

# Comparative realism as the best response to antirealism<sup>1</sup>

Theo A.F. Kuipers

*University of Groningen*

*T.A.F.Kuipers@rug.nl*

**ABSTRACT.** Arguments for and against scientific realism usually presuppose as the main epistemic claim about theories that we may have good reasons to conclude that they are true, or at least approximately true. The antirealist charges against this claim are not easy to counter. In this paper it is argued that the defense of realism is much easier if we relativize its epistemic claim in the light of (theories of) truth approximation. From this *comparative realist* perspective the main epistemic claim becomes that we may have good reasons to conclude that successor theories are closer to the truth than their predecessors. For example, although Einstein's theory of general relativity may still be false, and not even approximately true, we have good reasons to assume that it is closer to the truth than Newton's theory of gravitation. A similar relativization of claims that theoretical terms refer to things in the world is argued for in terms of 'being closer to the referential truth'. For both purposes it is also plausible to relativize 'being empirically successful' to the comparative notion of 'being (persistently) empirically more successful'.

Comparative realism hence is realism guided by the comparative perspective on success and on truth approximation, that is, the notions of 'more successful' and 'closer to the (observational, referential, and theoretical) truth', and their mutual relations. This approach is defended against the antirealist charges and compared with the main other realist responses. The main positive claim of comparative realism is that (theoretical) truth approximation provides the (stratified) default explanation and prediction of empirical progress between non-empirically equivalent theories and of 'aesthetic progress' between empirically equivalent theories. Here 'aesthetic progress' is understood in terms of the prevailing 'aesthetic canon' in the relevant field and period, that is, the prevailing non-empirical virtues of theories.

## 1 Introduction

Arguments for and against scientific realism usually presuppose as the main epistemic claim about theories that we may have good reasons to conclude that they are true, or at least approximately true. The antirealist charges against this claim are not easy to counter. In this paper it will be argued that the defense

---

<sup>1</sup>The author is with the Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, Netherlands. A Dutch version with more emphasis on 'Inference to the Best Explanation', entitled "Comparatief realisme: het beste antwoord op anti-realisme", appeared as target paper in *Algemeen Nederlands Tijdschrift voor Wijsbegeerte (ANTW)*, 100.3, 2008, 173-200. T.A.F.Kuipers@rug.nl, <http://www.rug.nl/filosofie/Kuipers>

of realism is much easier if we relativize its epistemic claim in the light of (theories of) truth approximation. From this comparative realist perspective the main epistemic claim becomes that we may have good reasons to conclude that successor theories are closer to the truth than their predecessors. For example, although Einstein's theory of general relativity may still be false, and not even approximately true, we have good reasons to assume that it is closer to the truth than Newton's theory of gravitation. A similar relativization of claims that theoretical terms refer to things in the world will be argued for in terms of 'being closer to the referential truth'. For both purposes it is also plausible to relativize 'being empirically successful' to the comparative notion of 'being (persistently) empirically more successful'.

Comparative realism hence is realism guided by the comparative perspective on success and on truth approximation, that is, the notions of 'more successful' and 'closer to the (observational, referential, and theoretical) truth', and their mutual relations. This approach will be defended against the antirealist charges and compared with the main other realist responses.

In Section 3 I will draw heavily upon the survey of (non-comparative) arguments pro and contra realism presented by James Ladyman in his textbook of [12] and his handbook exposition [13] of ontological, epistemological, and methodological positions. In particular, the *pessimistic meta-induction*, the *no-miracles argument* and *inference to the best explanation* and their counter arguments will be analyzed in detail and will lead to a number of conditions of adequacy for explicating the crucial expressions.

There are at least two ways to specify the truth approximation perspective, the quantitative approach of Niiniluoto [15, 16] and my own qualitative, comparative approach [6]. For the present purposes the latter, in a sense more strict, and hence more cautious, approach is sufficient. More specifically, in Section 2 I will sketch some of its main lines and in Section 4 I will show to what extent this approach fulfils the conditions of adequacy that have been gathered in Section 3. I will conclude with a final comparison of comparative realism with the main kinds of non-comparative realism that have been suggested to meet the antirealist charges. It concerns in particular the restriction of realism to entities, structures, 'mature' theories, theories with 'novel' predictive success, and 'essential' parts/aspects of theories, or stretching realism by stretching the underlying (causal) theory of reference.

The core of *comparative* realism becomes that the (comparative) phenomenon that one theory persistently is empirically more successful than another provides a good reason for the claim that the first theory is closer to the (theoretical and referential) truth than the second; the good, analytical, reason being that this very claim amounts to the generic *default* explanation for that comparative empirical phenomenon. Among empirically equivalent theories, it may occur that one theory persistently is more successful than another according to some non-empirical, notably aesthetic criteria. In this case there may be empirically justified reasons, however weak, for the truth approximation claim and the corresponding default explanation.

I conclude this introduction with a puzzling observation about the realism-antirealism debate. One important success of the first decades of (constructive) analytical philosophy was the discovery, notably by Russell, Carnap, Hempel, and Beth, of relations as a means to solve age-old problems by refined concept

explication. This concerned not least asymmetric relations, constitutive of comparative concepts, such as ‘longer than’, ‘caused by’, etc. Of course, conceptions of ‘better than’, hence of improvement or progress, are comparative as well.

Now it is very surprising that in the supposedly analytic realism – antirealism debate there is almost no sign of awareness of the possible relevance of this insight. One continues to talk about true and false theories and reference claims versus empirically adequate and inadequate theories, that is, observationally true and false theories, respectively. The retreat to ‘approximately true’ is of no help, for that remains basically non-comparative and hence it cannot capture progress. Moreover, it requires a necessarily arbitrary threshold. On the other hand, from the point of view of relations, it is highly plausible to think in terms of comparative notions, such as ‘closer to the truth’ and ‘more successful than’. To be sure, in the comparative approach theories will frequently not be straightforwardly comparable; but for ‘mixed’ cases the ‘principle of dialectics’ (that is, try to improve both) and the quantitative approach are plausible concretizations.

## 2 Truth approximation, some basics for comparative realism

In this short introduction to my favorite, qualitative theory of truth approximation, I will mainly restrict myself to its basic form, that is, without a (O-/T-) distinction between observational and theoretical terms and without other refinements. It is best represented within the structuralist theory of theory representation. Starting from a fixed vocabulary and a suitable similarity type of structures, let  $M_p$  indicate the set of structures of that type, also called the *conceptual possibilities* or *potential models* of the theory. Let the subset  $X$  of  $M_p$  indicate the set of models of theory  $X$ . Finally, assuming that our target is a fixed domain of physically or, more broadly, nomically possible constellations and events, let  $T$  indicate this domain ‘as seen through  $M_p$ ’, hence a subset of  $M_p$ , and be called the set of (intended) nomic possibilities or the domain of intended applications. According to the Nomic Postulate, we assume that such a unique fixed set exists, given  $M_p$  and an intended domain. Note that we do not yet suppose that we dispose of a general characterization of  $T$  as a subset of  $M_p$ . What we know is only that each intended application can be represented as a potential model. A general characterization of  $T$  is ‘the great unknown’ which theories, as represented by their models, are looking for. More formally, a theory is a triple of the form  $\langle M_p, X, T \rangle$ , together with the weak claim that  $T$  is a subset of  $X$  ( $T \subseteq X$ ) and the strong claim that  $T = X$ . A theory is said to be true (false) in the weak sense when the weak claim is true (false). It is easy to check that (a general characterization of)  $T$  represents the strongest true weak claim, and hence may be called ‘the truth’ in this context, that is, the truth about the given domain in the given vocabulary.

