THEO A.F. KUIPERS

# NAIVE AND REFINED TRUTH APPROXIMATION*

ABSTRACT. The naive structuralist definition of truthlikeness is an idealization in the sense that it assumes that all mistaken models of a theory are equally bad. The natural concretization is a refined definition based on an underlying notion of structurelikeness.

In Section 1 the naive definition of truthlikeness of theories is presented, using a new conceptual justification, in terms of instantial and explanatory mistakes.

In Section 2 general constraints are formulated for the notions of structurelikeness and truthlikeness of structures.

In Section 3 a refined definition of truthlikeness of theories is presented, based on the notion of structurelikeness, using a sophisticated version of the conceptual justification for the naive definition.

In Section 4 it is shown that 'idealization and concretization' is a special kind of potentially refined truth approximation.

## INTRODUCTION

The idea of truthlikeness or verisimilitude is that one theory can be closer or more similar to the truth than another. In 1974 Miller and Tichy proved that Popper's original definition was inadequate, for it did not leave room for false theories, i.e., theories having empirical counterexamples. In Kuipers (1982, 1984) I presented a naive structuralist definition of truthlikeness that left room for empirical counterexamples and that was moreover attractive in other conceptual, logical, and methodological respects. However, this naive definition is based on the assumption that the mistakes of the other type, which a theory can make, i.e., allowing mistaken models, are all equally bad. For this reason the naive definition does not seem to have real-life scientific examples, for in cases of scientific progress a theory with mistaken models is usually replaced by a theory with less mistaken, but nevertheless mistaken, models. A paradigmatic case is the theory resulting from a concretization of an idealized theory.

A sophisticated definition of truthlikeness should hence not only account for empirical counterexamples but also for the fact that one mistaken model may be more similar to a required model than another. For then there is room for improving a theory by introducing new, but fewer, mistaken models. Of course, a sophisticated definition should

*Synthese* **93**: 299–341, 1992.

reduce to the naive definition under the relevant assumptions. Finally, it should retain the attractive logical and methodological features of the naive definition.

In Section 1 the naive definition of truthlikeness of theories is presented, using a new conceptual justification, in terms of instantial and explanatory mistakes. It will briefly be shown what plausible formal properties it has, how it can explain established naive success differences, how it can justify methodological rules, how it works out for stratified theories, and which plausible quantitative version it has.

In Section 2 general constraints are formulated for the notions of structurelikeness and truthlikeness of structures. A number of specific examples are also given.

In Section 3 a refined definition of truthlikeness of theories will be presented, based on the notion of structurelikeness, using a sophisticated version of the conceptual justification for the naive definition. It will be shown that refined versions of all merits of the naive definition follow.

In Section 4 it is shown that 'idealization and concretization' is a special kind of potentially refined truth approximation. This is illustrated by Van der Waals's theory of gases. Moreover, it is indicated how idealization and concretization can function as a strategy in validity research around 'interesting theorems'.

## 1. NAIVE TRUTHLIKENESS OF THEORIES

### 1.1. *Preparations*

Let there be given a domain D of natural phenomena (states, situations, systems) to be investigated. D is supposed to be circumscribed by some informal, intensional description and may be called the primitive set of intended applications. Let there also be given a set Mp of *conceptual possibilities* or potential models designed to characterize D. It may be assumed that Mp is, technically speaking, a set of structures of a certain similarity type. In practice Mp will be the conceptual frame of a research program for D.

The confrontation of D with Mp, i.e., D seen through Mp, is assumed to generate a unique, time-independent subset $Mp(D) = T$ of all Mp-representations of the members of D, to be called the Mp-set of intended applications or the (Mp-)set of physical or *empirical possibilities*.

This assumption will be called the *frame-hypothesis* associated with ⟨D, Mp⟩. As a consequence, Mp − T contains the relevant empirical *im*possibilities. As a rule, T is unknown, or even the great unknown and hence the target of theory-directed research in the domain.

It is clear that T is Mp-dependent, hence T is *conceptually relative*. It is also clear that T depends on reality through D. However, it does not represent 'the actual world', i.e., some actual state, situation, or system, but 'the set of empirically possible worlds' (as far as D is concerned). For that reason, the present type of realism may be called *theoretical realism* instead of descriptive realism.

A theory is any combination of a subset X of Mp and the claim that T is equal to X, and will be briefly indicated by 'theory X' or just 'X'. Members of X are called models of theory X. Theory X is true or false when its claim 'T = X' is true or false, respectively. According to this definition there is only one true theory, viz., theory T itself. Hence T may be called 'the true theory' or 'the theoretical truth' or even 'the truth'.

T can easily be interpreted as 'the strongest law'. A (general) hypothesis is defined as the combination of a subset X of Mp and the (weak) claim that T is a subset of X (i.e., all empirical possibilities satisfy the conditions of X). Hypothesis X is true or false when its claim 'T ⊆ X' is true or false, respectively. Members of X are now also called models of hypothesis X. A true hypothesis is also called a law.

If Y is a subset of X, the claim of hypothesis Y implies the claim of hypothesis X. In that case hypothesis Y is also said to imply hypothesis X, in agreement with standard model-theoretic usage to say that a statement S1 logically implies the statement S2 iff the models of S1 form a subset of those of S2. If hypothesis Y implies hypothesis X, i.e., Y is a subset of X, it will not only be said that hypothesis Y (X) is stronger (weaker) than hypothesis X (Y), but also that theory Y (X) is stronger (weaker) than theory X (Y), although the full claims of these theories are mutually incompatible as soon as Y is a proper subset of X.

If (hypothesis) Y implies the law X, it is said to explain it. Hypothesis T is of course the strongest law, for it is not only true as a hypothesis, i.e., it is a law, but it implies, hence explains, all other laws.

A law of nature is traditionally understood to be a true impossibility statement, e.g., a perpetuum mobile is impossible. It is important to note that a hypothesis X in our sense is in fact a domain-relative version

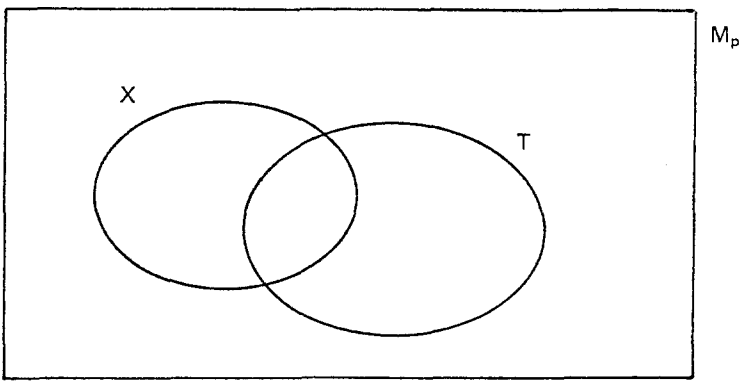


Figure 1.

of such a, potentially true, impossibility claim, viz., phenomena that would have to be represented by a member of Mp − X are claimed to be empirical impossibilities. In case hypothesis X is true, i.e., when we speak of law X, this claim is true. Note further that, when hypothesis X, whether true or false, fails to recognize an empirical impossibility x as such, i.e., when x belongs to X − T, this means that it fails to entail, and hence to explain, the law to the effect that x and similar conceptual possibilities are empirical impossibilities.

It should be stressed that the presented distinction between hypothesis and theory has nothing to do with theoretical terms. Here both a hypothesis and a theory may or may not have theoretical terms. Later on I will explicitly consider stratified hypotheses and theories, with an observational and a theoretical level. The crucial distinction between a hypothesis and a theory in this article is that the claim of hypothesis X is just one conjunct of the combined claim of the corresponding theory X.

Theory X can make two kinds of *mistakes* (see Figure 1). The members of T − X, if any, are *instantial* mistakes: empirical possibilities that are excluded by X; in other words, they are the empirically realizable counterexamples of X. Recall that X explains T when T includes X. Hence, the members of X − T, if any, may be called the *explanatory* mistakes of X (with respect to T): empirical *im*possibilities that are not excluded by X, that is, wrongly admitted models, also called mistaken models. Note that the explanatory mistakes form, by definition, a kind of counterexample that cannot be empirically realized. The set of all

mistakes of theory X is the union of these two sets T-X and X-T, which is technically called the symmetric difference between X and T, indicated by $X \Delta T$.

A theory does not only make mistakes but also makes *matches*. $T \cap X$ represents the instantial matches: empirical possibilities that are recognized as such by X or, in other words, they are the empirically realizable examples of X. Let cX indicate the complement of X with respect to Mp, i.e., $Mp - X$. X explains T, i.e., T includes X, is equivalent to cX includes cT. Hence, the members of $cT \cap cX$ are the explanatory matches: empirical *im*possibilities that are rightly excluded by X. The explanatory matches are examples that cannot be empirically realized. The union of the two sets of matches of theory X is of course equal to the complement of the total set of mistakes, viz., $c(T \Delta X)$.

Note that all mistakes and matches are ultimate, in the sense that they need not have been established. Established mistakes and matches will later come into the picture.

The following table gives a survey of the instantial and explanatory matches and mistakes of a theory.

TABLE I.

| | matches | mistakes | total (union) |
|---|---|---|---|
| instantial | $T \cap X$ | $T - X$ | T |
| explanatory | $cT \cap cX$ | $X - T$ | cT |
| total (union) | $c(T \Delta X)$ | $T \Delta X$ | Mp |

All concepts introduced thus far, and most of the ones to be introduced, can be illustrated by the following electric circuit (see Figure 2). Let $p_i$ for $1 \leqslant i \leqslant 4$ indicate that switch i is on ($\leftrightarrow$) and $-p_i$ that it is off ($\updownarrow$). Let $p_o(-p_o)$ indicate that the bulb lights (does not light).

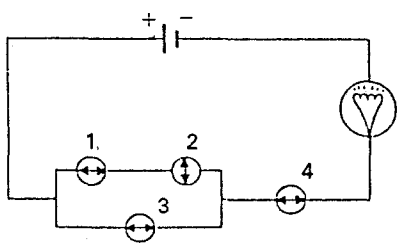


Figure 2.

It is assumed that the bulb is not defective and that there is enough voltage. An arbitrary conceptual possibility, for instance, can be represented by a set of unnegated $p_i$'s and the true theory about the circuit by the propositional formula $p_0 \leftrightarrow (((p_1 \,\&\, p_2) \vee p_3) \,\&\, p_4)$.

### 1.2. *The Naive Definition*

From now on X, Y, etc., refer to theories or just to the sets X and Y, depending on the context. When the hypotheses X and Y are intended it will be explicitly mentioned.

The naive definition of truthlikeness states that *theory Y is at least as similar (close) to the truth (T) as theory X*, indicated by NTL(X,Y,T), if the following two conditions are satisfied:

(Ni) $T - Y$ is subset of $T - X$
(Nii) $Y - T$ is subset of $X - T$

The *instantial clause* (Ni) says that the instantial mistakes of X include those of Y, and the *explanatory clause* (Nii) that explanatory mistakes of X include those of Y. Hence, it may be said that (Ni) and (Nii) require that Y instantiates and explains T at least as well as X, respectively.

