including a vocabulary, to solve that problem. A descriptive research program uses an observational conceptual frame, and may either exclusively aim at one or more true descriptions (as for example in most historiography), or it may also aim at the true (observational) theory. In the latter nomological type of descriptive program, however, the goal of a true theory is supposed to be achieved exclusively by establishing observational laws. Given that this requires (observational) inductive jumps, it is plausible to call such a descriptive program an inductive research program. Not surprisingly, such programs ‘approach the truth by induction’. For the establishment of observational laws, testing the relevant general observational hypothesis will result in true descriptions that either falsify the hypothesis or are partially derivable from it. According to the basic qualitative definition of ‘more truthlike’, assuming that accepted observational laws are true, any newly accepted observational law guarantees a step in the direction of the true theory. Hence, inductive research programs are relatively safe strategies of truth approximation: as far as the inductive jumps happen to lead to true accepted laws, the approach not only makes truth approximation plausible, it even guarantees it.

Let us now turn to the explication of the nature of explanatory or theoretical programs, which are by definition of a nomological nature. An explanatory program may or may not use a theoretical vocabulary. Even (nom) empiricists can agree that it is directed at establishing the true observational theory. If there are theoretical terms involved, the referential realists will add that it is also directed at establishing the referential truth. The theory realist will add to this that it is even directed at establishing the theoretical truth. Scientists working within such a program will do so by proposing theories respecting the hard core as long as possible, but hopefully not at any price. They will empirically evaluate these theories separately and comparatively. Theory choice will be governed by being persistently more successful, which is trivially functional for empirical progress. However, although that ‘rule of success’ is, moreover, demonstrably functional for all distinguished kinds of nomic truth approximation, it cannot guarantee, even assuming correct data, a step in the direction of the relevant truth. Though the basic notions of successfulness and truthlikeness are sufficient to give the above characterization of the typical features of explanatory research programs, they usually presuppose refined means of comparison, which are presented in Part IV of [Kuipers, 2000].

34In this terminology, ‘inductive’ is supposed to exclusively refer to observational induction and not to theoretical induction, as distinguished in Subsection 1.2.2.
3.4 Interaction between programs

Of the two general types of interaction between programs, competition and cooperation, successful interdisciplinary research seems to result as a rule from a special kind of asymmetric cooperation between research programs. Moreover, it may or may not be a matter of cooperation between a holistic and a reductionistic program. Finally, some educational lessons from the program-bound and interactive research strategies will be drawn.

3.4.1 Interaction between programs as a research strategy

It is plausible to distinguish two main types of interaction between research programs, viz., interaction by competition and interaction by cooperation. Of course, it is also possible that after a period of interaction of one kind, the interaction turns into one of the other kind.

We will first concentrate on competition. When two programs are directed at the same domain and problem, and both are still in the internal phase, competition will concern the adequacy of the core ideas. When both programs are already in the external phase, competition concerns the question of which program is best suited to solving problems external to science or problems raised by a third program. When one program is still in the internal phase and the other in the external phase, competition usually takes the form of a challenge by the first to the supposed domain of validity or degree of accuracy of the second. A well-known example of the last kind is Einstein’s questioning of Newton’s theory.

The three indicated types of competition can all be very stimulating. However, when competition occurs, it is seldom seen as an explicit research strategy. Interestingly enough, one is not even always aware of being steered by a competing program, or one is at least not willing to admit that this is the case. These facts explain why the question of whether a further articulation of a competing program may lead to even more stimulating interaction is not always raised.

Population genetics provides a nice example of competition between two programs in the internal phase [Dolman and Gramsbergen, 1981]. Concerning the problem of the origin and dynamics of variations in populations two programs can be distinguished, viz., the classical and the equilibrium program. The development of both programs cannot be described without bringing the stimulating interaction between them into the picture. Moreover, they gradually show a remarkable convergence, with the consequence that the competition increasingly transforms into cooperation in such a way that a fruitful synthesis has emerged. The same development, initiated by, among others, Smolensky [1988], has taken place in the interaction between symbolism and connectionism in cognitive science.

Now let us consider cooperation. As in the case of competition, the forms of cooperation can be divided according to the three combinations of phases in which the two programs are situated. We have already seen that a program in the external phase can offer its services to another program, in the internal or external phase, which is confronted with a problem that the program itself is unable to
solve. In Zandvoort’s appealing terminology [1986; 1988; 1995] the latter program then functions as the guide program and the former as supply program. The core theory of the supply program may either be specialized (finalized) to the domain of the guide program, or it may be used as observation theory providing relevant observations for the guide program. For the particular problem the cooperation is of a fundamentally asymmetric nature. This character does not exclude the fact that the roles can be interchanged in dealing with another problem, in particular when not only one but both programs are in the external phase.