Now it is plausible to define what it means, for fixed  $\langle M_p, T \rangle$ , that one theory  $Y$  is closer (or more similar) to the truth than another  $X$ , and hence what it means to say that  $Y$  amounts to truth approximation relative to  $X$ . Intuitively, when  $Y$  is moving from  $X$  in the direction of  $T$ . Formally, when  $Y - T \subseteq X - T$  ( $\emptyset 2$ -area empty in Fig.1) and  $T - Y \subseteq T - X$  ( $\emptyset 1$ -area empty) and at least once it should be a proper subset relation ( $\#1$ -area and/or  $\#2$ -area non-empty). In terms of symmetric differences ( $A \Delta B =_{df} (A \Delta B) \cup (B - A)$ ),  $Y \Delta T$  should be

a proper subset of  $X\Delta T$ . In [6] I have argued that, among other formulations, the two clauses of this definition amount to: (relative to  $T$ )  $Y$  has more true consequences than  $X$  and  $Y$  has more correct models than  $X$ , respectively. Of course, we get ‘at least as close to the truth’, when we omit the proper-subset conditions. We get an asymmetric form of ‘closer to’ by only requiring in addition that the first relation is a proper subset relation (#1- area empty). Knowing what

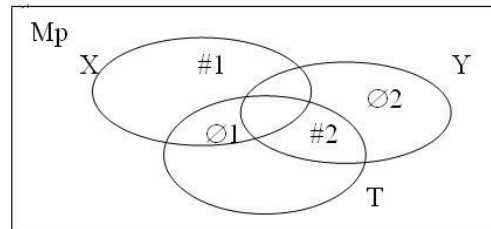


Figure 1: Theory  $Y$  is closer to the truth  $T$  than theory  $X$ , where  $T$  represents the domain as seen through  $M_p$

it means that one theory is closer to the truth than another is one thing, in view of the fact that we don’t know the truth, judging that this is in fact the case, in the light of our evidence is quite another. Although there is, of course, no theorem guaranteeing that (empirical) more-successfulness entails closer-to-the-truth, there is a theorem, the *Success Theorem*, guaranteeing almost the reverse entailment, viz., at-least-as-close-to-the-truth entails at-least-as-successfulness.

Let  $R$  represent the set of realized (described) applications at a certain time, that is, the experimentally or otherwise realized nomic possibilities at that time. Let  $S$  indicate the strongest law induced on the basis of  $R$ . Of course,  $R$  will partly be the result of testing hypothetical laws. If no descriptive mistakes have been made in representing the realized possibilities,  $R$  is not only a subset of  $M_p$ , but even of  $T$ , for nomic impossibilities can’t be realized, by definition. Moreover, in the structuralist representation  $S$  is also a subset of  $M_p$ , such that it has been concluded, inductively, and hence at most provisionally, that conceptual possibilities outside  $S$ , that is, in  $M_p - S$ , are nomically impossible. Finally, if our inductive jump from  $R$  to  $S$  is correct,  $S$  has to be a superset of  $T$ . To sum up:  $R \subseteq T \subseteq S \subseteq M_p$ , provided no descriptive and inductive mistakes have been made.

Let us call  $R/S$  the data set. We define what it means that  $Y$  is, relative to the data set  $R/S$ , (empirically) more successful than  $X$  as follows:  $X \cap R \subseteq Y \cap R$  and  $S \cup Y \subseteq S \cup X$ . The first condition amounts to ‘no loss of established examples’, to be called ‘instantially at least as successful’, and the second to ‘no loss of explained established laws’, to be called ‘explanatorily at least as successful’.

Now the already announced *Success Theorem* states, assuming that  $R/S$  is correct, that is,  $R \subseteq T \subseteq S$ , that  $Y$  is (empirically) at least as successful relative to  $R/S$  as  $X$  if  $Y$  is at least as close to the truth as  $X$ . This theorem is such that *persistently* being-more-successful is functional or instrumental for truth approximation, though not guaranteeing it, in the sense that it is very difficult, and without the O-/T-distinction even impossible, for  $X$  to remain at least as successful as  $Y$  when  $Y$  is in fact closer to the truth than  $X$ . More in detail we may argue as follows. Let  $Y$  be more successful than  $X$  relative to  $R/S$  at a certain time. This suggests the *Comparative Success Hypothesis*, according to which it is hypothesized that this will remain the case, whatever experiments we design and

perform. Testing of this hypothesis may result in the conclusion, at least for the time being, that this is in fact the case:  $Y$  persistently remains more successful than  $X$ . Note that this conclusion is a comparative inductive generalization, providing the paradigm situation for speaking of empirical progress, consisting of ‘instantial’ and ‘explanatory’ progress. That is, the paradigm case for applying the, essentially instrumentalist, ‘rule of success’, viz. replace, for the time being,  $X$  in favor of  $Y$ . Here  $Y$  may well be known to be false in view of  $R/S$ , for which reason the rule is called instrumentalist rather than falsificationist. Assuming empirical progress in this sense, it is very hard to imagine that  $Y$  is, despite appearances, not closer to the truth than  $X$ . More specifically, it holds that 1) the ‘closer to the truth hypothesis’, the *TA-hypothesis*, predicts and explains the empirical progress, 2) it cannot be the reverse, that is, it is impossible that  $X$  is closer to the truth than  $Y$ , and 3) if  $Y$  is in fact not closer to the truth than  $X$ , there is a specific burden to explain the supposed empirical progress so far. In the last case, without theoretical terms, hence without O-/T-distinction, the apparent empirical progress must have resulted from an unhappy test history of the comparative success hypothesis because the empirical progress conclusion can be broken by appropriate experiments. With O-/T-distinction, see below, the breaking of this conclusion is only excluded when  $Y$ , though presumably not closer to ‘the theoretical truth’ than  $X$ , happens to be nevertheless at least as close to ‘the observational truth’ as  $X$ . In sum, the TA-hypothesis provides the ‘*default explanation and prediction*’ of empirical progress.<sup>2</sup>

So far I have presented the naive or basic structuralist theory of truth approximation. It is basic in the sense that several idealizations have been made, requiring refinement or concretization. In [11] I have extensively illustrated the (philosophical) method of concept explication by idealization and concretization by the example of truth approximation. Let us review the main concretizations as they have been elaborated or indicated in [6].

One idealization was that we implicitly assumed that any established counterexample of a theory, that is, a realized possibility not being a model of the theory, is as bad for one theory as for any other. However, in chap.10 I have introduced the ternary relation of ‘(more) structurelikeness’, that is, the idea that one structure may be more similar to another than a third one. In this way, the possibility arises that a counterexample is less dramatic for one theory than for another, because the former has a model that is more similar to the ‘counter model’ (i.e., the potential model representing the counterexample), than any model of the latter theory. Adapted definitions of more-successfulness, empirical progress, and closer-to-the-truth lead again to the conclusion that the TA-hypothesis provides the default explanation and prediction of empirical progress. As a matter of fact we only know of real life scientific examples of (potential) truth approximation, e.g. the Law of Van der Waals as successor of the ideal gas law, when this concretization is introduced. Without this, we only know of toy examples.

---

<sup>2</sup>Surprisingly enough, the expression ‘default explanation’, let alone ‘default prediction’, does not seem to be in use in the philosophy of science. Here it is specifically used as an alternative to the expression ‘functional for truth approximation’. As Gerhard Schurz has remarked, the default explanation ‘closer to’ is also a kind of abductive explanation, with the interesting restriction that the genuine abductive step in this explanation is only from ‘at least as successful’ to ‘at least as close to’, for ‘not equally close’ is a deductive consequence of ‘more successful’.

Another idealization was the assumption that there is just one domain. In [6, chap.13] I have indicated that it is possible to introduce a domain vocabulary, as a subvocabulary of the (observational) vocabulary, such that it is possible to define the domain explicitly in that vocabulary, leaving the further behavior of the applications as the unknown to be specified. The advantage of this is that domains become comparable sets, that is, it enables us to say that an intended application of one theory is or is not an intended application of another. This opens the possibility of dealing with truth approximation by domain change besides by theory change, as indicated above, and hence by the combination of domain and theory change [10].

Returning to theory change, it is evident that theories will seldom be comparable in terms of ‘more successful’ and ‘closer to the truth’. These notions apply only in ideal cases. In practice it is important that in the case of ‘divided success’ between two theories, the crucial challenge is to improve both theories, that is, to design a third theory, a synthesis, that is more successful than both and, if so, plausibly closer to the truth than both<sup>3</sup>. This is called the ‘principle of dialectics’ in [6, chap.6]. In the theoretical context of the realism – antirealism debate, the move to the quantitative approach is more useful to deal with divided success. The basic form of being quantitatively closer to the truth is the condition, assuming finite  $M_p$ ,  $|Y\Delta T| < |X\Delta T|$ . For a refined quantitative approach the best proposal so far is of Ilkka Niiniluoto [15, 16], who has defined the so-called min-sum real-valued distance of a theory to the truth, which always enables comparison. Since the only point of the coming exposition is the principled possibility of straightforward empirical progress and truth approximation I will not go here into details of this approach.

Finally, we did not make a distinction between (relatively) observational and theoretical terms, but such a O-/T-distinction is of course crucial for the realism-antirealism debate,. In chap.9 of [6] I have spelled out how this distinction works out, with the main conclusion that empirical progress, of course, in observational terms, remains functional for truth approximation on the theoretical level, though with some greater risk of being wrong in this tentative conclusion.