Note that (Nii) implies:

(Nii*) When $X - T$ is empty, $Y - T$ is empty

that is, the claim that Y explains T as soon as X explains T. Note also that (Ni) and (Nii) together are equivalent to the claim that the mistakes of Y ($Y \,\Delta\, T$) form a subset of those of X ($X \,\Delta\, T$).

By NTL + (X, Y, T), I indicate that Y is more similar to T than X in the strict sense that the mistakes of Y form a *proper* subset of those of X. Here and later the strong verbal expressions 'closer to' or 'more similar to' will however also be used to refer to the corresponding weak notion. When the strict notion is meant it will be explicitly stated.

Some equivalent formulations of NTL are instructive. The numbers of the sets refer to Figure 3 (in which $M_p$ is not explicitly indicated). Very useful are (assuming priority of '$\cap$' and '$\cup$' over '$-$'):

(Ni)′ $X \cap T - Y = 2$ is empty
(Nii)′ $Y - X \cup T = 6$ is empty

The first tells that Y makes no extra instantial mistakes, and the second

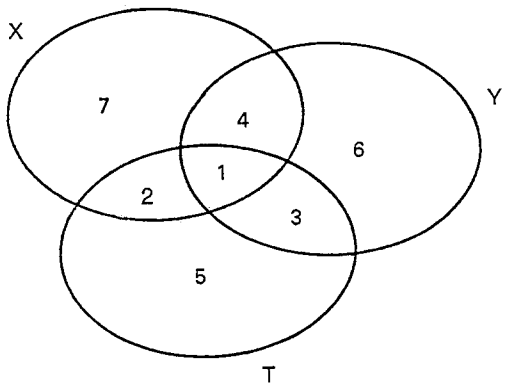


Figure 3.

that Y makes no extra explanatory mistakes. Consider also:

(Ni)'' $X \cap T$ is subset of $Y \cap T$
(Nii)'' $cX \cap cT$ is subset of $cY \cap cT$

telling that Y includes the instantial as well as the explanatory matches of X, respectively.

It is important to note that improving a theory X in the sense of finding a theory Y such that NTL + (X, Y, T) is not an easy task, due to the fact that both components are counteracting. This can be nicely illustrated by considering, e.g., just weakening of theory X: if Y is weaker than X (which was defined as: $Y \supseteq X$), then Y instantiates T at least as well as X, but X explains T at least as well as Y. Of course, strengthening a theory leads to the opposite tension.

### 1.3. *Some Formal Properties*

Any definition of the binary relation of truthlikeness between theories X and Y can be seen as a special case of a ternary relation of *theorylikeness* NTL(X, Y, Z) between theories X, Y, and Z by replacing the fixed true theory T by the variable theory Z. Then we get the general definition: $\mathrm{NTL}(X, Y, Z) =_{df} Z - Y \subseteq Z - X$ and $Y - Z \subseteq X - Z$, with plausible equivalent formulations.

NTL has several interesting properties. We list the main ones.

| | |
|---|---|
| *reflexivity*: | NTL(X, X, Y) (left) NTL(X, Y, Y) (right) |
| *antisymmetry*: | NTL(X, Y, Z) and NTL(Y, X, Z) imply X = Y (left)<br>NTL(X, Y, Z) and NTL(X, Z, Y) imply Y = Z (right) |
| *symmetry*: | NTL(X, Y, Z) implies NTL(Z, Y, X) (central) |
| *transitivity*: | (E.g.) if NTL(W, X, Z) and NTL(X, Y, Z), then NTL(W, Y, Z) (left) |

Hence, from left reflexivity, left antisymmetry, and left transitivity it follows that NTL(X, Y, Z) is for fixed Z, a partial ordering of theories. As a consequence, a sequence of theories converging to the truth is perfectly possible.

Some other interesting properties are:

| | |
|---|---|
| *centeredness*: | NTL(X, X, X) |
| *centering*: | If NTL(X, Y, X) then X = Y |
| *specularity*: | If NTL(X, Y, Z) then NTL(cX, cY, cZ) |
| *concentricity*: | If $X \subseteq Y \subseteq Z$ then NTL(X, Y, Z) and NTL(Z, Y, X) |
| *context neutrality*: | If X, Y, and Z are subsets of Mp and Mp itself is a subset of a larger set of conceptual possibilities Mp*, then NTL(X, Y, Z) implies NTL*(X, Y, Z) |

### 1.4. *Success Increase of New Theories and Its Explanation*

Up to now I have dealt with the logical problem of defining truthlikeness, assuming that T is at our disposal. In actual scientific practice we don't know T; it is the target of our theoretical and experimental efforts. Before we turn our attention to methodological rules guiding these efforts, it is fruitful to explicate the idea that one theory is more successful than another and to show that this can be explained by the hypothesis that the first theory is more similar to the truth than the second.

The success of a theory will have to be expressed in terms of the data to be accounted for. The data up to a certain moment t can be represented as follows. Let R(t) indicate the set of realized possibilities

up to t, i.e., *the accepted instances*, which have to be admitted. Note that there may be more than one realized possibility at the same time, before or at t, with plausible restrictions for overlapping domains.

Up to t there will also be some accepted hypotheses, *the (explicitly) accepted laws*, which have to be explained. On the basis of them *the strongest accepted law* to be explained is the hypothesis S(t) associated with the intersection of the sets constituting the accepted hypotheses. Of course, S(t) is, via the laws constituting it, in some way or other based on R(t); minimally we may assume that R(t) is not in conflict with S(t), that is, R(t) is a subset of S(t). In the following, however, I shall need the much stronger *correct-data-hypothesis* $R(t) \subseteq T \subseteq S(t)$, guaranteeing that R(t) does not contain explanatory mistakes, and that hypothesis S(t) does not make instantial mistakes and hence is true as a hypothesis and may hence rightly be called a *law*. Note that every L containing S(t) is a true hypothesis following from S(t), i.e., a law which is explicitly or implicitly accepted.

Now it is possible to explicate the success and problems of a theory X at time t. $R(t) - X$ indicates the set of established instantial mistakes of X, the *instantial problems of X*, whereas $X \cap R(t)$ indicates the set of established instantial matches of X, the *instantial success of X*. When X explains S(t) it is a subset of S(t) and cS(t) of cX. Hence, $X - S(t)$ represents the established explanatory mistakes, the *explanatory problems of X*, and $cS(t) \cap cX$ the set of established explanatory matches of X, the *explanatory success of X*.

For comparative judgements of the success of theories the following two clauses are obvious. Theory Y is *instantially at least as successful as* X iff Y instantiates R(t) at least as well as theory X in the sense that the instantial problems of Y form a subset of those of X, that is, Y has no *extra* instantial problems, or, equivalently, the instantial success of X is a subset of that of Y. Formally:

$$(Ni)_s \quad R(t) - Y \subseteq R(t) - X$$
$$(\equiv X \cap R(t) - Y = 2.1 = \phi \equiv X \cap R(t) \subseteq Y \cap R(t))$$

Theory Y is *explanatorily at least as successful as* X iff Y explains S(t) at least as well as theory X in the sense resulting from replacement of 'instantial' by 'explanatory' in the verbal phrases, and hence formally iff:

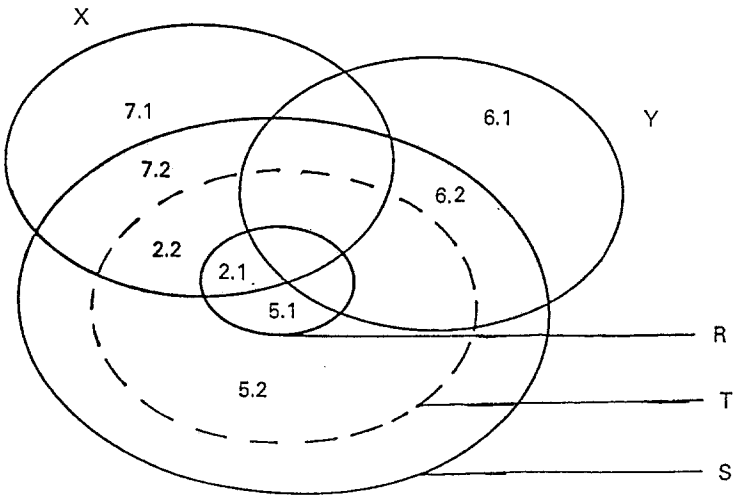


Figure 4.

$$(\text{Nii})_s \quad Y - S(t) \subseteq X - S(t)$$
$$(\equiv Y - X \cap S(t) = 6.1 = \phi \equiv cS(t) \cap cX \subseteq cS(t) \cap cY)$$

The conjunction of the instantial and explanatory clauses form the general definition of the statement that one theory is at a certain time at least as successful as another, relative to the data R(t)/S(t). This situation is depicted in Figure 4 (in which T is indicated by an interrupted circle to stress that it is unknown): the sets 2.1 and 6.1 are empty.

We arrive at a crucial argument: when one theory is at a certain moment at least as successful as another, simply called *success-dominance*, this fact can be derived from, and hence explained by, the following two hypotheses: the *TA*, (*truth-approximation*), *hypothesis*, that the first is at least as similar to the truth as the second, and the already introduced *correct-data-hypothesis*. All notions in this argument have been explicated, and the proof of its validity is only a matter of elementary set-theoretical manipulation, as will be clear from the following survey of the argument:

| | | | | |
|---|---|---|---|---|
| (Ni) | $T - Y \subseteq T - X$<br>$R(t) \subseteq T$ | (Nii) | $Y - T \subseteq X - T$<br>$T \subseteq S(t)$ | (truth-approximation)<br>(correct-data) |
| $(Ni)_s$ | $R(t) - Y \subseteq R(t) - X$ | $(Nii)_s$ | $Y - S(t) \subseteq X - S(t)$ | (success-dominance) |

It is also easy to prove in addition that, if both hypotheses are true, the first theory will always remain at least as successful as the second, i.e., in the face of new supplementary data, resulting in $R(t) \subseteq R(t') \subseteq T$ and $T \subseteq S(t') \subseteq S(t)$, for $t'$ later than $t$. From this it immediately follows that the TA-hypothesis is a decent, comparative, empirical hypothesis which can be further tested and falsified or confirmed.

As a rule a new theory will not include all instantial success of the former and/or not all explanatory success, let alone both forms of success. The idea is that the relative merits can now be explained on the basis of a detailed analysis of the relative 'position' to the truth, but for these cases a general argument is obviously not possible.

The basic argument makes clear that and how empirical progress is possible within a conceptual frame Mp for a domain D. It is important to note that the specific TA-hypothesis presupposes the *frame-hypothesis* of the research program that $\langle D, Mp \rangle$ indeed generates the unique, time-independent set T of empirical possibilities. The frame-hypothesis creates as it were the possibility that there may occur theories closer to the truth than others, and that if theories are more successful than others it may be (but need not be) for that reason. In other words, although each specific example of empirical progress is explained on the basis of the corresponding specific TA-hypothesis, the possibility of the generic phenomenon of empirical progress is explained on the basis of the frame-hypothesis associated with $\langle D, Mp \rangle$.

Two successive generalizations bring us to the explanation of the success of the natural sciences in general. First, the frame-hypothesis is true for all possible conceptual frames Mp with respect to the natural domain D. Second, the frame-hypothesis is true for all frames for all natural domains. I do not claim that these generalizations don't have exceptions. If they are true in the majority of cases, they serve their purpose.