A typical form of the type of asymmetric interaction is provided by design research programs in the internal phase. They frequently function as guide programs, with descriptive and explanatory programs in the role of supply programs.

Besides the foregoing type of cooperation, in which at least one program is in the application phase, cooperation is possible between two programs in the internal phase, in which case they frequently stimulate each other in rotation, alternating in the role of guide and supply program. In this case, as in the case of two programs in the external phase, the cooperation is (although with respect to specific problems asymmetric) on the whole symmetric: the programs co-evolve [Bechtel, 1986; 1988a; Bechtel and Hamilton, this volume].

For instance, at least on the basis of accounts given in physics textbooks, one easily gets the impression that the interaction between phenomenological thermodynamics and statistical mechanics is a classical example of this type of cooperative co-evolution. However, it is well known that the intentions of the researchers concerned were of a much more competitive nature. Apparently, this does not exclude the fact that the result of competitive interaction can make it plausible that the intention to cooperate could have been at least as productive.

The example shows that researchers themselves may be inclined to perceive the interaction between two explanatory programs as competitive rather than as cooperative. However, when one program evidently is of a descriptive nature and the other of an explanatory nature directed at the first, the interaction between them can easily be conceived by the researchers as cooperative: it is a paradigmatic kind of co-evolution.

In the foregoing the basic aim of cooperation between programs was the solution of a problem encountered in one of the programs by the other. Of course, other goals of cooperation occur. For example, programs may jointly strive to articulate an overarching theory, or a synthesis of theories. Recall that the latter was the case in the example taken from population genetics. Still one other important form of cooperation involves bridging the gap between two theories, requiring a third so-called interfield theory [Darden and Maull, 1977; Bechtel, 1988a; Kuipers, 2001, Chapter 6; Bechtel and Hamilton, this volume].

For all mentioned kinds of interaction the programs may be empirical programs of the same or different type. Moreover, interaction may also involve an empirical program of a certain type and an explicative program of a philosophical or mathematical nature. Current interactive researches within cognitive science and between cognitive and neuroscience are of the latter nature. Some of the cur-
rent philosophical mind-brain-body theories and theories of representation are not only of an explicative nature, they play at least some interactive role with some empirical programs [Bechtel, 1988a/b; Bechtel and Hamilton, this volume].

3.4.2 Interdisciplinary research

It is worthwhile to consider Zandvoort’s model of cooperation in tandem with his model for successful interdisciplinary research [Zandvoort, 1986; 1988; 1995].

IR-model: interdisciplinary research consists of some research programs, belonging to one or more disciplines, cooperating according to the following rules of the game:

- one program is the guide program which raises problems of a theoretical or experimental nature in the others,
- the other programs are supply programs, which have successfully passed their evaluation phase and hence can try to solve the problems provided by the guide program.

Compared with the popular ideas about interdisciplinary research, the above model has three fundamental differences. First, interdisciplinary research is not so much a matter of global cooperation between disciplines but, more specifically, cooperation between research programs. Second, it is a matter of asymmetric cooperation: one program poses the problems, the others try to supply solutions, and if successful they have the last word. Third, effective supply programs typically are in the external phase. Note that, if the guide program also has already passed its evaluation phase and is not a design program, then it will usually be directed at a science external problem of a technological or societal nature.

The IR-model suggests that the failure to start successful interdisciplinary research may well be due to the lack of relevant supply programs in the external phase. Moreover, it may be due to the collision of cognitive and social factors: in addition to the necessity of ‘cognitive asymmetry’ there is an inclination to as much ‘social symmetry’ as possible: all participants are supposed to deliver contributions of equal importance.

The stress on asymmetric cooperation between programs needs a counterbalance on the level of disciplines. It is conceivable that all interdisciplinary research directed at some science external problem area (e.g., health, environment, education) develops into a state in which one discipline provides all the guide programs, whereas the other participating disciplines provide only supply programs (the hierarchical model). Alternatively, it is also possible that there arises on the level of disciplines a more symmetric situation (the symmetric model). On the level of science and research policy, when setting up long-term strategic interdisciplinary research in some science external problem area, it seems very important to start with the symmetric model. The reason is that it is easy to imagine that starting from the symmetric situation purely scientific reasons may gradually lead to a
hierarchical situation, whereas it will be much more difficult to reach a symmetric situation starting from a hierarchical one, let alone to reverse that hierarchy. For reasons of completeness we conclude by noting that the IR-model does not seem to be appropriate as a point of departure for the investigation of a science external problem area when one wants to obtain short-term practical results. In planning that type of research ad hoc considerations seem to be unavoidable.