For present purposes I will introduce the main features of this type of ‘stratified (basic) truth approximation’. Let  $M_p$  now indicate the set of theoretical (cum observational) conceptual possibilities of the vocabulary. Let subset  $T_t$  of  $M_p$  indicate the unique subset of nomic theoretical possibilities, the theoretical truth. Let  $M_{pp}$  indicate the set of partial or observational conceptual possibilities and  $T_o$  its unique subset of nomic observational possibilities, the observational truth. We assume that  $T_t$  and  $T_o$  can be characterized in some way. Let  $\pi$  project (subsets of)  $M_p$  onto (subsets of)  $M_{pp}$ , stripping the (clauses involving) theoretical components. There are good reasons to assume that  $T_o = \pi(T_t)$ . Of course, theories  $X$  and  $Y$  are empirically equivalent when  $\pi(X) = \pi(Y)$ . The relations of ‘empirically more successful’, and hence of ‘empirical progress’, and of ‘closer to the (observational) truth’ between theories  $X$  and  $Y$  now refer to these notions between  $\pi(X)$  and  $\pi(Y)$  as defined above, but now on the level of  $M_{pp}$ . On the other hand, ‘ $Y$  is closer to the (theoretical) truth than  $X$ ’ remains defined on the

---

<sup>3</sup>At first sight, a more basic problem of comparability seems to arise from so-called incommensurable vocabularies. However, as long as there is the suspicion that two terms may have a different meaning it is plausible and possible to include both terms, with indices, in a fusion of vocabularies.

(theoretical cum observational) level of  $M_p$ , by the otherwise unchanged basic, that is, set theoretic definition. Finally, it is possible to define ‘the referential truth’ as, roughly, the conjunction of (un)negated reference claims of theoretical terms that are correct according to  $T_t$ , that is, those terms that do play or do not play a role in the shaping of  $T_t$  relative to  $T_o$ . To each theory, a (total) reference claim is associated by a similar condition. In sum, the above assumptions enable definitions of ‘closer to the observational, referential, and theoretical truth’, and hence of ‘observational, referential, and theoretical truth approximation’.

Besides the previously mentioned Success Theorem, now restricted to the observational level, some more TA-theorems and -conjectures become relevant, notably the (conditional) Projection Theorem and the Stratified Success Theorem, which is a kind of combination. In Section 4, I will give some more information. Together they support the main claim of comparative realism: theoretical truth approximation provides the (stratified) default explanation and prediction of empirical progress between non-empirically equivalent theories and of ‘aesthetic progress’ between empirically equivalent theories. Here ‘aesthetic progress’ will be defined in terms of the prevailing ‘aesthetic canon’ in the relevant field and period, that is, the prevailing non-empirical virtues of theories.

### 3 Antirealist arguments to be met

In this section I will heavily draw upon James Ladyman’s [12, 13] in looking for the main arguments that have been put forward against realism. I’ll start with a brief preview of the arguments and the responses that naturally follow from the comparative realist point of view. Recall, that this point of view conceives realism guided by the comparative perspective on success and on truth approximation, that is, the notions of ‘more successful’ and ‘closer to the truth’, and their mutual relations.

Laudan’s so-called pessimistic meta-induction is rebutted by the claim that it is not at all a problem that previously the best theories turned out to be false (and even did not refer). For the main question is whether they remained, at least as a rule, more successful than their predecessor theories and hence can still be argued to be closer to the truth than (and refer at least as well as) their predecessors. Hence, similarly for our currently the best theories, they will presumably turn out to be false, but the interesting question remains whether they will continue to be more successful than their predecessors, so that we can still hold them as closer to the truth.

Regarding the arguments against Putnam’s so-called no-miracles argument in favor of realism, the comparative response is that the former will miss their point when the latter is replaced by the observation that the success of science would be miraculous on anything but a scientific realist view in the sense that specific occurrences of (persistent) empirical progress would be miraculous if not, as a rule, due to truth approximation. This, notwithstanding Van Fraassen’s Darwinian analogy that, as a rule, the best theories survive because they are selected for that reason.

Finally, the arguments (notably of Van Fraassen) against inference to the best explanation (IBE) are rebutted by the claim that IBE has so far not been adequately explicated. The proposed comparative explication reads: inference to

the best explanation as the closest to the truth among the available explanations or theories, even among empirically equivalent theories.

In the course of the presentation, there will be derived a number of conditions of adequacy an acceptable explication of comparative realism, including IBE, will have to satisfy. In Section 4, it will be investigated to what extent they can be satisfied.

### 3.1 Arguments from theory change

#### 3.1.1 The pessimistic meta-induction

One of the main arguments against realism is known as ‘the pessimistic meta-induction’ and is due to Larry Laudan. Ladyman [12, pp.236–237], [13, p.345] presents it as follows (my italics):

- (i) There have been *many empirically successful* theories in the history of science which have subsequently been rejected and whose central theoretical terms do not refer according to our best current theories.
- (ii) Our best current theories are no different in kind from those discarded theories and so we have no reason to think they will not ultimately be replaced as well.

So, by induction we have positive reason to expect that our best current theories will be replaced by new theories according to which some of the central theoretical terms of our best current theories do not refer, *and hence, we should not believe in the approximate truth or the successful reference of the theoretical terms of our best current theories.*

According to Ladyman the most common response is that the realist claims should be restricted to mature theories and/or theories having novel predictive success, for which (i) would not apply, and hence the inductive argument would be blocked. This retreat has its own problems, see below. Another way to block this argument is to liberalize the realist elements in the argument according to the comparative perspective:

- (iCR) There have been (some, several, many<sup>4</sup>) theories in the history of science *that were and still are empirically more successful than their predecessor theories*, which have *nevertheless* subsequently been rejected and whose central theoretical terms do not *all* refer according to our best current theories.
- (iiCR) (= (ii)) Our best current theories are no different in kind from those discarded theories and so we have no reason to think they will not ultimately be replaced as well.

So, by induction we have positive reason to expect that our best current theories will be replaced by new theories according to which some of the central theoretical terms of our best current theories do not refer, *however we may still* (instead of: and hence, we should not)

---

<sup>4</sup>Depending on whether we will use qualitative or quantitative versions of ‘more successful’ and ‘closer to the truth’.



*believe that our best current theories are closer to the truth than their predecessor theories or that their theoretical terms more successfully refer than those of their predecessor theories.*

Hence, it is not plausible anymore to think that the second part of the suggested, partially bracketed conclusion of the CR-version of the inductive argument, starting with “(and hence, we should not) believe that . . .” is still ‘valid’. By consequence, it seems plausible to think that, starting from the basic explications in Section 2, when making a distinction between (relatively) observational and theoretical terms, refined explications can be given of the crucial phrases, viz. ‘empirically more successful’, ‘closer to the truth’ and ‘more successfully refer’, such that the following conditions of adequacy are satisfied, in which we add ‘novel predictive success’ in the second one for later purposes.

**CA1** The explications of the crucial phrases should leave room for the possibility that false theories, even with some non-referring theoretical terms, are not only (persistently) empirically more successful but even closer to the (theoretical) truth than and refer at least as well as other theories.

**CA2** The explications should be such that ‘being (persistently) empirically more successful’ is, *as a rule*, due to ‘being closer to the theoretical truth’, which on its turn entails, as a rule, ‘referring at least as well’ and ‘novel predictive success’.

Or, in a plausible ‘progress version’:

**CA2** The explications should be such that empirical progress is, *as a rule*, due to theoretical truth approximation, which on its turn entails, as a rule, referential truth preservation and novel predictive success.

In Section 2 we have already indicated some basic explications that allow ‘false theories closer to the truth than other ones’ and therefore ‘being as least as successful’, however, not yet with the distinction between observational and theoretical terms. In Section 4 we will try to meet these conditions while taking this distinction into account.

### 3.2 The antirealist attack on the no-miracles argument

As mentioned, according to Ladyman, the usual realist response to the pessimistic meta-induction is the retreat to mature theories and/or theories having novel predictive success. The weakness of both, *prima facie* rather ad hoc, retreats is, of course, that they require precise non-ad hoc definitions of ‘mature theories’ and ‘novel predictive success’, both of which tasks have so far not generated a proposal that is agreed upon. Note that comparative realism remains to take all theories seriously.