### 1.5. *Methodological Rules*

Let us return to one particular $\langle D, Mp \rangle$ and the corresponding frame-hypothesis. From the foregoing it immediately follows that the *rule of*

*success* (RS) "If theory Y is more successful than theory X, then choose theory Y (for the time being)" is functional for approaching the truth in the sense that Y may still be closer to the truth than X, which would explain that Y is at least as successful, and that X cannot be closer to the truth than Y, for otherwise X could not be less successful. Hence, RS can be justified on the basis of the frame-hypothesis as a prescriptive rule. In my opinion it can be conceived as the (naive) explication of the hallmark of scientific rationality. Of course, as long as one does not dispose of explicit knowledge of T it is impossible to have a rule of success that can guarantee that the more successful theory is closer to the truth.

The following rules are heuristic rules, of which it is easy to see that they stimulate new applications of RS. To begin with, the *rule of content* (RC) "Aim at success-preserving strengthening *or*, *pace* Popper, weakening of your theory", where strengthening a theory amounts to considering a subset and weakening to the opposite. Further, the *rule of testing* (RT) "Aim at establishing new counterexamples (instantial mistakes) of your theory, and new laws which cannot be explained by your theory". Finally, I would like to mention the *rule of dialectics* (RD) for two theories that escape RS because of divided success, "Aim at a success-preserving synthesis of two RS-escaping theories".

It is important to note that RS is not a rule of inference in the sense that it does not conclude that the more successful theory is true (as a hypothesis, let alone as a theory). It suggests at most the provisional conclusion that the more successful theory is closer to the truth than the other. More generally, putting RS in its generalized form "Choose the most successful theory among the available theories", it suggests at most the provisional conclusion that the most successful theory is the closest to the truth, which might be called the rule of TA-inference (as opposed to truth-inference).

It is interesting to compare the rule of TA-inference to the so-called 'inference to the best explanation' (BE-inference), which prescribes, according to one plausible reading, to conclude, provisionally, that the most successful theory is true (as a hypothesis), provided it has not yet been falsified. Several conceivable objections to and limitations of this rule of truth-inference do not apply to the suggested rule of TA-inference. The two main ones are the following. First, in contrast to BE-inference, TA-inference is not restricted to the case that the most successful theory has not been falsified. Second, TA-inference suggests

that being-the-closest-to-the-truth is something relative to other (i.e., the available) theories, which is unproblematic. However, BE-inference suggests that being-true(-as-hypothesis) is something relative to other theories, and that is close to a contradiction.

### 1.6. *Laws and the Explanatory Clauses*

The explanatory truthlikeness and success clauses were mainly motivated in terms of (established) explanatory mistakes. In this subsection we shall see that both clauses are also in perfect agreement with intuitions about the explanation of (accepted) laws.

I start by analyzing the explanatory truthlikeness clause

(Nii) $Y - T$ is subset of $X - T$

which coud be interpreted as claiming that *Y explains T at least as well as X* in the sense that it makes no extra explanatory mistakes with respect to T. From (Nii) it follows that Y makes for no law extra explanatory mistakes, formally:

(NLii) For all laws L, $Y - L$ is subset of $X - L$

However, (NLii) and (Nii) are even equivalent, for (NLii) reduces to (Nii) for the special case $L = T$. Hence, (Nii) may, because of this equivalence, be paraphrased by the intuitively appealing claim that *Y explains all laws at least as well as X*.

From (NLii) immediately follows:

(NLii*) For all laws L when $X - L$ is empty $Y - L$ is empty

When $X - L$ is empty this means that X explains L. Hence, (NLii*) says that *all laws explained by X are explained by Y*. Note that (NLii*) includes as a special case

(Nii*) When $X - T$ is empty $Y - T$ is empty

of which I already noted that it can be read as: Y explains T when X does.

It is not difficult to prove that (NLii*) implies (Nii), and hence that it is equivalent to (Nii) and (NLii). (NLii*) in fact claims generally that all Z that include $T \cup X$ (then it is a law explained by X) include $T \cup Y$. But this is only possible when $T \cup Y$ is a subset of $T \cup X$, and this is equivalent to (Nii).

In the form (NLii*), (Nii) can be seen as an explication of the idea that Y explains T at least as well as X in terms of laws. Calling all laws explained by a theory its ultimate law-explanatory success and all laws not explained by it its ultimate law-explanatory problems, (NLii*) amounts to the claim that Y's ultimate law-explanatory success includes that of X, or, equivalently, that *X's ultimate law-explanatory problems include those of Y*, that is, Y has no extra ultimate law-explanatory problems. Note the close analogy with the different interpretations of (Nii) in terms of the explanatory matches and (extra) mistakes presented at the beginning.

So much for the explanatory clause of truthlikeness. Now I turn to that of success, i.e., $(\text{Nii})_s$, which explicated that Y is explanatorily at least as successful as X by claiming that *Y makes no extra explanatory mistakes with respect to the strongest accepted law S(t)*. I call law L an *accepted* law when it is implied by S(t). It is easy to check that if we substitute in (NLii) 'accepted laws' for 'laws' we get an equivalent version $(\text{NLii})_s$ of $(\text{Nii})_s$ which can be paraphrased by the appealing claim that *Y explains all accepted laws at least as well as X*. By the same substitution in (NLii*) we get a claim $(\text{NLii*})_s$, equivalent to $(\text{NLii})_s$ and $(\text{Nii})_s$, telling that *Y explains all accepted laws explained by X*. Also, with reformulations of $(\text{NLii*})_s$ in terms of 'accepted', instead of 'ultimate', law-explanatory success, problems and extra problems follow immediately.

In sum, the explanatory clauses are in perfect agreement with intuitions about the explanation of laws by a theory that is closer to the truth or more successful than another.

### 1.7. *Truthlikeness and Success of Stratified Theories*

Thus far it might seem that my conceptually relative point of departure leads to an extreme form of relativistic (theoretical) realism. However, this would only be the case if I exclude constraints between different conceptual frames for the same domain. In this subsection I will deal with the relation between an observational and a theoretical (cum observational) level, that is, an observational and a theoretical conceptual frame for our domain, where the distinction between observational and theoretical components is of course assumed to be not of the classical, absolute form but of a sophisticated, theory-relative kind.

Let Mp indicate the set of conceptual possibilities with theoretical

and observational components (potential models) and Mpp the corresponding set of conceptual possibilities without theoretical components (potential partial models). Let $\pi$ be the projection function from Mp onto Mpp such that for x in Mp, $\pi(x)$ is the result of dropping the theoretical components in x (technically speaking, leading to a substructure of x). For a subset X of Mp, $\pi X$ indicates the set of all projections of all members of X.

Application of the frame-hypothesis to $\langle D, Mp\rangle$ and $\langle D, Mpp\rangle$ leads to the existence of unique, time-independent subsets $T = Mp(D)$ of Mp and $T_0 = Mpp(D)$ of Mpp, representing 'the theoretical (cum observational) truth' and 'the observational (or partial) truth', respectively. It is not guaranteed that $\pi T = T_0$. That $\pi T$ is a subset of $T_0$ is a semantic fact, to be called *T-projection*: if something is empirically possible, its observational part will also be empirically possible. Mp is said to be *complete* with respect to $\langle D, Mpp\rangle$ if $T_0$ is also a subset of $\pi T$, and hence $\pi T = T_0$.

Here (and in the corresponding subsection of Section 3) I will restrict the attention to $\pi T$. I will assume that the data are formulated as subsets of Mpp, but I will assume in addition that they are correct with respect to $\pi T$, i.e., R(t) is a subset of $\pi T$ and $\pi T$ of S(t).

If Mp is complete with respect to $\langle D, Mpp\rangle$, the results also follow trivially for $T_0$. If Mp is incomplete, the question is whether $\pi T$ and R(t), being subsets of $T_0$, and $T_0$ of S(t), guarantee the replacement. Note, as I will assume, that it remains plausible to assume that R(t) is a subset of $T_0$, but it need not be a subset of $\pi T$. From these mutual relations it follows, generally speaking, that the results along the explanatory line can be extrapolated unproblematically, but not along the instantial line without qualifications.

The important question is whether truthlikeness on the theoretical level is projected on the observational level: Does NTL(X, Y, T) imply NTL($\pi X$, $\pi Y$, $\pi T$)? Let us first consider the explanatory clause and assume that Y explains T at least as well as X. It is easy to check, using the general fact that the emptiness of a set of the form $X - Y \cup Z$ guarantees the emptiness of $\pi X - (\pi Y \cup \pi Z)$, that the explanatory clause on the observational level follows indeed: $\pi Y$ explains $\pi T$ at least as well as $\pi X$. And, hence, using the previous results on success explanation, the explanatory success clause also follows: $\pi Y$ explains S(t) at least as well as $\pi X$.

Due to the many-one character of projection, the instantial side is

not so easy. Under the naive condition, the 'projection-step' is invalid, as is easy to check: there may be an extra explanatory mistake of X and a common instantial mistake that are projected on an element of $(\pi X \cap \pi T) - \pi Y$. To exclude this just amounts to filling the gap between the instantial clauses on the theoretical and the observational levels. Hence, there seems no interesting condition that guarantees the projection. In Section 3 we shall see that the situation changes considerably when we include similarity-considerations between structures.

### 1.8. *Quantitative Truthlikeness*

Let us return to unstratified theories. The naive notion of truthlikeness that has been defined is a comparative or qualitative notion in the sense that it is purely based on *sets* of mistakes. It is plausible to define naive quantitative truthlikeness in terms of the *numbers* of mistakes. That is, I define 'the (naive) dissimilarity or *distance* of X from T', indicated by $\mathrm{NTD}(X, T)$, $|T - X| + |X - T| = |T \Delta X|$.

$\mathrm{NTD}(X, T)$ is not only a quasi-distance function in the sense that $\mathrm{NTD}(X, T) \geq 0$, but it is even a proper distance function for it satisfies in addition: $\mathrm{NTD}(X, T) = 0$ iff $X = T$, $\mathrm{NTD}(X, T) = \mathrm{NTD}(T, X)$and $\mathrm{NTD}(X, Y) + \mathrm{NTD}(Y, T) \geq \mathrm{NTD}(X, T)$. Moreover, NTL and NTD are compatible in the sense that $\mathrm{NTL}(X, Y, T)$ guarantees $\mathrm{NTD}(Y, T) \leq \mathrm{NTD}(X, T)$.

## 2. STRUCTURELIKENESS AND TRUTHLIKENESS OF STRUCTURES

### 2.1 *Basic Assumptions*

Up to now I have been dealing with the problem of truthlikeness of theories and more generally theorylikeness. But there is also a problem of truthlikeness of structures and more generally structurelikeness. In the circuit example (Figure 2), for instance, it is clear that there is, given the conceptual frame, not only just one true theory characterizing the set of empirically possible states of that particular circuit. There is also just one true description of the actual state of the circuit as it is depicted, $p_0 \,\&\, p_1 \,\&\, - p_2 \,\&\, p_3 \,\&\, p_4$, according to the standard propositional representation. In general, in addition to the frame-hypothesis leading to the assumption that there is just *one true theory*, I will assume

that, given a conceptual frame, every particular situation or state of affairs (of a system) in the domain, every actual world so to speak, has just one correct representation or *one true description*. By consequence, with each experiment, i.e., with each realization of an empirical possibility, there is associated a unique true description within the conceptual frame.