3.4.3 Interaction of holistic and reductionistic research programs

One of the most exciting forms of competition and cooperation occurs between reductionistic and holistic research programs involving essentially the same domain. In most cases the interaction can be described in Zandvoort’s terms of a guide program and one or more supply programs. Moreover, in one sense or another the guide program is reduced, i.e., there is a reduction of concepts, or laws, or both. In this case, the guide program is called holistic and the supply program reductionistic, terms which are of course relative qualifications. The reduction of laws and concepts may or may not conform to one of the kinds of reduction distinguished in the pluralistic models for the reduction of laws and concepts that has been presented in [Kuipers, 2001, Chapter 3 and 5, respectively].

None of these basic forms of reductive interaction implies the elimination of the guide program. One of them only amounts to one or more corrections in the tentative laws that guided the research. The other two are variants of a non-eliminative reduction of a higher to a lower level. In these cases the laws and concepts of the higher level of the guide program are reduced to laws and concepts of lower levels, without explaining away the higher level.

Below we give a number of illustrative examples of studies of interacting research programs, where reductionistic and holistic perspectives play a major role. All studies mentioned are related to Groningen. Although they represent a type of ‘cognitive studies’ in philosophy of science which is typical for Groningen (see [Kuipers and Mackor, 1995]), numerous examples from elsewhere could, of course, be given. In a detailed study Janssen [1993] analyses the so-called micro-economic foundation of macro-economic concepts and laws. The current micro-foundation is a non-reductionistic attempt to interaction between the descriptive guide program of macroeconomics and the explanatory program of neo-classical microeconomics, in particular general equilibrium theory. According to Janssen the results are problematic because the supposed individualistic foundation of general equilibrium theory is doubtful. He sketches another way to explain macro- and micro-economic laws and theories in a strictly methodological-individualistic way. In this approach game-theoretic adaptation of utility theory plays a crucial role: it serves as a supply program.

Looijen [1995; 1998/2000] investigates the structure and dynamics of ecological communities. His working hypothesis is that the cooperation between three kinds of research programs could be improved. On the extremes sides one there are holistic guide programs describing the structure and dynamics of communities and
radical reductionistic programs that try to explain these patterns in terms of the species composing the communities and their environmental needs could. They could well cooperate with moderately reductionistic (or moderately holistic, for that matter) programs that try to explain these patterns using theories about the interactions between the composing species, such as predation and competition.

Guichard [1995; 1997] starts by documenting that stress researchers, though striving at cooperation between psychological and biological research programs, were unsuccessful in the strategies employed to achieve this end. The reason for this systematic failure seems to have been that these strategies essentially presupposed a dualistic explication of the mind-body problem. Guichard argues that monistic explications, in particular of a materialistic-reductionistic nature, are more appropriate for such cooperation. His intervention uses a philosophical explication program to get new perspectives for cooperation between (relatively holistic) psychological guide programs and (relatively reductionistic) biological supply programs. More specifically, Guichard argues that the proper function theory of Ruth Millikan provides the ideal ‘interfield theory’ for this purpose.

Mackor [1995; 1997] argues along the same lines, but more generally, that a conceptual unification is possible not only between biology and psychology but that this can be extended to sociology by analyzing meaningful and rule-guided behavior in terms of Millikan’s notion of proper functions. The result is a new, naturalistic, mildly reductionistic perspective on the spectrum of disciplines and their possibilities of cooperation. From this perspective, the most important boundary between the sciences runs between physics and chemistry on the one hand and biology on the other, although, successful cooperation between them is never excluded.

Festa [1993; 1995] initiates a confrontation between three research programs that were developing almost independently: inductive logic or inductive probability theory (Carnap), truth approximation (Popper) and Bayesian statistics. In the first place he shows that inductive logic, despite Popper’s dismissive attitude about it, can be considered as a part of the truth approximation program, directed at the approximation of the true, objective chances. In the second place Festa elaborates the claim that, using De Finetti’s representation theorem, (relatively holistic) systems of inductive probabilities can be reduced to special types of Bayesian statistics, which can hence be used as a supply program for further development of inductive logic. Last but not least, he shows that it is possible to define an optimum inductive system in terms of the available prior information about the domain to be investigated.