Besides such technical problems with these rejoinders there is a more general problem with such retreats. The antirealist counter argument to such a retreat is that one does not need the (unrestricted) inductive argument to undermine the so-called no-miracles argument in favor of realism, that is, “the idea that the success of science would be miraculous on anything but a scientific realist view” [13, p.342]. Illustrated by the ether theory of light and the caloric theory

of heat one may claim, according to [12, p.244], [13, p.346] (as quoted, with my bracket insertions [] and deletion ( )):

- (a) Successful reference of its central theoretical terms is a necessary condition for the approximate truth of a theory.
- (b) There are [at least two] examples of theories that (were mature and) had novel predictive success but which are not approximately true.
- (c) [Hence,] Approximate truth and successful reference of central theoretical terms is not a necessary condition for the novel-predictive success of scientific theories

So, the no-miracles argument is undermined since, if approximate truth and successful reference are not available to be part of the explanation of some theories' novel predictive success, there is no reason to think that the novel predictive success of other theories has to be explained by realism.

I have some reservation against this argument as such. Although (a), (b) and (c) may be granted<sup>5</sup>, the conclusion drawn from them seems too strong. Why would maturity and novel predictive success not be possible without approximate truth and successful reference? One could imagine that there are logical gaps leaving room for such cases. Of course, in such cases the novel predictive success would have to be explained in another, specific way. However, the suggestion that such cases prevent the use of approximate truth and successful reference in other specific cases for the explanation of novel predictive success seems unjustified or, at least no such justification is reported by Ladyman. In other words, why is it excluded that *prima facie* similar cases have different explanations?

However this may be, the current realist responses to the antirealist rejoinder are twofold according to [12, pp.245–248], [13, pp.346–348]. (I) Stretch realism by stretching the (causal) theory of reference such that the relevant abandoned theoretical terms refer after all or (II) Restrict realism to those theoretical claims about unobservables that feature in an *essential* way in the derivation of novel predictions. For example, Stathis Psillos [19] uses (I) to save the ether theory as referring after all, viz. to the electromagnetic field, and (II) to explain the success of the caloric theory without 'caloric' as a referring term.

The comparative perspective does not need such interventions, although even within this perspective there may be *other* good reasons for such moves. However, in the present context the point is whether the abandoned theories were and are closer to the truth than their respective predecessor theories. And this may well be the case in several historical examples, notably not only for the ether theory but also for the caloric theory.

Hence, let us turn to the CR-versions of the antirealist premises (a) – (c) enabled by CA1&2. This will lead to a defensible CR-version of the no-miracles argument.

- (aCR) Being closer to the truth entails, as a rule, referring at least as well, and novel predictive success ('novel', now in whatever favorite sense).

<sup>5</sup>However, even with respect to (a), there may be room for doubt, for it very much depends on the definition of 'approximate truth' whether there is the logical possibility of an approximately true theory without reference of (all of) its theoretical terms.

- (bCR) Exceptionally there (may) have been examples of theories that (were mature and) had novel predictive success relative to their predecessor theories but which were not closer to the truth and even not referring at least as well.
- (cCR) Being (persistently) more successful is, as a rule, due to being closer to the truth, even more so when novel predictive success is involved.

Note that (aCR) and (cCR) are taken care of by CA2. Note also that the caloric and the ether theory may well fit in (aCR) and (cCR) relative to their predecessors. But, by the insertion 'as a rule', (aCR) and (cCR) leave room for other exceptions (bCR) and suggest a cautious CR-version of the no-miracles argument.

*CR-no-miracles argument:*

As a rule, (persistently) empirically more successful theories are closer to the truth than, and refer at least as well as their predecessors.

If this were not the case, the regular occurrence of persistently empirically more successful theories, including some novel predictive success, would be miraculous.

Occasionally, other, case-specific, explanations of persistent empirically more successfulness, novel or not, may be appropriate.

Hence, the CR-no-miracles argument suggests that truth approximation provides a kind of *default explanation and prediction* of (persistent) empirical more successfulness, including some novel predictive success. It is attractive to reformulate the CR-version in terms of progress.

*CR-no-miracles argument, progress version:*

As a rule, (persistent) empirical progress is due to theoretical truth approximation and referential truth preservation, if not referential truth approximation.

If this were not the case, the regular occurrence of (novel) empirical progress would be miraculous.

Occasionally, other, case-specific, explanations of (novel) empirical progress may be appropriate.

And, hence, the argument suggests that (theoretical) truth approximation provides a kind of *default explanation and prediction* of (novel) empirical progress.

For standard realism the ether theory of light and the caloric theory of heat give rise to problems because in both cases the crucial theoretical terms do not refer in a straightforward sense according to our present lights, whereas the theories were mature and had quite some novel predictive success. It surely is an intriguing question how these theories can combine both features. However, they may not be so very relevant for the realism debate since these problems do not arise for comparative realism. What matters in this perspective are only comparative questions, that is, questions in comparison with previous and successor theories. For example, in the case of the ether theory the focus in the debate is on (Young and) Fresnel's revised version of Huygens' wave theory of light in ether, by replacing his longitudinal waves by transversal waves. Instead of trying to answer, as a defense of realism - however interesting it may be - the question

as to how Fresnel's theory could have impressive novel predictive success, by our present lights, without reference of the crucial theoretical term 'ether', some of the plausible CR-questions are:

Was Fresnel's theory empirically more successful than its predecessor theory, viz. that of Huygens, and its competitor theory, the particle emission theory of Newton?

As is well known, at least with respect to Foucault's quasi-crucial velocity of light experiment, the result was in favor of Fresnel's theory relative to both Huygens' version of the wave theory and Newton's particle alternative. Hence, the next comparative realist questions are:

If so, could this be explained by the hypothesis that Fresnel's theory is closer to the truth than, and hence, presumably, refers at least as well as, those of Huygens and Newton?

Not knowing the truth, such questions can only be answered in the abstract. However, assuming that the relativistic version of Maxwell's electromagnetic theory, that is, Einstein's special theory of relativity encompassing Maxwell's theory, is the truth (or at least a close approximation of the truth), we can form potential answers to these questions.

Can the greater empirical success of Fresnel's theory be explained by the fact that Fresnel's theory is closer to the theory of Maxwell (-Einstein) than those of Huygens and Newton, and hence, presumably, refers at least as well?

Assuming precise enough explications, these questions can be answered. It is at least plausible to think that Fresnel's theory is closer to Maxwell's theory than Huygens' theory. Moreover, it may well be that the sequence of theories of Huygens, Fresnel, Maxwell, Maxwell-Einstein is a matter of truth approximation. However, it does not seem plausible that the sequence of theories of Newton, (Huygens), Fresnel, Maxwell (-Einstein) can be reconstrued as such. Newton's theory is indeed incommensurable with versions of the wave theory, at least without the leap to the first versions of quantum physics, suggesting that the photon theory of light is closer to the truth than both Newton's theory and (non-quantum) versions of the wave theory.<sup>6</sup> I leave it as a challenge to raise similar questions about the caloric theory of heat.

### 3.3 Inference to the best explanation

So-called inference to the best explanation (IBE) plays an important role in the defense of standard realism. Ladyman [12, p.209], [13, p.341] describes this (putative) rule of inference as follows:

---

<sup>6</sup>Several more encompassing truth approximation questions are suggested by the set of success comparisons of electrodynamic theories that is provided by a table in [18, p.282](reprinted and commented upon in [6, p.118] and [7, p.236]). It represents the records in the face of 13 general experimental facts of the special theory of relativity (STR) and six alternative electrodynamic theories that were largely developed and defended in the 19<sup>th</sup> Century, viz., three versions of the ether theory and three emission theories.

“...where we have a range of competing hypotheses all of which are empirically adequate to the phenomena in some domain we should infer the truth of the hypothesis which gives us the best explanation of those phenomena.”

Here it is understood that one hypothesis provides a better explanation than another when either the first explains more of these phenomena than the second or when they explain the same phenomena but the first satisfies some additional non-empirical evidential criteria, or epistemic virtues, whereas the second does not satisfy all of them or to a lesser degree. Ladyman (p.340) lists as such non-empirical criteria simplicity, non-ad hocness, novel predictive power, elegance, and explanatory power, [3] mentions simplicity, elegance, inner beauty, fertility, coherence with background theories and/or metaphysical suppositions.<sup>7</sup> Following [14], I will call such non-empirical criteria *aesthetic criteria* when they are put forward by scientists for evaluative purposes. McAllister deals in particular with: symmetry, simplicity, and visualizability. [23] deals in his chap.6, entitled “Beautiful theories”, not only with simplicity and symmetry, but also with inevitability or rigidity.