Hence, the traditional problem of explicating the idea of truthlikeness concerns on closer inspection two intuitive phrases, viz., "one description is more similar to the true description than another" and "one theory is more similar to the true theory than another". In the previous section I dealt with the second problem, neglecting the fact that it is plausible to take into account a possible underlying notion of likeness of descriptions. This will be done in the next section. In the present section I will just deal with the idea of truthlikeness of descriptions or, equivalently from the structuralist point of view, truthlikeness of structures and more generally structurelikeness.

Let x, y, z indicate structures in Mp, and s(x, y, z) indicate that y is at least as similar (close) to z as x. The true structure of the context will be indicated by t. I will not aim at a general definition of structurelikeness, for a precise definition will have to depend on the specific nature of the conceptual possibilities.

When s(x, y, z), y is said to be *between*, or an *intermediate* of, x and z. Structurelikeness is not generally assumed to be symmetric: s(x, y, z) does not generally imply s(z, y, x). As a consequence, being in between or an intermediate may be a directed notion: if y is between x and z in the sense of s(x, y, z), this does not yet imply that y is also between z and x in that sense of structurelikeness.

Structures x and z are said to be connected or *related*, r(x, z), iff there is y such that s(x, y, z). It follows also that r is not by definition symmetric, i.e., r(x, z) does not automatically imply r(z, x). But r(x, z) is already guaranteed by s(x, x, z) or s(x, z, z). Hence, the basic idea behind r(x, z) is not the existence of a *proper* intermediate, but only that x and z have at least so much in common that it makes sense to talk about (proper and improper) intermediates: in other words, they may also be said to be *comparable*. For instance, all pairs of propositional structures (see below) will be comparable if Mp contains only structures constituted by one set of elementary propositions, as in the case of the circuit example. But as soon as structures based on subsets of this set are also taken into consideration, not all pairs are comparable

anymore. p & q and − p & − q have an intermediate, e.g., p & − q, but p & q and p don't. The case of concretization below (Section 4) provides another example of this situation, e.g., not every Van der Waals gas model is a concretization of every ideal gas model (they may deal with different sets of states and/or different numbers of moles). Hence, they don't need an intermediate. In general, a minimal condition for comparability seems to be that the two structures have the same base- or domain-sets.

It would have been possible to introduce r(x, z) as a primitive term. But I have not done so because it always seems possible to read r(x, z) directly from the relevant specific definition of s(x, y, z). In the next section we shall see that naive truthlikeness as defined in the previous section is in fact based on *trivial* structurelikeness, indicated by t(x, y, z), defined by x = y = z. (The symbol 't' is used in this text as a time variable, and to indicate trivial structurelikeness and 'the descriptive truth', but confusion need not arise). Note that in the case of trivial structurelikeness two structures are only related when they are equal. Hence, naive truthlikeness is based on the idea that two different structures are never comparable. This point was already essentially noted in Kuipers (1987a) and also in Oddie (1986, Ch. 3).

### 2.2. *Properties*

Let us assume some plausible properties of the notion of structurelikeness. s is *centered* iff s(x, x, x), and *centering* iff s(x, y, x) implies x = y. s is said to be *conditionally left/right reflexive* iff s(x, y, z) implies all kinds of left and right reflexivity, i.e., s(x, x, y), s(x, x, z), s(y, y, z), and s(x, y, y), s(x, z, z), s(y, z, z), respectively. Note that r(x, z) now implies s(x, x, z) and s(x, z, z). Together these properties are called *the minimal s-conditions*. Note that being centered implies that r is reflexive, but the converse does not hold.

s is called *symmetric* when s(x, y, z) implies s(z, y, x) and *antisymmetric* when s(x, y, z) and s(z, y, x) imply x = y = z, in which case centering trivially follows. If s is symmetric, then r is symmetric as well; and if s is antisymmetric, then r is antisymmetric. Note that the converse implications do not hold.

There are many ways in which s can be *transitive*, e.g., left transitivity: s(w, x, z) and s(x, y, z) imply s(w, y, z). However, none of these ways implies that r is transitive, as a laborious survey makes clear, nor does

the transitivity of r imply any of these ways. Moreover, and this is even more important to note, r can be transitive without assuming that the middle term is the intermediate: if r(x, y) and r(y, z), then r(x, z) may generally be the case, without implying that r(x, z) is due to s(x, y, z).

As far as r is concerned, it is useful to state in sum that r may well be an equivalence relation or a partial ordering, without strong implications for s. In the case of an equivalence relation, comparability generates equivalence classes of comparable structures. In the case of a partial ordering, directed sequences of comparable structures arise.

In Section 4 we shall see that concretization provides a good example of antisymmetric structurelikeness, generating a partial ordering of comparable structures. In the present section I will restrict the attention to some symmetric examples of structurelikeness, generating in all cases comparability as an equivalence relation, sometimes in the trivial sense that all structures are comparable.

### 2.3. *Examples of Symmetric Structurelikeness*

A typical symmetric example is the case of *propositional structures*. Given a fixed set of elementary propositions a structure is identified with a propositional constituent, i.e., an arbitrary conjunction of all elementary propositions, each of them negated or unnegated. It is easy to see that such a constituent can be represented, for example, by the set of its unnegated elementary propositions. The plausible specification of structurelikeness is then as follows: s(x, y, z) iff the symmetric difference between y and z is a subset of that between x and z. Note that this corresponds formally to the naive definition of theorylikeness on the level of structures. Moreover it is easy to check that all propositional structures (generated by a fixed number of elementary propositions) are comparable and that s satisfies not only the minimal s-conditions of being centered, centering, and left/right reflexivity, but also symmetry: if s(x, y, z) then s(z, y, x).

As in the case of theorylikeness, there is also a plausible quantitative variant, which was already proposed in Tichý (1974): the distance between two propositional constituents may be defined as the size of the indicated symmetric difference set.

It is just a technical exercise to generalize the qualitative definition to propositional constituents based on different sets of elementary prop-

ositions. Comparability of structures then coincides with being based on the same set of elementary propositions.

It is now also plausible to formulate symmetric structurelikeness for *first-order structures*. Let Mp consist of structures $\langle \ldots D_i \ldots; \ldots R_j \ldots \rangle$ of a fixed similarity type, with $D_i$'s as domain-sets and the $R_j$'s as relations defined on one or more of them. Structurelikeness s(x, y, z) is now defined by the requirement that the corresponding domains are the same ($D_i(x) = D_i(y) = D_i(z)$ for all i) and that the corresponding relations are as follows: the symmetric difference between $R_j(y)$ and $R_j(z)$ is a subset of that between $R_j(x)$ and $R_j(z)$ for all j. It is easy to check that s satisfies the minimal s-conditions, that it is symmetric, and that the corresponding notion of comparability between two structures amounts to having the same domain-sets, i.e., the first requirement.

Note that three different conceptual structures x, y, and z, related by s(x, y, z), can only be realized at different moments, due to the fact that s(x, y, z) requires that they have the same domain.

For a definition of quantitative structurelikeness between first-order structures, see Niiniluoto (1987, Ch. 10.3). Also see Oddie (1986, Ch. 3), who examines examples of symmetric structurelikeness between propositional and first-order structures.

I conclude this section with some symmetric examples of elementary *real number structures* specifying one or more ordered real numbers. Real numbers of 'the same dimension' will be indicated by numbered x's, etc.

For one dimension it is plausible to define structurelikeness by 'x1 ≤ x2 ≤ x3 or x3 ≤ x2 ≤ x1', indicated by $s^1$(x1, x2, x3). For two dimensions one possible definition reads '$s^1$(x1, x2, x3) and $s^1$(y1, y2, y3)', indicated by $s^2(\langle x1, y1\rangle, \langle x2, y2\rangle, \langle x3, y3\rangle)$. In both cases the minimal s-conditions and symmetry are satisfied, and all structures are comparable.

The latter property, and only that one, is not shared by another possible definition of structurelikeness for two dimensions, viz., '$s^1$(x1, x2, x3) and y1 = y2 = y3'. In that case, only 'horizontal pairs' are comparable. Restricted comparability of real number structures may or may not make sense, depending on the nature of the properties the real number are supposed to represent.

Note that I would have obtained antisymmetric, directed variants of

real number structurelikeness if I had started for one dimension with 'x1 ≤ x2 ≤ x3' as opposed to 'x1 ≤ x2 ≤ x3 or x3 ≤ x2 ≤ x1'.

## 3. REFINED TRUTHLIKENESS OF THEORIES

### 3.1. *Basic Definition*

It is clear that the naive definition of truthlikeness of theories does not exploit the idea that one structure may be more similar to a second than a third, i.e., the idea of an underlying notion of structurelikeness. Let us assume that there is such an underlying ternary relation of structurelikeness s and that it satisfies the minimal s-conditions introduced in Section 2: being centered, centering (together: s(x, y, x) iff x = y), and conditional left and right reflexivity (s(x, y, z) implies, e.g., s(x, x, y) and s(y, z, z)).

I will present a refined definition of truthlikeness of theories that turns out to have many plausible and desirable properties if we neglect queer theories and restrict our attention to so-called convex theories. A set X is called *convex* (with respect to s) if it is closed for intermediates, i.e., if for all x and z in X and all y if s(x, y, z), then y is in X.

Note that there are already elementary examples of non-convex theories. Assuming propositional structurelikeness as defined in the previous section, it is, for instance, easy to see that the propositional theories indicated by 'p → q' and 'p ↔ q' are non-convex, for the non-model 'p & − q' is between the models 'p & q' and '− p & − q'. However, it is doubtful whether theories that are non-convex, with respect to the relevant underlying notion of structurelikeness, play an important role in science proper. If, for instance, T is not convex, this has the consequence that there is an empirical *im*possibility (in the relevant sense) between two empirical possibilities, and this is unlikely as far as nature is continuous.

But the restriction to convex theories is certainly not unproblematic. One of the referees is probably right by suggesting that the true theory about the atomic weights of the isotopes of uranium, if reconstructed along the adopted structuralist lines, is not convex. The other referee notes that the property of convexity has already received considerable attention in the truthlikeness debate (see in particular Oddie (1987) for this and further references).

However this may be, many of the results to be presented are simply invalid or need technical qualifications if the relevant local convexity assumption is not satisfied; but I will not specify such qualifications. Instead of just dealing throughout with convex theories or specifying the situation for non-convex cases, I have chosen the middle course of locally indicating when convexity is a necessary condition for the reported result. The first option can be obtained from the present text by just skipping all local convexity assumptions and assuming convex theories throughout.

I will introduce the refined definition of "Y is at least as similar to T as X" again by an instantial and an explanatory clause.