The general outline and the examples suggest at least three different research strategies for attacking a domain on the macro-level of some macro-/micro-level distinction. The radical holistic strategy is to try not only to describe but also to explain the phenomena at the macro-level in terms of that level, or even higher levels, and to refrain from lower level theories. The radical reductionistic strategy is to try not only to explain but also to describe the macro-phenomena in micro-terms as much as possible. The third strategy is a mixed one: according to the
mixed strategy, one describes the macro-phenomena and their possible relations in macro-terms, and tries to explain them in micro-terms as far as possible, and hence in macro-terms as far as necessary.

These three strategies suggest equally many philosophical (ontological cum epistemological) positions with respect to a certain ‘macro-domain’. They are formulated as general statements about the possibility of reduction of the concepts and laws of that domain. Radical reductionism is the belief that every macro-concept and macro-law can be reduced. Radical holism is the belief that all (interesting) concepts and laws of the domain cannot be reduced. Finally, restricted reductionism (and holism!) is the belief that some concepts and laws may be reducible, but others may not be.

It is important to note that the terminologically corresponding strategies and positions are not strictly coupled, except perhaps that radical philosophical holism leaves only room for the radical holistic strategy. The converse is not self-evident. There are excellent examples of research according to the radical holistic strategy, e.g., phenomenological thermodynamics and macroeconomics, where it is seen as a compatible, but separate task to try to reduce the macro-concepts and -laws, i.e., to work according to the radical reductionistic strategy. To be sure, it has to be conceded that reductionistic strategies in general have been very successful in the history of science.

However, the radical reductionistic strategy often leads to impressive minute research, which, however, nobody is waiting for. On the other hand, the radical holistic strategy frequently degenerates into hardly testable and transferable insights.

In line with these roughly formulated impressions, it is plausible to formulate the working hypothesis that the mixed strategy will in many cases be the best strategy. For, to reduce concepts and laws of a certain domain, they have first to be established. In its turn, it is frequently the case that the search for concepts and laws has been considerably stimulated by reductionistic questions.

The mixed strategy provides an important form of interaction of research programs. In the case of reductive interaction the guide program is of course a program on the macro-level, whereas at least one supply program is supposed to deal with the micro-level. When the supply program is, like the guide program, in the internal phase, the reductive interaction is symmetric, when it is in the external phase it is asymmetric.

Zandvoort’s paradigm example of a supply program in the natural sciences is the NMR(nuclear magnetic resonance)-program, engaged in asymmetric reductive cooperation with chemical and biological research programs. An important asymmetric example in the social sciences is the utility maximization or rational choice program. It has proved its strength in microeconomics and nowadays it cooperates with guide programs in macroeconomics [Janssen, 1993] and macro-sociology (explanatory sociology). An historical example of the symmetric reductive type is the interaction between phenomenological thermodynamics and statistical mechanics.
3.4.4 *Program pluralism as an education and research strategy*

Program-bound research has one main disadvantage. One can readily become very indoctrinated with a program. The postgraduate schooling in program-bound research may well degenerate into the delivery of program-bound researchers. As a counterbalance to the importance of program-bound research, a program pluriform education and subsequent research career seems equally important.

Based on the work of Kuhn and Lakatos it may be inferred that mature science frequently consists of dogmatic research, i.e., research sticking to the hard core of ideas, executed by dogmatically inclined researchers. Although such practices can be rational provided one aims at progress within the boundaries of the program, it is also our conviction that science would profit still more if non-dogmatic researchers perform dogmatic research.

The way to learn to do program-bound research, without becoming a prisoner in one program, is to get research experience in at least two programs. They need not be competing programs. According to Kuhn, such program pluralism is almost impossible, due to the Gestalt-switch that it is claimed to require. Although this thesis, together with the so-called incommensurability thesis, has been criticized severely ([Franklin, 1986; Hintikka, 1988], to mention a few), it is important to stress that it may be at least as instructive to undertake research in two programs that might cooperate.

The short-term effects are very positive. To begin with, when one alternatingly does research in two programs, the period engaged in one program may function at the same time as a form of productive breathing space for the work on the other. It may even occur, to say it in popular terms, that the right hemisphere is further stimulated to do its work. For, as suggested by the literature on creativity and serendipity [Van Andel, 1994], successful, or even unsuccessful, attempts at problem solving in one program may be transformed into successful solutions for the problems of the other.