According to the (standard) reading of IBE suggested by Ladyman, the competing hypotheses are not only assumed to be empirically adequate to the phenomena in the relevant domain in the sense that they are compatible with them, that is, they have not been falsified by one of them, but all of these hypotheses are even assumed to explain *all* phenomena in the domain. In this case, the best explanation is that one among those hypotheses that satisfies the non-empirical criteria the best.

IBE is of course not deductively valid, but it is used as a kind of inductive or abductive argument at two levels. At the local level it is used to escape in specific cases of empirically equivalent theories from the so-called ‘(strong) underdetermination by empirical data’. It is supposed to enable specific theoretical truth and reference claims. At the global level it is used to defend realism as the best explanation of the overall success of scientific theorizing. (According to the strong version of the no-miracles argument it is even the only explanation.) The global use can be defended against Van Fraassen’s Darwinian analogy [21, p.40] that, as a rule, the best theories survive because they are selected for that reason, by noting that realism predicts and explains the survival so far but also the continuation of success. Against the charge of circularity the global use has been defended by David Papineau and Stathis Psillos in a similar way as Braithwaite and Carnap defended the inductive defense of induction against the circularity charge, viz. by claiming that it is only ‘rule circular’, not ‘premise circular’ [12, p.218], [13, p.343].

However, IBE is debated in general as a rule of inference, where the following explication<sup>8</sup> is more or less presupposed, with a weak and a strong version.

<sup>7</sup>See [17] for a discussion of ten empirical and non-empirical virtues of theories.

<sup>8</sup>In earlier publications (e.g. [6, 9]) I have introduced a distinction between the ‘the best theory’ and the stronger notion of ‘the best explanation’, viz., being the best theory which has not yet been (empirically) falsified. Although it may sound a bit strange to talk about the best explanation even when it is known to be false, it makes the presentation much easier. Moreover, since I will now take non-empirical criteria into account for empirically equivalent theories it now meets the main criticism of [3, note 9, p.291] of my previous version of ‘inference to the best theory’.

**IBE-E1** If an explanation has so far proven to be the best one among the available empirically adequate theories, then conclude, for the time being, that

- it is true, that is, true as a hypothesis about the domain, or even
- it is the truth, that is, the strongest true hypothesis about the domain.

Elsewhere (e.g. [6]) I have put forward three general objections to this ‘implicit explication’:

- (1) It is restricted to unfalsified theories, for empirically adequate theories have by definition only true observational consequences. Hence there is no possibility of conclusions to progress by false, but less false, theories.
- (2) It is asymmetric in that it deals in the premises with the best of the available theories and in the conclusion with being ‘true/the truth’ simpliciter, that is, without reference to the available theories. In other words, the premises are comparative, whereas the conclusion is not.
- (3) It lacks justification of the conclusion
  - (3.1) when there is just one empirically the best theory available, let alone
  - (3.2) when there are empirically equivalent (EE-) theories among empirically the best ones (i.e. in the case of strong underdetermination).

Van Fraassen has raised three specific objections that may be seen as specifications of (3.2), hence dealing with EE-theories. They are known as:

- (i) The Argument from Indifference
- (ii) The Argument from the Best of a Bad Lot
- (iii) The Argument from Bayesianism

and will soon be paraphrased.

These objections suggest the following extra condition of adequacy for the comparative realist explication tasks extended with IBE:

**CA-IBE** An explication of IBE should also deal with the case that all available theories are false or even have been falsified (1), it should be symmetric between the formal nature of the premise(s) and the conclusion (2), and the (truth-related) conclusion should have some analytical justification and/or allow some empirical justification (3), when there is only one best theory available (3.1), and when there are EE-ones (3.2)

It is easy to check that the following revised explication of IBE satisfies at least CA-IBE1 and CA-IBE2.

**IBE-E2** If an explanation has so far proven to be empirically the best one among the available theories, then conclude, for the time being, that it is the closest to the truth among the available theories.

As announced before, Section 4 will indicate to what extent CA1 and CA2 can be satisfied and hence CA-IBE3.1, for that is essentially implied by them. However, IBE-E2 does not yet take EE-best theories into account, hence, it does not yet meet CA-IBE3.2.

There are some famous examples of EE-theories discussed by scientists themselves. Van Fraassen [21] has also discussed some rather artificial examples. The best-known and by no means artificial case is of course provided by the various so-called interpretations of quantum mechanics. The first one being the Copenhagen interpretation (put together principally by Niels Bohr) followed by the hidden variable interpretation of David Bohm. Several other interpretations are discussed today. Another case is provided by the existence of at least four EE-versions of classical mechanics, which will be discussed in Section 4.1.3. Hence, there is a serious problem of how to choose between EE-theories. IBE has been put forward not least for this problem of choice. Hence, let us look at Van Fraassen's objections, as formulated by Ladyman [13, pp.344-345]<sup>9</sup>, and evaluate them.

- (i) The Argument from Indifference: "...since there are many ontologically incompatible yet empirically equivalent theories, we have no reason to choose among them and identify one of them as true. ..."

This argument is to be rejected, for although there may be no compelling reasons, it disqualifies any role of non-empirical criteria as truth indicators.

- (ii) The Argument from the Best of a Bad Lot: "...we are to think that the collection of hypotheses that we have under consideration will include the true theory. The best explanatory hypothesis we have may just be the best of a bad lot, all of which are false. ..."

This is evidently a variant of our general asymmetry objection (2), but now specifically intended, though this remains here implicit, for EE-theories. However, also for EE-theories it is rather plausible. It is taken care of by CA-IBE2.

- (iii) The argument from Bayesianism: "...any rule for the updating of belief that goes beyond the rules of Bayesian conditionalisation [...] will lead to probabilistic incoherence."

This is also plausible and hence should be respected.

So let us reformulate CA-IBE3.2 in detail and in accordance with these evaluations, in particular, in contrast to (i).

**CA3** As far as EE-explanations are concerned, an explication of IBE should only take those prevailing aesthetic criteria into account, *if any*, for which there are empirical or analytical reasons, however weak, to assume that they are truth-conducive in the field.

Note for the moment already that the prevailing aesthetic criteria in a certain field at a certain time constitute what [14] has aptly called the aesthetic canon, arrived at by, and equally aptly called, *aesthetic induction*.

In accordance with (iii) we will require:

**CA4** In a probabilistic version of an adequate explication of IBE 'EE-updating' of the relevant belief should remain within the confines of Bayesian conditionalisation.

---

<sup>9</sup>It is to be noted that the formulation of the Argument from Indifference in [12, p.219] is a bit confusing, for it seems to coincide with the Argument from the Best of a Bad Lot (p.220). The Argument from Bayesianism is not dealt with in [12].

A merely probabilistic version will not do in general, for probabilistic updating makes only differential sense for unfalsified theories. However, probabilistic estimation of the distance from the truth [15, 16], using Bayesian updating, is generally applicable.

Of course, IBE-E2 trivially satisfies CA3 and CA4 when there are no EE-best theories. For the case that there are we extend it with an extra clause

### **IBE-E3**

- (3.1) (=IBE-E2). If an explanation has so far proven to be empirically the best one among the available theories, then conclude, for the time being, that it is the closest to the truth among the available theories.
- (3.2) If some empirically equivalent explanations have so far proven to be empirically the best ones among the available theories, and if among them there is one that is according to the relevant aesthetic canon the best one, then conclude, for the time being, that it is the closest to the truth among the available theories.

It is easy to check that (3.2) satisfies, like (3.1), CA-IBE1 and CA-IBE2. Section 3 will indicate to what extent it also satisfies CA-IBE3.2, that is, CA3 & CA4.

This concludes our survey of the main antirealist arguments and which conditions of adequacy comparative realist explications of the crucial notions have to satisfy in order to meet them. The resulting conditions of adequacy to be satisfied are on the one hand CA1 and CA2 and on the other hand CA3 and CA4, for CA-IBE was argued to be taken care of by IBE-E3 (meeting CA-IBE1 and CA-IBE2) and the resulting four conditions (meeting CA-IBE3).

## **4 Evaluation of comparative realism**

In the previous section we have derived four desiderata for an adequate explication of a comparative realist stance that is claimed to be better defensible than other responses to the discussed antirealist charges. In this section the first task is the separate evaluation of comparative realism, that is, to show to what extent realism guided by the comparative perspective on success and on truth approximation, fulfils these desiderata. The second task is its comparative evaluation, that is, to what extent is it superior to the other responses to antirealism.