The *refined instantial clause* again expresses the intuitive idea that Y instantiates T at least as well as X, but now not just in the sense of full instantial matches, but also in the sense of approximate matches:

(Ri) For all x in X and z in T if r(x, z), then there is y in Y such that s(x, y, z)

It is easy to check that (Ri) implies the corresponding naive instantial clause (Ni) because s satisfies centering. Hence, it is a strengthening of the naive clause. A reformulation of (Ri) is instructive. Due to conditional reflexivity, (Ri) is equivalent to:

For all x in $X - Y$ and z in $T - Y$ if r(x, z) then there is y in Y such that s(x, y, z)

that is, for every extra model of X comparable to an instantial mistake of Y, Y has a model which is at least as similar to that mistake.

The *refined explanatory clause* will not be a strengthening but a weakening of the corresponding naive one, which required that $Y - X \cup T$ was empty. In the context of structurelikeness it is plausible to leave room for members of $Y - X \cup T$, i.e., extra explanatory mistakes of Y, provided they are between $X - T$ and T, that is, all y in $Y - X \cup T$ have to be between a member of $X - T$ and one of T. This clause guarantees as it were that Y is moving up from X to T, without detour. This results in the following clause:

(Rii) For all y in $Y - X \cup T$ there are x in $X - T$ and z in T such that s(x, y, z)

It is evident that the naive explanatory clause implies the refined one. (Rii) expresses the idea that every extra explanatory mistake of Y is at

least as similar to some empirical possibility as some explanatory mistake of X. This may also be paraphrased as the claim that every extra explanatory mistake of Y guarantees the existence of some explanatory mistake of X that is at least as serious in the weak sense that the former is at least as similar to some empirical mistake as the latter. Hence, (Rii) expresses in this weak refined sense that Y explains T at least as well as X.

It would be possible to strengthen (Rii) by adding "and there is no $z'$ in T such that $s(y, x, z')$" which leads to a strong refined sense of the idea that Y explains T at least as well as X. However, this did not lead to further conceptual elucidation, nor to elegant results. Given the fact that I also like to keep the explication as weak as possible, the present version of the refined explanatory clause is the most attractive one.

(Rii) is the general formulation, also appropriate for non-convex theories, in the sense of also leading to centering for such theories (see below). When T is convex, (Rii) can be simplified to:

> For all y in $Y - X \cup T$ there are x in X and z in T such that $s(x, y, z)$

for the convexity of T assures that the guaranteed x in X is in $X - T$. Hence, in this case, (Rii) amounts to the claim that (every member of) $Y - X \cup T$ is between (a member of) X and (a member of) T. The inclusion of non-convex cases precludes this otherwise highly plausible conceptual justification of (Rii).

The resulting definition of refined truthlikeness, i.e., *Y is at least as similar (close) to T as X*, indicated by RTL(X, Y, T), imposes both the clauses (Ri) and (Rii), and may be paraphrased as: Y instantiates and explains T at least as well as X.

It is easy to state and prove the desirable reducibility of the refined to the naive definition. If s is just trivial structurelikeness t – which was defined by $t(x, y, z)$ iff $x = y = z$ – refined truthlikeness reduces to naive truthlikeness. That is, indicating RTL(X, Y, T) based on $s = t$ by $RTL_t(X, Y, T)$, it is easy to prove the following *reduction theorem*: NTL(X, Y, T) iff $RTL_t(X, Y, T)$. That (Ri) reduces to (Ni) for $s = t$ follows immediately from the fact that it implies this, as we have seen already, and that the condition is vacuous for different members of X and T, for $r_t(x, z)$ implies $x = z$, i.e., different structures are never comparable on the basis of trivial structurelikeness. On the other hand,

(Rii) reduces trivially to (Nii) for $s = t$, for in that case there cannot be a member of Y outside $X \cup T$ between two different members of $X \cup T$. Note in passing that all sets are trivially convex with respect to t.

As already hinted upon for the explanatory clause, the definition of refined truthlikeness might be sharpened by strengthening one or both clauses, but I do not see convincing reasons to do so. Moreover, it is important to note that RTL(X, Y, T) is a general definition in the sense that it is not based on a particular specification of structurelikeness. Any specification that is appropriate for the particular type of structures of the context leads to the relevant specific form of RTL for that context. It might be such that such a specification also makes some sharpening of the clauses plausible.

### 3.2. *Formal Properties*

For the formal properties to be considered I jump to the general definition of refined *theory*likeness that is obtained from the formal version of the refined definition of *truth*likeness by replacing the true theory T by the arbitrary theory Z. In Figure 3, T is also supposed to be replaced by Z. It will turn out that RTL satisfies almost all properties that have been listed as properties of NTL, with some qualifications, in particular for symmetry.

Illuminating is the *sufficient condition property* (SC-property): if sets $X \cap Z - Y = 2$ and $Y - X \cup Z = 6$ and $Z - X \cup Y = 5$ and/or $X - Y \cup Z = 7$ are empty, then RTL(X, Y, Z). From this property immediately follow the following properties. *Concentricity*: if X is a subset of Y and Y of Z or if Z is a subset of Y and Y of X, then RTL(X, Y, Z), with the immediate consequence that RTL is *centered*, i.e., RTL(X, X, X). Moreover, concentricity implies (unconditional) *left and right reflexivity*: RTL(X, X, Y) and RTL(X, Y, Y), respectively. Hence, all theories are comparable, in the sense that for all X and Z there is Y such that RTL(X, Y, Z). It is also easy to prove that RTL satisfies *centering*, i.e., if RTL(X, Y, X) then $X = Y$. As a consequence, RTL satisfies, like NTL, the three properties that were called the minimal s-conditions for the underlying notion of structurelikeness. That these likeness notions share these minimal formal properties is plausible and desirable: theorylikeness may well function as structurelikeness for likeness of higher-order theories: sets of theories of the

present exposition. Note in passing that centering would only follow for convex X, if I had not required in (Rii) that x is an explanatory mistake of X, but simply that it is a model of X.

Another interesting property RTL shares with NTL is *context neutrality*. Let X, Y, and Z be subsets of Mp. If Mp itself is a subset of an extended set of conceptual possibilities Mp* and if s* is an extension of s (with r* and RTL* based on s*), then RTL(X, Y, Z) iff RTL*(X, Y, Z). The proof uses the fact that conditional reflexivity of s already guarantees for all x and z in Mp that r(x, z) iff r*(x, z). Hence, theorylikeness is not disturbed by enlargement or diminution of the set of conceptual possibilities, as long as the theories themselves are not changed.

Let us turn to *(anti-)symmetry*: first the central versions. Whereas naive theorylikeness was trivially symmetric, it now depends on the specific nature of the underlying notion of structurelikeness whether the ternary relation of theorylikeness is symmetric in the sense that RTL(X, Y, Z) implies RTL(Z, Y, X), or not. If it is not symmetric it may be antisymmetric: if RTL(X, Y, Z) and RTL(Z, Y, X) then X = Y = Z, in which case centering (RTL(X, Y, X) implies X = Y) immediately follows.

It is easy to check that RTL(X, Y, Z) is symmetric when it is based on symmetric structurelikeness, provided both X and Z are convex. In other words, the refined definition guarantees symmetry transport from the level of structures to the level of convex theories. However, antisymmetry transport from the level of structures to that of theories is not guaranteed by the refined definition. In Section 4 we shall see that theorylikeness based on (antisymmetric) concretization of structures provides an antisymmetric example.

Turning to non-central symmetry notions, left antisymmetry is, in view of the possibility of sequences of theories straightforwardly converging to the truth, the most interesting notion. Under certain conditions it is not difficult to prove that left antisymmetry is transported from structurelikeness to theorylikeness: to be precise, if structurelikeness is left antisymmetric and if it is *(de)composable*, defined by s(x, y, z) iff r(x, y) and r(y, z), and if X and Y are convex, then RTL(X, Y, Z) and RTL(Y, X, Z) imply X = Y. That s is *de*composable in the sense that s(x, y, z) implies r(x, y) and r(y, z) follows immediately from conditional reflexivity and the definition of r. That s is *composable* in the sense that r(x, y) and r(y, z) together imply s(x, y, z) is a substan-

tial condition. But, as we shall see in Section 4, it is trivially satisfied by the ternary relation of concretization.

NTL satisfied all kinds of *transitivity*, of which, again in view of truth convergent sequences, left transitivity is the most important one. It is not difficult to prove that left transitivity of structurelikeness is transported to refined theorylikeness, that is, it guarantees: RTL(W, X, Z) and RTL(X, Y, Z) imply RTL(W, Y, Z). Other, but similar, transitivity results follow easily. Combining the results about left reflexivity, left antisymmetry, and left transitivity we get that RTL(X, Y, Z) is a partial ordering of convex theories for fixed Z if s(x, y, z) is (de)composable and a partial ordering for fixed z. Hence, under these conditions a sequence of convex theories converging to the truth is perfectly possible.

There is one property of NTL that is not at all shared by RTL, viz., *specularity*: RTL(X, Y, Z) does not generally imply RTL(cX, cY, cZ). This is directly related to the fact that NTL deals with instantial and explanatory mistakes in essentially the same way, whereas RTL introduces a basic asymmetry between these kinds of mistakes. The first proposal for a definition of refined truthlikeness (Kuipers, 1987a) stuck to the symmetric treatment of the two kinds of mistakes. It amounted to (Ri), and as second clause (Ri) applied to the complements of X, Y, and T/Z. A number of objections raised by van Benthem (1987) to NTL and that refined proposal were essentially due to the symmetric treatment of instantial and explanatory mistakes, which prevented, for instance, the allowance of extra explanatory mistakes for the better theory.

### 3.3. *The Child's Play Objection*

Let us consider a famous objection to the naive definition due to Graham Oddie (1981). He formulated this objection in a discussion about David Miller's original version of the naive definition in which the true theory is identified with the one-element set {t} containing the unique structure t representing the actual world. Oddie noted that it would be child's play to replace a theory which is, in my terminology, false as a hypothesis, in the sense of a theory not containing t, by one that is closer to the truth, viz., by just strengthening the theory. In my 'empirical possibility version' of the truth and the naive definition, this objection can still be made when a theory does not contain any empirical

possibility: if $X \cap T = \phi$ and Y is a strengthening, i.e., a subset, of X, then NTL(X, Y, T). However, as is easy to see, according to RTL this child's play is excluded, in particular by (Ri). By strengthening a theory, mistaken models may well be dropped that are necessary for an intermediate between a mistaken model and an instantial mistake of the original theory. This would not be a problem if such, in this respect problematic, strengthenings could be distinguished from non-problematic strengthenings; but this is of course impossible without knowing T.

### 3.4. *Being More Successful, Its Explanation, and the Rule of Success*

An important question is whether there can also be given a plausible refined definition of 'more successful' such that the corresponding rule of success is functional for approaching the truth in the refined sense. The answer to this question is positive; to show this it is crucial to point out that the adapted TA(Truth-Approximation)-hypothesis and the correct-data-hypothesis can explain that one theory is more successful than another.

Recall that at time t, R(t) indicates the instances to be admitted and S(t) the strongest law to be explained, and the correct-data-hypothesis guarantees that R(t) is a subset of T and T of S(t).