This short-term favorable effect concerns the stimulation of largely unconscious processes, the second to be mentioned results from the conscious stimulation of interaction. If two programs have something to do with each other, knowledge of both leads in a natural way to questions of cooperation or competition. For instance, one may ask whether one program may be of help for the other. Can it solve a problem the other is unable to solve? Additionally, competition questions may arise. If on closer inspection both programs essentially claim to be able to solve the same problem, and if the one has already succeeded in doing so, the other cannot remain behind.

At first sight one may find the second favorable short-term effect of doing research in two programs unimportant. For competition and cooperation questions also may arise in the research team, or in the study of the literature, as specific research questions may be transmitted from one individual researcher to the other. In theory such exchanges should occur, but in practice they are limited, because it is understandable and even productive that one primarily works in program-bound
research groups. Moreover, as far as competition questions are concerned, there is the additional fact that it is only attractive to raise such questions when the researcher has acquired affinity with both programs. Hence, it seems plausible that the suggested interaction questions are best stimulated by promoting research training and further practice in more than one program.

The plea for pluralistic research experience should not, however, be misunderstood as a plea for unlimited and diffuse eclecticism. On the contrary, it is a plea for experience within and interaction between a limited number of (analytically) well-distinguished research programs.

The favorable long-term effects of this plural-program-bound research experience also seem advantageous. To join new developments in international research it is of great importance that one has learned about methodological multiplicity and the various perspectives from which one can investigate the world, not only in theory but also in practice. Plural-program-bound research experience stimulates the flexibility of the individual scientist in his further research career. The desirability of this flexibility for scientific research in general and interdisciplinary research in particular is obvious.

CONCLUDING REMARKS

In this last section we have argued that the notion of a research program is very useful in globally describing scientific research and in specifying global research strategies. In Sections 1 and 2 some of the main products of scientific research, usually resulting from isolated or cooperative (descriptive and explanatory) programs, have been studied in some detail, viz., laws and theories. Let us finally come back to the relation between the main subjects of the first and the third section, that is, the distinction between observational laws and proper theories, i.e., the law-distinction, and research programs, respectively.

It is the (relatively) short-term dynamic role of the law-distinction that relates to the development of research programs. The first context in which the law-distinction is used in short-term dynamics is just one explanatory research program, more in particular a (proper) theoretical program. Here a proper theory is revised on the basis of its successes and failures in explaining and predicting observational laws. The second context is the interaction between a theoretical program and a descriptive program. Here the two programs develop hand in hand, each challenging the other with new results. On the one hand, there are potential observational laws that have been predicted by the theoretical program and which have to be tested within the descriptive program. On the other hand, there are observational laws that have been established independently, in particular inductively, within the descriptive program (which may be called to that extent an inductive program) and that have to be explained by the theoretical program.

35For example [Kuipers, 2000] grew out of research and teaching in at least two rather different, if not opposing, research programs, the Carnap-Hintikka program of inductive (confirmation) logic and Popper's program of truth approximation.
Finally, the third context is of course the competition between two theoretical programs claiming to be superior with respect to the explanation and prediction of observational laws for a certain domain. Of course, the latter laws may or may not hang together within a descriptive program. In all three contexts the relevant theoretical programs are in the internal phase, whereas the relevant descriptive programs may or may not already be in the external phase.

Although the law-distinction may seem at first sight to be parallel to the different internal goals of descriptive and explanatory research programs, it is actually not so. To be sure, proper theories cannot be the internal goal of descriptive programs. However, the true description, aimed at by a descriptive program, may concern the true general description, i.e., the true observational theory in the sense of the (relative to the program) complete set of observational laws of the domain, which is an improper theory. The complicating fact is that this true observational theory may also be the true theory at which an explanatory program is aiming, in order to explain and further predict certain observational laws, usually those restricted to some sub-vocabulary of the explanatory program. In terms of the distinction between theoretical and (non-)theoretical terms, the failing parallel is even easier to formulate. Whereas a descriptive program by definition can not introduce new theoretical terms, an explanatory program may introduce such terms, and then be called a theoretical program, but it need not introduce such terms.

ACKNOWLEDGMENTS

The Sections 1, 2, and 3 are strongly based on Chapters 2, 12, and 1, respectively, of (Kuipers, 2001). The final version profited a lot from the critical remarks of Martin Mahner.

BIBLIOGRAPHY


[Schurz, 1990] G. Schurz. Paradoxical consequences of Balzer’s and Gähde’s criteria of theorit-