### **4.1 In defense of the main claim of comparative realism**

Recall that we have summarized the way in which comparative realism is claimed to fulfill the desiderata a couple of times by its ‘main claim’: truth approximation provides the default explanation and prediction of empirical and aesthetic progress. This claim is supposed to be realized by the stratified (basic) explication of the notions of ‘more successful’ and ‘closer to the truth’, and the resulting mutual relations. Let us start with noting a number of terminological equivalencies that will be presupposed in the defense of the claim.

Empirical progress combines instantial progress and explanatory progress, and a new theory amounts to empirical progress relative to its predecessor when it is concluded on the basis of ‘sufficient’ comparative testing that the



new theory is persistently empirically more successful than the old one. Empirical success preservation provides weak versions of these kinds of progress.

Aesthetic progress of a new theory relative to an old one, according to the prevailing aesthetic canon and in particular among empirically equivalent (EE-)theories amounts to an increase of having desirable-aesthetic properties and a decrease of having undesirable ones. Below I will indicate that empirical progress and aesthetic progress have in some formal sense the same structure. Aesthetic success preservation is again a weak version.

In the expression ‘closer to the truth’, ‘the truth’ always refers to the strongest true theory about a given domain in a given vocabulary and ‘truth approximation’ (TA) is short for ‘getting closer to the truth’. In the stratified context we get the equivalences: getting closer to the theoretical/ observational/ referential truth = theoretical/ observational/ referential truth approximation. In this context, ‘truth approximation’ and TA without qualification are supposed to refer to ‘theoretical truth approximation’. The weak version of ‘theoretical truth approximation’ is called ‘theoretical truth preservation’. Similarly, the weak version of ‘referential truth approximation’ is ‘referential truth preservation’.<sup>10</sup>

We will first deal with CA1 and CA2, in which EE-theories and aesthetic progress are not yet at stake, and later with CA3 and CA4. In both cases the claim is of course that the crucial notions can be explicated such that (it is reasonable to assume that) these conditions of adequacy are satisfied. In the first case the justification is primarily analytical, allowing empirical illustrations. In the second case the justification will be a combination of empirical and analytical arguments. The analytical justification will be based on a number of ‘TA-theorems and -conjectures’. The empirical justification will be twofold. Moreover, I will first restrict the attention to the cautious, very demanding, explications of the crucial notions, that is, the stratified (basic) explications.

#### 4.1.1 Empirical progress

Let us first focus on matters of empirical progress in relation to theoretical truth approximation, and address matters of referential truth preservation later. For convenience, I repeat the first two conditions of adequacy restricted in this way:

**CA1-EP** The explications of the crucial phrases should leave room for the possibility that false theories are not only (persistently) empirically more successful but even closer to the (theoretical) truth.

**CA2-EP** The explications should be such that empirical progress is, *as a rule*, due to theoretical truth approximation, which on its turn entails, as a rule, novel predictive success.

In Section 2 I have already specified the stratified basic definitions of ‘empirically more successful’, hence of ‘empirical progress’, and of ‘closer to the theoretical truth’. From these definitions it is easy to see that CA1-EP is satisfied, for the definitions do not make a crucial difference between theories that are true or false, in the (weak) sense, that is, theories that don’t exclude nomic theoretical possibilities and theories that do, respectively.

<sup>10</sup>Note that we do not call ‘theoretical, referential, observational truth approximation’ progress of that kind, for we like to reserve the term ‘progress’ for kinds of progress that do not presuppose to know the relevant truth, such as empirical and aesthetic progress.

Regarding CA2-EP, recall, the (empirical) Success Theorem, which can be formulated for the stratified case as follows:

**Stratified Empirical Success Theorem**

Theoretical truth *preservation* of one theory relative to another entails explanatory success preservation and, as a rule, instantial success preservation. Exceptions to instantial success preservation concern instantial pseudo successes<sup>11</sup>, to be explained, of the old theory, dropped by the new theory.

Theoretical truth *approximation* entails in addition, assuming empirically non-equivalent theories, that at least some extra (empirical) success can be obtained, and hence the prospect of empirical progress including novel facts.

The proof of the explanatory part in the first clause is straightforward, as in the unstratified case. However, the instantial part is now not unconditional, for in the stratified situation there may be ‘instancial pseudo successes’ of  $X$  relative to  $Y$ . Such a success amounts to an observational nomic possibility that is allowed by  $X$  and not by  $Y$  (hence breaking observational truth approximation of  $Y$  relative to  $X$ ), which, however, on the theoretical level must be due to the allowance by  $X$  of a (theoretical) nomic *impossibility*, which is not allowed by  $Y$ . Of course, any instantial success of  $X$  or  $Y$  may be a pseudo success, which is something which we cannot observationally distinguish from a genuine instantial success. However, this is only obscuring comparative theory evaluation as far as *extra* instantial successes are concerned. By the way, the epistemic asymmetry between explanatory and instantial successes may well explain why explanatory successes of proper theories (theories with theoretical terms) seem more impressive than counterexamples.

From this theorem we may conclude that theoretical truth approximation provides the realist default explanation and prediction of empirical progress. For, assuming at least some initial extra explanatory success, the hypothesis of theoretical truth approximation, the TA-hypothesis, enables the explanation and prediction of explanatory progress, including some (differential) novel predictive success, as we will see. Moreover, the hypothesis enables the conditional explanation and prediction of instantial progress, including some (differential) novel predictive successes, as we also will see. The condition being that the old theory does not have or get pseudo successes. If the TA-hypothesis is in fact false, new experiments will break the empirical progress conclusion in the long run. However, as far as breaking the instantial progress conclusion is concerned this may be due to pseudo successes.

In some more detail, the situation is as follows, assuming empirical progress, that is, a well confirmed comparative success hypothesis (CSH) of some theory  $Y$ , relative to  $X$ , on the basis of evidence  $R/S$ . If  $Y$  is in fact not closer to the theoretical truth than  $X$ , then there is some theoretical nomic possibility, say  $x$ , excluded by  $Y$  but included by  $X$  and/or some theoretical nomic impossibility, say  $y$ , allowed by  $Y$ , but not by  $X$ . In the first case there is room for an experimental counterexample of  $Y$  which is not a counterexample of  $X$ , viz. when the projection of  $x$  is also excluded by  $Y$ , and hence is no pseudo success of  $Y$ . In

<sup>11</sup>Called ‘extra lucky observational hits’ in previous work.

the second case an observational law, according to which at least  $y$  is impossible, can be experimentally established that is an explanatory success of  $X$ , but not of  $Y$ , provided the projection of  $y$  is also excluded by (the projection of)  $X$ , and hence no pseudo success of  $X$ .

Regarding the claim of some (differential) novel predictive success the situation is as follows. It is rather difficult to imagine that, at the start of testing a CSH, see Section 2, where the success superiority of  $Y$  relative to  $X$  is based on previous findings (R/S), these findings exhaust every kind of thing there is to know of the observable differences between  $X$  and  $Y$ . In other words, except in finite cases, and perhaps some very simple infinite cases, at least some room for differential ‘novel predictive success’ is plausible, in whatever sense ‘novel’ precisely is meant.

The foregoing not only shows that the TA-hypothesis provides the default explanation and prediction of empirical progress when the crucial notions are explicated in the stratified basic way, but also that CA2-EP then is satisfied.

#### 4.1.2 Referential truth preservation

The first two conditions of adequacy also deal with reference claims, specifically:

**CA1-Ref** The explications of the crucial phrases should leave room for the possibility that false theories, even with some non-referring theoretical terms, are closer to the (theoretical) truth and refer at least as well as other theories.

**CA2-Ref** The explications should be such that theoretical truth approximation entails, as a rule, referential truth preservation.

In Section 2 I have already indicated the definition of ‘reference’ and of ‘closer to the referential truth’. Here I will elaborate these indications a bit further, partly because of some weak spots in earlier formulations, but the main ideas remain the same.

In the definition of a subset  $X$  of  $M_p$ , constituting theory  $X$ , with strong claim “ $X = T_t$ ”, a theoretical term may or may not play a substantial role in the sense that if we project  $X$  on the conceptual possibilities generated by the  $M_p$ -vocabulary minus that term, and then recombine these possibilities again with that term in all possible ways, the resulting subset of  $M_p$  may or may not be a proper superset of  $X$ . If it is,  $X$  is said to claim that the term does refer, for it makes a difference. If it is not,  $X$  is said not to refer, for it does not make a difference.