I will first give the refined definition of "*theory Y is, relative to R(t)/S(t), at least as successful as theory X*" and then paraphrase the clauses (for the numbering of sets, see Figure 4):

- $(Ri)_s$ For all x in X and z in R(t) if r(x, z), then there is y in Y such that s(x, y, z)
- $(Rii)_s$ For all y in $Y - X \cup S(t) = 6.1$ there are x in $X - S(t)$ and z n S(t) such that s(x, y, z)

Note first the strong analogy between this definition and the refined definition of truthlikeness. T has been replaced by R(t) or S(t) in a systematic way. The first clause says that Y represents R(t) at least as well as X. Hence, Y may be said to be instantially at least as successful as X. The second clause states that for every established extra explanatory mistake of Y there is an established explanatory mistake of X that is at least as serious with respect to some model of S(t).

It is easy to check that the refined instantial success clause is stronger than the naive one $(Ni)_s$, hence established extra instantial mistakes remain forbidden for Y, formally: $X \cap R(t) - Y = 2.1 = \phi$. What $(Ri)_s$

substantially adds to $(Ni)_s$ will be specified in the next subsection. On the other hand, $(Rii)_s$ is substantially weaker than the corresponding naive clause $(Nii)_s$. Established explanatory mistakes of Y are now not forbidden, but they have to lie between an explanatory mistake of X with respect to S(t) and a model of S(t). Finally, it is easy to check that the two refined success clauses reduce to the corresponding naive ones when s is assumed to be trivial.

Assuming correct data, i.e., R(t) is a subset of T and T a subset of S(t), and assuming that S(t) is convex, the two refined success clauses follow trivially from the corresponding RTL-clauses. Accordingly, analogous to the naive case, being more successful in the refined sense can be explained on the basis of the refined truth approximation hypothesis, assuming correct data and convex S(t). Note that R(t), fortunately, does not need to be convex for this result.

Again it is important to realize that the TA-hypothesis can be further tested by realizing new empirical possibilities and/or establishing new laws. That is, it is a decent, empirical hypothesis.

In the same way as in the naive case, the explanation of the possibility of specific progress presupposes the frame-hypothesis that $\langle D, Mp \rangle$ indeed generates the frame-relative, but otherwise unique, time-independent set T of empirical possibilities. It is easy to reformulate the two successive generalizations of the frame hypothesis, which were required for the naive explanation of the success of the natural sciences in general, leading to a refined version of that general explanation.

From the foregoing it follows that the refined rule of success, prescribing to choose the more successful theory in the refined sense, is again functional for approaching the truth in the sense that the chosen theory may still be closer to the truth (which would explain its being at least as successful) and that the rejected theory cannot be closer to the truth (for otherwise it would not be less successful). It is also easy to check that the adapted versions of the heuristic rules discussed in the naive case, the rules of content/testing/dialectics, are in their turn functional for the rule of success in the sense that they stimulate new applications of the latter.

### 3.5. *Further Conceptual and Technical Analysis of the Clauses*

It is instructive to analyze the defining clauses in some more detail. Numbers will refer to Figures 3 and 4. The instantial clause (Ri) claims

something about X/T-pairs $\langle x, z\rangle$, hence the sets 1, 2, 3, 4, 5, and 7 are involved. The following reformulation of (Ri), presupposing conditional reflexivity, suggests a decomposition into a conjunction of three components:

(Ri′) For all x in X and z in T if r(x, z), and if neither x nor z is in Y, then there is y in Y such that s(x, y, z)

First, as already noted, (Ri) implies, due to centering of s, the naive instantial condition formally telling that $2 = X \cap T - Y$ is empty and conceptually that Y does not introduce new instantial mistakes, that is, mistakes not already made by X. Hence, $\langle x, z\rangle$-pairs from 2 are excluded. The second component implied by (Ri) concerns X/T-pairs of which at least one belongs to Y, and hence to sets 1,3, or 4. In this case conditional reflexivity already guarantees that there is a member of Y in between, which is built into (Ri′). The third component implied by (Ri) concerns x in $X - Y \cup T = 7$ and z in $T - X \cup Y = 5$ of which it guarantees that if they are comparable, then they have a member of Y in between. In terms of mistakes this comes down to the requirement that when an extra explanatory mistake of X is comparable with a common instantial mistake of X and Y, then Y has a member in between. The conjunction of the three components is trivially equivalent to (Ri′), and hence, assuming conditional reflexivity, the three components together imply (Ri).

The instantial success clause $(Ri)_s$ can similarly be rewritten as:

$(Ri')_s$ For all x in X and z in R(t) if r(x, z) and if neither x nor z is in Y, then there is y in Y such that s(x, y, z)

and decomposed. What $(Ri)_s$ substantially adds to $(Ni)_s$ is analogous to the distinguished third component of (Ri): if an extra model x of X which may still be mistaken, i.e., x in $X - Y \cup R(t) = 7 \cup (2\text{–}2.1) = 7 \cup (2.2)$, is comparable with an established common instantial mistake z, i.e., z in $R(t) - X \cup Y = 5.1$, Y should have a model in between.

Let us turn our attention to the refined explanatory truthlikeness clause:

(Rii) For all y in $Y - X \cup T$ there are x in $X - T$ and z in t such that s(x, y, z)

in particular in relation to the explanation of laws. It claims something about members of $Y - X \cup T = 6$, that is, extra explanatory mistakes.

They are not excluded, as in the naive case, but they have to be harmless in the sense that they have to be between X − T and T, or, equivalently, (Rii) guarantees the existence of an explanatory mistake of X that is at least as serious. In this sense (Rii) was said to explicate the idea that *Y explains T at least as well as X*.

Recall that a set L including T is called a law (for it is true-as-hypothesis), and that it is explained by theory X if it also includes X. (Rii) immediately implies:

(RLii) For all convex laws L and for all y in $Y - X \cup L$ there are x in X − L and z in L such that s(x, y, z)

Analogous to (Rii) this may be paraphrased as: *Y explains all convex laws at least as well as X*.

When T is convex, (RLii) trivially implies (Rii), because T itself is a law, hence, if T is convex, (RLii) and (Rii) are equivalent.

From (RLii), and hence from (Rii), immediately follows

(RLii*) For all convex laws L if X − L is empty, then Y − L is empty

That is, *Y explains all convex laws explained by X*.

(RLii*) again indicates a way to formulate the idea that Y explains T at least as well as X in terms of laws: as far as convex laws are concerned, Y's ultimate law-explanatory success includes that of X, or, equivalently, *X's ultimate law-explanatory problems include those of Y*, that is, Y has no extra ultimate law-explanatory problems.

In contrast to the naive case, non-convexity phenomena, even if we assume that T is convex, prevent one from proving that (RLii*) implies (Rii), in which case it would be equivalent to (Rii) and (RLii).

Note that (RLii*) implies, for the special case L = T, that Y − T is empty when X − T is, provided that T is convex. However, from (Rii) the unconditional special case directly follows:

(Rii*) If X − T is empty then Y − T is empty

That is, Y explains T as soon as X explains T.

Analogous to the naive case, it is possible to restrict all formal and informal explanatory claims to *accepted* laws, i.e., laws implied by S(t). Among others, Y explains all accepted convex laws at least as well as X, and hence it explains all accepted convex laws explained by X. Also with reformulations of the latter in terms of 'accepted', instead of

'ultimate', law-explanatory success, problems and extra problems always follow immediately, as far as convex laws are concerned.

Recall, finally, that the refined clause reduces to the naive clause when trivial structurelikeness is substituted for s. As I already noted, non-convex sets cannot exist under trivial structurelikeness. (Rii), (RLii), and (RLii*) then become equivalent to their naive versions, which were proven to be mutually equivalent. In that case the convexity condition can also be skipped in all informal claims. Similar results follow about accepted laws in case of trivial structurelikeness.

In sum, the explanatory clauses are again in perfect agreement with intuitions about the explanation of laws, which are not too peculiar, by a theory that is closer to the truth or more successful than another.

I conclude this subsection with some technical observations that may be useful for certain applications.

If s is unconditionally reflexive, and hence all pairs of structures are comparable, (Ri) reduces to the following:

For all x in X and z in T there is y in Y such that s(x, y, z)

Further, assuming X and T non-empty and T convex, (Rii) is equivalent to:

For all y in Y there are x in X and z in T such that s(x, y, z)

As a consequence, if we define the reflexive closure of s by:

rcs(x, y, z) iff y = x or y = z or s(x, y, z)

(Ri) and (Rii) may be reformulated for convex T as follows:

For all x in X and z in T if r(x, z), then there is y in Y such that rcs(x, y, z)
For all y in Y there are x in X and z in T such that rcs(x, y, z)

respectively, where r(x, z) remains the original comparability relation.

### 3.6. *Stratified Theories*

Let us now investigate how the refined definition works out for theories that are stratified in terms of a distinction between theoretical and (relatively) non-theoretical or observational components. The main

question is again whether truthlikeness on the theoretical level (including theoretical and observational components) is preserved on the observational level.

Recall that $\pi$ indicated the projection function from Mp, the set of conceptual possibilities including theoretical components, onto Mpp, the set of structures without theoretical components. Recall also that, although T-projection, i.e., $\pi T$ is a subset of $T_0 = Mpp(D)$ is plausible, Mp need not be complete with respect to $\langle D, Mpp\rangle$ i.e., $T_0 = \pi T$ need not be the case. To avoid complications in this respect I restricted the attention to $\pi T$ and correct data with respect to $\pi T$, i.e., $R(t) \subseteq \pi T \subseteq S(t)$. Now it is also plausible to assume that s satisfies *s-projection*: $s(x, y, z)$ implies $s_0(\pi(x), \pi(y), \pi(z))$. Note, contrary to what one might think at first sight, that the projection $\pi X$ of a convex set X need not be convex, nor the other way around.

Let us start again with the explanatory clause. Let Y explain T at least as well as X, i.e., (Rii). Then it follows immediately that $\pi Y$ explains $\pi T$ at least as well as $\pi X$, provided that $\pi T$ is convex. And again, now using the refined results on success explanation, the explanatory clause also follows: $\pi Y$ explains S(t) at least as well as $\pi X$, provided S(t) is convex. Note that for the last result $\pi T$ need not be convex.

Let us now consider the instantial clause and recall that projection failed in the naive case and that no interesting sufficient (extra) condition existed. In the refined case, straightforward projection is also invalid, but there are now interesting sufficient conditions, i.e., conditions which together with s-projection guarantee the projection of the instantial clause. I will consider three of them, in order of decreasing strength.

It will be useful to write down explicitly the assumption that Y instantiates T at least as well as X:

(Ri) For all x in X and z in T if $r(x, z)$, then there is y in Y such that $s(x, y, z)$

and what I am eager to prove, i.e., $\pi Y$ instantiates $\pi T$ at least as well as $\pi X$:

$(Ri)_0$ For all $x_0$ in X and $z_0$ in $\pi T$ if $r_0(x_0, z_0)$, then there is $y_0$ in $\pi Y$ such that $s_0(x_0, y_0, z_0)$

Note first that, due to the nature of $\pi$, $(Ri)_0$ is equivalent to

$(Ri)_0'$ For all x in X and z in T if $r_0(\pi(x), \pi(z))$, then there is y in Y such that $s_0(\pi(x), \pi(y), \pi(z))$

It is easy to check that $(Ri)_0'$ follows directly from the assumption that all theoretical(-cum-observational) structures are comparable:

(A1) For all x and z in Mp r(x, z)

Note that (A1), which trivially implies that all observational structures are also comparable, is a rather strong condition. But there may well be cases where it is satisfied: for instance, the case of propositional structures based on fixed finite sets of observational and theoretical elementary propositions.

A weaker sufficient condition assumes that observational comparability guarantees theoretical comparability:

(A2) For all x and z in Mp if $r_0(\pi(x), \pi(z))$, then r(x, z)

Although (A2) is weaker than (A1), it is also a strong condition. Because comparability presupposes at least that the domain-sets are the same, (A2) excludes observationally comparable structures with different theoretical domain-sets.

A still weaker sufficient condition leaves this possibility open. But, whereas (A1) and (A2) were general conditions, the weakest sufficient condition is specific in the sense that it imposes a restriction on the theories in question. As is easy to check, $(Ri)_0'$ is also guaranteed if X and T are *mutually exhaustive with respect to* $r_0$:

(A3) For all x in X and z in T if $r_0(\pi(x), \pi(z))$, then there are x′ in X and z′ in T such that $\pi(x') = \pi(x)$ and $\pi(z') = \pi(z)$ and r(x′, z′).

It is interesting to consider the nature of this condition in some detail. Let ME(X/Z) indicate that X and Z are mutually exhaustive with respect to $r_0$ in the sense defined by (A3). As we have seen in Section 2, it may well be that r is either an equivalence relation or a partial ordering. It is not difficult to prove that ME is in this case also an equivalence relation or a partial ordering on subsets of Mp, respectively. In the equivalence case the theories considered may all be mutually exhaustive, in which case all or none of them belong to the equivalence class associated with T. In the case of a partial ordering, the

theories considered may form an ordered sequence which may or may not end with T, but many of such paths will end in T.

In the naive case there were no interesting sufficient conditions for projection of the instantial clause. From (A1), (A2), and (A3) it is easy to see why. Naive truthlikeness was based on trivial structurelikeness, which implied that two different structures are always incomparable. As a consequence, (A1) is excluded as soon as Mp contains more than one element, and (A2) and (A3) both reduce to the condition that $\pi$ is a one-one function, which is of course an improper extreme case of stratification. (See note added in proof.)

If the instantial clause is projected, it follows also directly that $\pi$Y instantiates R(t), which was assumed to be a subset of $\pi$T, at least as well as X. Hence, if Y instantiates T at least as well as $\pi$X and if (A1) or (A2) or (A3) hold, then $\pi$Y is instantially at least as successful as $\pi$X.

In sum, in contrast to the naive case, projection of refined truthlikeness, with the relevant success consequence, is guaranteed under some interesting conditions.

### 3.7. *Refined Quantitative Truthlikeness*

The refined definition is, like the naive one, qualitative, now in the specific sense that it is based on a comparison of sets of structures in terms of a ternary relation of structurelikeness and not in terms of a quantitative distance function between structures. Although I am sceptical about the use-value of quantitative truthlikeness, I will formulate a plausible quantitative version of refined truthlikeness.

Let d be a quasi-distance function on Mp, i.e., $d(x, y) \geqslant 0$, such that $d(x, y) = 0$ iff $x = y$, $d(x, y) = d(y, x)$. Let sd be based on d in the sense that $sd(x, y, z)$ iff $d(x, y) \leqslant d(x, z)$ and $d(y, z) \leqslant d(x, z)$.

The most plausible definition for *the quantitative distance between X and T* seems to be the following (where I confine myself to countable theories, extrapolation to non-denumerable theories is possible): $RTD(X, T) = \Sigma_{z \in T} d_{min}(z, X) + \Sigma_{x \in X} d_{min}(x, T)$. Here the minimum distance between, e.g., z and X is defined as the minimum of $d(z, x)$ for all x in X comparable with z. I assume that all pairs of theories are such that for every member of the one there is at least one comparable member of the other.

As in the qualitative case, RTD reduces to the corresponding naive quantitative notion under the appropriate conditions. The trivial distance function corresponding to trivial structurelikeness (which was defined by $t(x, y, z)$ iff $x = y = z$) can be defined as follows: $d_t(x, y)$ is 0 or 1 depending on whether x is equal to y or not. It is easy to check that $RTD_t(X, T)$, i.e., $RTD(X, T)$ based on $d_t$, then reduces to $NTD(X, T) = |T \Delta X|$, i.e., the naive quantitative notion.

An interesting question is whether or under what further conditions qualitative truthlikeness is compatible with quantitative truthlikeness in the sense that $RTL(X, Y, T)$, based on sd, implies $RTD(Y, T) \leqslant RTD(X, T)$. In contrast with the corresponding naive case, this is not generally the case. However, it is guaranteed when (Rii) is strengthened to:

(Rii-Q) There is a one-one function from $Y - X \cup T$ to $X - Y \cup T$ such that for all y in $Y - X \cup T$ there is z in T such that $s(f(y), y, z)$

This strong version of (Rii) is trivially satisfied in the case of naive truthlikeness. However, it need not be satisfied in other cases, e.g., in the context of concretization (see Section 4). This kind of example makes me sceptical about the use-value of any quantitative approach, at least as far as quantitative approaches in the line of the presented qualitative structuralist approach are concerned.

## 4. APPLICATION: IDEALIZATION AND CONCRETIZATION

In this section I will study a special kind of theorylikeness, viz., theorylikeness based on idealization and concretization. From the general exposition it then trivially follows that concretization of theories can be a truth approximation strategy. This will be illustrated by the transition of the theory of ideal gases to that of Van der Waals. Then I will outline how concretization is also an important strategy in the investigation of the domain of validity of an interesting theorem and in particular whether it is true for the actual or even the empirically possible worlds.

### 4.1. *Idealization and Concretization*

Concretization or factualization, as it has been presented by the Polish philosophers Wladislaw Krajewski (1977) and Leszek Nowak (1980), is basically a relation between real-valued functions. Hence, let us assume that the conceptual structures to be considered contain one or more real-valued functions, with or without one or more real constants. Structure y is called a *concretization* of x and x an *idealization* of y, indicated by con(x, y), if y transforms, directly or by a limit procedure, into x when one or more constants or functions occurring in y uniformly assume the value 0. It is easy to see that it is a necessary condition for con(x, y) that x and y have the same domain-sets. Moreover, it is easy to check that con is reflexive, antisymmetric, and transitive. In a subsection to follow, the example is presented of a Van der Waals gas model as a concretization of an ideal gas model.

Concretization is primarily a binary relation, but for my purposes I need the plausible ternary version leading to a *concretization triple*: ct(x, y, z) iff con(x, y) and con(y, z). I will assume ct as the underlying notion of structurelikeness. The relation of relatedness based on ct is easily seen to be equivalent to con. Note that we have here a clear example in which relatedness is not symmetric, but directed. Note also that ct is trivially decomposable. Truth-/theory-likeness based on this ternary relation will be indicated by $RTL_{ct}(X, Y, Z)$ and $RTL_{ct}(X, Y, T)$, respectively. It is easy to check that ct satisfies the minimum conditions of being centered, centering, and conditional left and right reflexivity. Moreover, it is antisymmetric (central, left, and right) and it satisfies all conceivable kinds of transitivity, e.g., left: if ct(w, x, z) and ct(x, y, z), then ct(w, y, z).

My next task is to define the binary relation of concretization between theories. Again I will do this as weakly as possible: Y is a concretization of X and X an *idealization* of Y, indicated by CON(X, Y), iff all members of X have a concretization in Y and all members of Y have an idealization in X. At first sight one might think that the second clause should be strengthened to: and all members of Y have a *unique* idealization in X. However, this would exclude, e.g., 'inclusive' concretization triples $\langle X, Y, Z\rangle$ with X as a subset of Y and Y of Z and CON(X, Y) and CON(Y, Z).

It is trivial that CON is reflexive and transitive. However, it need not be antisymmetric, contrary to what one might expect. But sufficient

for antisymmetry of CON(X, Y) is that X and Y are convex (i.e., closed for intermediates). Now it comes down to: if con(x, y) and con(y, z), i.e., ct(x, y, z), and x and z in X, then y in X.

The ternary relation of concretization of theories I define again as weakly as possible: CT(X, Y, Z) iff CON(X, Y) and CON(Y, Z). It is easy to check that CT has the properties of being centered, centering for convex sets, conditional left and right reflexivity, and antisymmetry (central, left, right) for convex sets and all conceivable forms of transitivity. As a consequence, CT(X, Y, Z) is for fixed Z a partial ordering as far as convex theories are concerned.

The main question is whether or under what conditions CT(X, Y, Z) implies $RTL_{ct}(X, Y, Z)$. It turns out that some conditions have to be added to guarantee this implication, but there are some alternative possibilities. I am, of course, primarily interested in conditions on X and/or Y or their combination, for in the crucial case we do not dispose of Z, i.e., T. One sufficient combination of conditions is the following: Y should be convex as well as *mediating*, the latter condition being defined as: if z is a concretization of x and if x has a concretization in Y and z an idealization in Y, then Y also provides an intermediate for x and z; or, more formally, if con(x, z) and if there are y and y′ in Y such that con(x, y) and con(y′, z), then there is y″ in Y such that con(x, y″) and con(y″, z) (i.e., ct(x, y″, z)).

Note that both conditions only concern Y. Although being mediating is a more specific property than convexity, it is not a very restrictive condition in the present context. Note also that it follows that any X can be an idealized starting point for successive concretization. However, the starting point X will usually even be *closed for idealizations*, in the sense that if x in X and con(x′, x) then x′ in X. It is easy to check that this trivially implies that X is convex and mediating.

Let us formally state the main claim: it is (easily) possible to prove the following *Concretization as Theorylikeness Theorem* (C → TL-Theorem): if CT(X, Y, Z) and if Y is convex and mediating, then $RTL_{ct}(X, Y, Z)$. In words: the intermediate theory of a concretization triple is closer to the third than the first, assuming that it is convex and mediating.

We may define stronger versions of concretization triples such as CT*(X, Y, Z) = CT(X, Y, Z) and Y convex and mediating or even CT**(X, Y, Z) = CT*(X, Y, Z) and X and Z also convex. According to the theorem both are special kinds of theorylikeness. Moreover, it was

already mentioned that CT(X, Y, Z) is antisymmetric (in the central sense) as soon as the three sets are convex; hence CT** is an antisymmetric special type of theorylikeness.

### 4.2. *Truth Approximation*

A direct consequence of the C→ TL-Theorem is that, if theory Y is a concretization of theory X, if Y is convex and mediating, and if the true set of empirical possibilities T is a concretization of Y, then Y is closer to the truth than X. This may be called the *Truth Approximation by Concretization (TAC-)Corollary* – a major goal of this section, viz., to show that and in what sense concretization may be a form of truth approximation. All conditions for truth approximation can be checked, except of course the crucial heuristic hypothesis that T is a concretization of Y.

To get 'good reasons' to assume that this crucial hypothesis is also true it is important that the concretization has some type of (necessarily insufficient) justification, of a theoretical or empirical nature, suggesting that the account of the new factor is in the proper direction. In this respect it is plausible to speak of theoretical and/or empirical concretization. The famous case of Van der Waals to be presented evidently is a case of theoretical concretization, followed by empirical support.