The definition still needs some qualification. It seems adequate primarily for theoretical terms of which  $X$  claims that they refer to attributes (properties, relations, and functions); attribute terms for short. For theoretical terms claimed by  $X$  to refer to classes of entities, used as domain-sets for attribute terms - entity terms for short - this definition does not seem to work. However, it is plausible to say that a (theoretical) entity term refers according to  $X$  if and only if there is at least one (theoretical) attribute term that is claimed to refer and that uses the entity term as (one of) its domain-set(s). The consequence is that an entity term does not refer according to  $X$  if there are no attribute terms using it as domain-set. This is plausible, for in this case it is difficult to

see how that theoretical domain could play a substantial role in the definition of  $X$ , which is precisely the reason for the detour via attribute terms. If this detour is not possible, the theoretical entities hang in the air according to  $X$  as unconstrained entities, not distinguishable from genuine fictitious entities, at least not with the means provided by the vocabulary. For example, the Higgs-particles, postulated for theoretical reasons by the so-called Standard Model in elementary particle physics, are supposed to have theoretical properties with observational consequences. One expects to observe the latter in the upcoming experiments in 2008 with the Large Hadron Collider (LHC) at CERN, near Geneva.

Of course, assuming that  $T_t$  can be defined with the means of  $M_p$ ,  $T_t$  also claims, according to this definition of reference claims, of certain theoretical terms that they refer and of others that they do not, but now always rightly of course. Note that this definition of (non-)reference amounts to a direct reference criterion for attribute terms and an indirect one for entity terms. Note in passing that the very possibility to define reference of entity terms on the basis of the definition of reference of attribute terms is a good reason to extend the idea of 'entity realism' to 'referential realism'. Note finally that the given definition of reference is explicitly  $T_t$ -relative.

According to the *Nomic Postulate*, applied to  $M_{pp}$  as well as  $M_p$ , we have assumed that unique fixed sets  $T_o$  and  $T_t$  exist, given an intended domain. Moreover, we have implicitly assumed that both can be characterized in some way by the available vocabularies. Regarding  $T_t$ , however, the Nomic Postulate is somewhat dubious, for it seems to presuppose that the reference of the theoretical terms is given beforehand. More cautiously than we have done before, we have to assume that, if we would have some kind of superhuman observational access to  $M_{pp}$ -relevant theoretical entities and attributes it is possible to couple none, some or all theoretical terms constituting  $M_p$  to them such that  $T_t$  arises. Of course, it is not at all evident that this can be done only in one way. But we may assume that the number of terms that can be meaningfully coupled at once has a maximum and that this maximal coupling is unique and gives rise to  $T_t$ . In other words, we assume that there is unique maximal referential use of the theoretical vocabulary and that this is which we are looking for. The conjunction of the (non-)reference claims of  $T_t$  is called 'the referential truth' and the conjunction of all (non-)reference claims of theory  $X$  is called its (total) referential claim. Finally, theory  $Y$  is closer to the referential truth than theory  $X$  if  $Y$ 's referential claim agrees on more terms with the referential truth than that of  $X$ . Referential truth preservation is defined as: theory  $Y$  is at least as close to the referential truth as theory  $X$  if  $Y$ 's referential claim agrees on at least the same terms with the referential truth as that of  $X$ .

Let us call two theories referentially equivalent when they have the same (total) referential claim, and let us call them referentially semi-equivalent when they have the same number of positive reference claims. The uniqueness assumption about  $T_t$  now amounts to the assumption that a referentially semi-equivalent theory to  $T_t$  is referentially equivalent to  $T_t$ . Of course, referential (semi-) equivalence does not imply nor exclude empirical equivalence.

From the above it is clear that being (in the weak sense) true or false, even with some non-referring theoretical terms, does not play some crucial role in the definitions of theoretical truth approximation and referential truth preservation. Hence, CA1-Ref is satisfied. Regarding CA2-Ref, we submit the following

(diagnostic) conjecture:

**Referential Truth Conjecture:**

Theoretical truth approximation entails, as a rule, referential truth preservation. If not, the old theory has *extra lucky* reference claims.

Note first that we do not add a second clause to the conjecture, similar to that of the Stratified Empirical Success Theorem, which would now amount to: theoretical truth *approximation* entails in addition (to theoretical truth preservation), as a rule, referential truth approximation. In the present case, it is difficult to argue for it, and plausibly so, because this would (almost) exclude the possibility of improving theories without improving its reference.

Let me explain what lucky reference claims, mentioned in the conjecture, are. One might think that the assumption that  $Y$  is closer to  $T_t$  than  $X$  does entail the claim that  $Y$  is referentially at least as close to the truth as  $X$ . A specific reason to expect this is the fact that the theoretical claim of a theory, e.g. " $X = T_t$ ", entails its referential claim " $V_r(X) = V_r(T_t) = V_r$ ", where  $V_r(X)$  indicates the set of theoretical terms that refer according to  $X$ , and  $V_r$  is the one corresponding to  $T_t$ . However, a proof for the suggested unconditional conjecture is not possible, for interesting reasons. Suppose that  $Y$  is closer to  $T_t$  than  $X$ . What we would like to prove amounts to the following: all  $\tau$  in  $V_r(Y) - V_r(X)$  belong to  $V_r$  and all  $\tau$  in  $V_r(X) - V_r(Y)$  do not belong to  $V_r$ . Let us suppose that  $\tau$  in  $V_r(Y) - V_r(X)$  does not belong to  $V_r$ . Hence,  $Y$  wrongly claims that it does, whereas  $X$  rightly claims that it does not. However,  $X$ 's claim may be based on a wrong aspect of its theoretical claim " $X = T_t$ ", whereas  $Y$  may not yet be so good that its theoretical claim implies the reference claim for the right reasons. It is important to note that the suggested proof already fails in the case of just one theoretical term. Similar arguments can be given for the possibility that  $\tau$  in  $V_r(X) - V_r(Y)$  belongs to  $V_r$ . In both cases we might say that  $X$  has lucky reference claims relative to  $Y$ , that is,  $X$  has true reference claims, whether or not precisely for the right reasons, where  $Y$  has false reference claims.

In sum, we have to admit the possibility that  $X$  has such lucky (non-) reference claims relative to  $Y$ . This does not mean, of course, that 'to be as close to the theoretical truth' does not provide good arguments for believing 'to be at least as close to the referential truth'. As can be learned from the suggested proof attempt, the suggested entailment is only violated when a theory bases (part of) its referential claim on wrong reasons in precisely the right direction, which will be rather exceptional because it is rather artificial.

So far, for the analytical justification of the claim that the stratified (basic) definition of truth approximation satisfies to a considerable extent the conditions of adequacy regarding empirical progress and referential truth preservation, viz. CA1 and CA2. After dealing with aesthetic progress, I will come back on the possibility of empirical support of this claim.

#### 4.1.3 Aesthetic progress

For the moment we will focus on CA3, for CA4, dealing with probabilistic versions of IBE, evidently exceeds the boundaries of the qualitative approach.

**CA3** As far as EE-explanations are concerned, an explication of IBE should only take those prevailing aesthetic criteria into account, *if any*, for which there

are empirical or analytical reasons, however weak, to assume that they are truth-conducive in the field.

As we already noted, the prevailing aesthetic criteria in a certain field at a certain time constitute what James McAllister [14] has called the aesthetic canon, arrived at by aesthetic induction. In contrast to observational criteria, where we may assume, and have argued in some detail in the first subsection, that they are, as a rule, truth-conducive in the comparative sense, empirical information will have to support that certain aesthetic criteria are truth-conducive. However, before we enter the empirical justification, we will first show that there may well be truth-conducive non-empirical criteria at all. We will show this by a theorem that is a general version of (the explanatory part of) the success theorem. Of course, ‘the truth’ has not only observational consequences, but also theoretical (and mixed) consequences. A (purely) theoretical consequence of  $T$  amounts to a superset  $C$  of  $T$ , of which the projection equals  $M_{pp}$ , it does not exclude something observational. Note that every consequence of  $T$  can also be interpreted as a property or feature of all nomic possibilities, to be called a nomic property<sup>12</sup>.

*Nomic Theorem*

Theoretical truth preservation (whether or not among EE-theories) entails being at least as successful regarding nomic properties.

A generally interesting question is which non-empirical properties are nomic, hence truth-conducive, properties. However, since observational differences will have priority in truth approximation evaluations, the use of this theorem naturally focuses on EE-theories. Hence, the main question becomes which non-empirical properties may be or have been supposed to be truth-conducive in the indicated sense.