### 4.3. *Application to Gas Models*

The transition from the theory of ideal gases to Van der Waals's theory of gases has frequently been presented as a paradigmatic case of concretization. The challenge of any sophisticated theory of truthlikeness hence is to show that this transition can be a case of truth approximation.

For this purpose I start with formulating the relevant models in elementary structuralist terms. $\langle S, n, P, V, T\rangle$ is a *potential gas model* (PGM) iff S represents a set of thermal states of n moles of a gas and P, V, and T are real-valued functions defined on S representing pressure, volume, and (empirical absolute) temperature, respectively.

Specific gas models are PGM's satisfying an additional condition. The *ideal gas models* (IGM) satisfy in addition $P(s)V(s) = nRT(s)$ for

all s in S, or simply PV = nRT, where R is the so-called ideal gas constant. For *gas models with mutual attraction* (GMa) there is a non-negative real (number) constant a, within a certain fixed interval, such that $(P + (n^2a/V^2))\ V = nRT$. For *gas models with non-zero volume of molecules* (GMb) there is a non-negative real constant b, within a certain fixed lnterval, such that $P(V - nb) = nRT$. Finally, in the case of *Van der Waals gas models* (WGM) there are non-negative real constants a and b, within the previously mentioned two intervals, such that $(P + (n^2a/V^2))\ (V - nb) = nRT$.

Note first that it is a neccesary condition for con(x, y) (x and y in PGM) that $S_x = S_y$. Note also that IGM, GMa, GMb, and WGM have been defined such that they are all convex and mediating.

It is easy to check that IGM, GMa, and WGM as well as IGM, GMb, and WGM constitute a concretization triple: an element of WGM transforms into an element of GMa and GMb by substituting the value 0 for b and a, respectively. The resulting elements of GMa and GMb transform into elements of IGM by substituting 0 for a and b, respectively.

Due to the C→ TL-Theorem it follows that GMa and GMb are both closer to WGM than IGM. As a consequence, if WGM were to represent the true set of empirically possible gases, GMa and GMb would be closer to the truth than IGM. Finally, and most importantly, the TAC-Corollary guarantees that WGM is closer to the truth than IGM, assuming the heuristic hypothesis that the true set of empirically possible gases is, in its turn, a concretization of WGM.

Let us finally confront the strengthening of (Rii) to (Rii-Q) which seemed a plausible way to get an acceptable notion of quantitative truthlikeness with the gas model example. It is easy to check that, for instance, the concretization triple of (sets of) gas models ⟨IGM, GMa, WGM⟩ does not satisfy (Rii-Q), simply due to the fact that the set GMa is (much) larger than IGM. As a consequence, it may well be the case that all plausible distance functions in this case, if any, are such that RTD(GMa, WGM) is larger than RTD(IGM, WGM), notwithstanding the fact that GMa is closer to WGM than IGM, not only according to my qualitative definition of refined truthlikeness, but also according to a generally accepted informal judgement.

See, however, Niiniluoto (1986) for a completely different quantitative approach to concretization in general and the Van der Waals case in particular.

### 4.4. *Validity Research*

Scientific research is not always directed to describing the actual world or characterizing the set of empirically possible worlds. It may also primarily aim at proving interesting theorems for certain conceptual possibilities, as Hamminga (1983) showed for neo-classical economics.

Let a certain Mp be chosen, let T indicate the (unknown) subset of empirical possibilities, and let R indicate the (unknown) subset (of T) of realized empirical possibilities, possibly containing just one element, the actual possibility.

Let IT indicate an 'interesting theorem', that is some insightful claim, of which it is interesting to know whether it is true for the empirical possibilities, or at least the realized possibilities. Let V(IT), or simply V, indicate the set of conceptual possibilities for which IT can be proved. V is called the *domain of (provable) validity* of IT, and it is assumed not to be already explicitly characterized.

A frequent type of scientific progress is the following. Suppose that it was earlier proved that IT holds for X, i.e., that X is subset of V. The new result is that Y, which includes X, also is, like X, included in V. Due to concentricity of the naive and refined theorylikeness notions it follows in this case that N/RTL(X, Y, V). The ultimate purpose of this type of research was to find out whether T or at least R are subsets of V. Of course, the larger V has been proven to be, as in the described case, the greater the chance, informally speaking, that R or even T are subsets of V. However, just enlarging the proven domain of validity does not necessarily go in the direction of R and T. For this purpose concretization is the standard strategy.

Let it first be shown that X is a subset of V, and later that a concretization Y of X (CON(X, Y), X need not be a subset of Y) is also a subset of V. It then trivially follows that RTL(X, X ∪ Y, V). If, moreover, Y is convex and mediating, it follows from the heuristic hypothesis that Y is a concretization of T (CON(Y, T)), using the C → TL- Theorem, that RTL(X, Y, T). Hence, we have proved IT for a set Y which is more similar to T than X, which increases the chance that IT holds for T, *ipso facto* for R.

A complex form of validity research concerns the case that IT is not fixed, but that realistic factors are successively accounted for. Formally this is also a form of concretization. IT2 is called a concretization of IT1 if V(IT2) = V2 is a concretization of V(IT1) = V1.

Now suppose that IT1 is proven for X. The relevant heuristic strategy is to look for a concretization Y of X and a concretization IT2 of IT1 such that IT2 can be proved for Y. The heuristic hypotheses are that T is a concretization of Y and that there is a concretization IT* of IT2 such that IT* holds for T and hence for R. This makes sense because if Y and IT2 are convex and mediating, it not only follows that Y is closer to T than X but also that V2 is closer to V* than V1. Hence, in this case we are not only on the way to T but also to IT*.

The concretization of the theory and corresponding theorem of Modigliani and Miller concerning the capital structure of firms by Kraus and Litzenberger turns out to be a perfect example of this kind of approximation of a provable interesting truth (Cools, Hamminga, and Kuipers, forthcoming).

## CONCLUDING REMARKS

In this article I have presented conceptually plausible definitions of naive and refined truthlikeness of theories, the latter based on an underlying notion of structurelikeness. Taking into account the assumed fixed character of the conceptual frame (the set of conceptual possibilities), it allows minimally the conclusion that conceptually relative but otherwise objective truth approximation is possible in a sophisticated sense, for example, by concretization. Moreover, it justifies the corresponding, intersubjectively applicable, methodological rule to choose, whenever possible, the more successful of two theories.

In sum, my threefold explication of the idea of truthlikeness is coherent and conceptually attractive and fruitful. It is tempting to mention one other fruit, the plausible explication of the everyday expression "the truth lies in between": in terms of structurelikeness, when the true description t is concerned: s(x, t, z) & s(z, t, x); and in terms of naive or refined theorylikeness when the true theory T is concerned: NTL(X, T, Z) & NTL(Z, T, X) and RTL(X, T, Z) & RTL(Z, T, X), respectively.

Moreover, as we have seen, my explication allowed the explanation of the global success of (natural) science by assuming that conceptual frames for natural domains contain, as a rule, a unique, time-independent subset of empirical possibilities. For this so-called frame-hypothesis is sufficient to explain local progress in success on the basis of (frame-relative) truth approximation. In Kuipers (1989) I already

presented the global argument as far as naive truthlikeness is involved. However, it is not only important to have shown here that this argument can be easily extrapolated to refined truthlikeness, but also that especially the notion of refined theorylikeness can relativize all kinds of incommensurability claims between theories formulated within different conceptual frames. As a crucial example in this respect, I have demonstrated that refined truthlikeness of stratified theories is under general conditions projected on observational theories. Hence, refined truthlikeness clearly paves the way for fundamental, or at least pragmatic, commensurability of related theories.

## NOTES

* I would like to acknowledge that van Benthem (1987) played in several respects a crucial role in the research for a new refined definition. Moreover, I like to thank David Miller, Ilkka Niiniluoto, and two referees for their comments on an earlier version. One of the referees notes that the Miller version of the naive approach (in model-theoretic terms, and identifying the truth with the truth about the actual world) has been criticised on several occasions for its failure to accommodate likeness between structures by Oddie (notably Oddie, 1986). The idea that likeness between structures should be a guiding idea behind truthlikeness is said to be a constant theme of Oddie's work. All this may well be true, but I should add however that Oddie's publications did not play any role in my research. The local references to Oddie are based on the suggestions by the referee.

## REFERENCES

Benthem, J. van: 1987, 'Verisimilitude and conditionals', in Kuipers (1987b), pp. 103–28.

Cools, K., B. Hamminga, and T. Kuipers: forthcoming, 'Truth Approximation by Concretization in Capital Structure Theory', *Idealization in Economics* (ed. B. Hamminga), *Poznan Studies*, Rodopi, Amsterdam.

Hamminga, B.: 1983, *Neoclassical Theory Structure and Theory Development*, Springer, Berlin.

Krajewski, W.: 1977, *Correspondence Principle and Growth of Science*, D. Reidel, Dordrecht.

Kuipers, T.: 1982, 'Approaching Descriptive and Theoretical Truth', *Erkenntnis* **18**, 343–87.

Kuipers, T.: 1984, 'Approaching the Truth with the Rule of Success', *Philosophia Naturalis* **21**, 244–53.

Kuipers, T.: 1987a, 'A Structuralist Approach to Truthlikeness', in Kuipers (1987b), pp. 79–99.

Kuipers, T. (ed.): 1987b, *What is closer-to-the-truth?*, *Poznan Studies*, Vol. 10, Rodopi, Amsterdam.

Kuipers, T.: 1989, *How to Explain the Success of the Natural Sciences, Proceedings of the 13th Wittgenstein Symposium*, Hölder-Pichler-Tempsky, Vienna, pp. 318–22.
Niiniluoto, I.: 1986, 'Theories, Approximations, and Idealizations', in R. Barcan Marcus et al. (eds.), *Logic, Methodology and Philosophy of Sciences VII*, North-Holland, Amsterdam, pp. 255–89 (revised and extended version in: 1990, J. Brzezinski et al. (eds.), *Idealization I: General Problems*, Rodopi, Amsterdam, pp. 9–57).
Niiniluoto, I.: 1987, *Truthlikeness*, D. Reidel, Dordrecht.
Nowak, L.: 1980, *The Structure of Idealization*, D. Reidel, Dordrecht.
Oddie, G.: 1981, 'Verisimilitude Reviewed', *The British Journal for the Philosophy of Science* **32**, 237–65.
Oddie, G.: 1986, *Likeness to Truth*, D. Reidel, Dordrecht.
Oddie, G.: 1987, 'Truthlikeness and the Convexity of Propositions', in Kuipers (1987b), pp. 197–216.
Tichý, P.: 1974, 'On Popper's Definition of Verisimilitude', *The British Journal for the Philosophy of Sciences* **25**, 155–60.

Department of Philosophy
University of Groningen
A-weg 30
9718 CW Groningen
The Netherlands
e-mail: T.A.F.Kuipers@philos.rug.nl

## ADDED IN PROOF

This is not quite correct for (A3). That reduces to:

> For all $x$ in $X$ and $z$ in $T$ if $\eta(x) = \eta(z)$, then there is $y$ in $X \cap T$ such that $\eta(x) = \eta(y) = \eta(z)$.

This seems to be an interesting, sufficient condition, not merely filling the gap, i.e., a condition suggested to be impossible at the end of Section 1.7.