We have already suggested, following McAllister and, for example, Weinberg, calling the non-empirical properties of interest aesthetic criteria, for such properties are as a rule also aesthetically appreciated. The empirical justification of claims to the effect that certain aesthetic criteria are truth-conducive in the field may be given along two different lines, and their combination. In the first line, due to Igor Douven, the claim is tested with hindsight, here called ‘aesthetic testing’, in the second line, the claim is based on ‘aesthetic induction’ as introduced by McAllister.

Douven [1–3] has described a way to (bootstrap) test applications of IBE on EE-theories with hindsight, which amounts briefly to the following. Although at the time of the application of IBE the theories were empirically equivalent, this may change in the course of time. New developments, e.g. the development of a sophisticated type of microscopes or the building of higher energy particle colliders, may lead to “a shift in the boundary between what can and cannot be observed” [3, p.297], which he aptly called an OUD-shift in his [1], where OUD is short for ‘Observable-Unobservable Distinction’. In our approach this amounts to enriching  $M_{pp}$  to some  $M'_{pp}$ . Assuming that the new microscope or collider is veridical, which is the bootstrap condition of the testing, the interesting question

<sup>12</sup>Similarly, we could introduce ‘strongly non-nomic’ properties, properties corresponding to consequences of the negation of  $T$ , i.e. supersets of the complement of  $T$ , enabling a kind of mirror theorem to the nomic theorem. See [7, chap.10] for a general exposition of truth approximation in terms of two kinds of features.

becomes whether the previously best EE-theory, based on certain non-empirical properties, turns out to be empirically the most successful theory according to the new observational evidence acquired with the new means. If so, this confirms the claim that the (still) non-empirical properties were in fact truth-conducive.

The second line of empirical justification is provided by the ‘naturalistic-cum-formal’ analysis of the relation between aesthetic criteria, empirical success and truth [8]. It starts with the notion of ‘aesthetic induction’. As McAllister has shown in his very inspiring book *Beauty and Revolution in Science* [14], our aesthetic judgments are subject to change. We are not only inclined to find the heliocentric worldview of Copernicus more beautiful, because it is simpler, than the geocentric view of Ptolemy, but we are also inclined to find Kepler’s elliptic planetary orbits at least as beautiful as Copernicus’ circular orbits. However, ellipses are undoubtedly more complicated than circles, and this is precisely the reason why they were found less beautiful, if not problematically ugly, at the time. Moreover, I would like to add, aesthetic criteria not only change with time within a discipline, but may also differ greatly between disciplines. For example, after expressing in an interview series [5] his agreement with Steven Weinberg [23] about the importance of beauty considerations, Stephen Gould hastens to stress that his criteria for beauty totally differ from those of Weinberg. Whereas Weinberg mentions ‘inevitability’ of desired consequences as his dominant aesthetic criterion – as exemplified by Einstein’s theory, which, in contrast to Newton’s theory, made the inverse square in the law of gravitation inevitable – Gould stresses that, besides diversity, unrepeatable contingencies and irregularities are the sources of his ultimate aesthetic satisfaction.<sup>13</sup>

Having observed the variation of aesthetic criteria, McAllister’s main claims are as follows. First, scientists normally use aesthetic criteria in addition to empirical criteria for theory evaluation. Second, and most importantly, the aesthetic criteria of the time, the ‘aesthetic canon’, is based on ‘aesthetic induction’ regarding non-empirical features of paradigms of empirically successful theories which scientists have come to appreciate as beautiful. Third, aesthetic criteria can play a crucial, schismatic role in scientific revolutions. Since they may well be wrong, they may, in the hands of aesthetic conservatives, retard empirical progress, but this does not occur in the hands of aesthetically flexible, ‘revolutionary’ scientists.

My earlier analysis [8] of the relation between the empirical success and the beauty of scientific theories elaborates and supports McAllister’s claims. Like McAllister, I focus on non-empirical aesthetic features, that is, features with aesthetic value but without empirical content, although empirical features may also be aesthetically valued. In the spirit of naturalized epistemology, one may first argue that the phenomenon of aesthetic induction may be a variant of the so-called ‘mere-exposure effect’ [24], viz. non-conscious affective priming [25]. In this perspective one may decompose the notion of aesthetic induction into aesthetic induction proper, or affective induction, and a related cognitive (meta-) induction. Together they lead to correlations between non-empirical features that are found beautiful on the one hand and empirically successful theories on the other. Such correlations may be called ‘beauty-success correlations’. The corresponding received or ‘canonical’ aesthetic features are non-empirical features that have ac-

---

<sup>13</sup>This is not to suggest that standard examples of aesthetic features mentioned by physicists do not play a role in biology. For example, Gould mentions order a number of times and Sober [20] points out that simplicity in the form of parsimony plays a considerable role in taxonomy.

quired (positive) aesthetic value and (empirical success related) meta-inductive support. This makes it plausible to explicate the notion of an ‘aesthetic feature’ more precisely as an aesthetically (positively) valued non-empirical feature. Now it is possible to argue that the co-production of the two types of induction is functional for empirical progress as far as the cognitive meta-induction is reliable. That is, aesthetic features may rightly have become indicative of empirical success.

In view of the nomic theorem and our formal analysis of the relation between (more) empirical success and truth approximation, indicated in the first subsection, such features become indirectly indicative of truth approximation, that is, truth-conducive, provided they are of a certain formal nature. Since aesthetic features correlate with empirical success and empirical success with truth approximation, it has some plausibility to submit that such features of EE-theories are nomic properties, in which case they strongly correlate with truth approximation. However, this formal account only holds for aesthetic features of a certain formal kind, to be called distributed features. In order to possibly be a nomic property, such a feature should apply to all models of the relevant theory, in which case they can also be represented as a consequence of the theory. To be sure, the link between distributed aesthetic features and truth approximation is weak, but it makes some sense. It essentially consists in the combination of the stratified empirical success theorem and the nomic theorem. Below we will briefly discuss which aesthetic features are or may be distributed.

Of course, aesthetic induction will cover (positive) results of previous aesthetic testing, whether or not such testing has been undertaken explicitly for this purpose. Hence, like aesthetic induction, aesthetic testing contributes to the relevant ‘aesthetic canon’. This strengthens the conclusion that there are good, though not strong, reasons to assume that the aesthetic canon, as far as consisting of distributed features, may rightly hypothesize what are truth-conducive theoretical nomic properties, and hence for the following conjecture:

*Aesthetic Success Conjecture*

Theoretical truth *preservation* among EE-theories entails preservation of theoretical nomic properties, and hence, in view of the nature of aesthetic testing and aesthetic induction, as a *weak* rule, aesthetic success preservation according to the aesthetic canon as far as distributed. If not, which may appear by new aesthetic testing or a major breaking of aesthetic induction, previous aesthetic testing and induction prove to have their limits.

Theoretical truth *approximation* among EE-theories entails increase of theoretical nomic properties, and hence, as an at least as weak rule, aesthetic progress according to the aesthetic canon as far as distributed.

The important remaining question is which aesthetic features are, or can have a distributed form. Let us review the features that have been mentioned earlier, now grouped in five categories. Each of them may have other examples.

- inner beauty, elegance,
- simplicity, symmetry, visualizability, inevitability or rigidity, coherence with metaphysical suppositions,



- explanatory power, novel predictive power, fertility,
- coherence with background theories,
- non-ad hocness, parsimony.

I will deal with these categories in the listed ordering. The *first* category lists features that ask for some kind of objective specification, such as those of the other categories. The *second* category typically represents features that are non-empirical and admit an objective specification. Moreover, although they can apply to theories as wholes they can also apply, or not, to separate conceptual possibilities, and hence to all models of a theory, or not, that is, they can distributively apply to a theory, or not. Finally, their precise nature and their aesthetic value may typically vary historically and between disciplines. In sum, they typically form the kind of features to which the aesthetic success conjecture is oriented. The *third* category typically contains features that first of all promise to be accounted for in empirical success: ‘explanatory power’ in explanatory success, ‘novel predictive power’ in novel predictive success, instancial or explanatory, and finally ‘fertility’ frequently is used to express the two previous kinds of power. In all three cases, it may well be that the aesthetic value they have, derives from these promises of empirical virtues. Although such promises may turn out to be realizable, it is important to note that such empirical pay-offs can only arise when there are such promises. All this holds, whether or not the features may apply in a distributed form, which seems to be possible for ‘novel predictive power’, and hence for a certain kind of fertility. The *fourth* category is a kind of mirror category of the previous one. Incoherence with background theories is a ‘promise’ of empirical problems, the stronger the more established the background theories are. Hence, this type of incoherence may well have become negatively valued aesthetically as a predictor of empirical problems. However, it may also be a (very weak) predictor of a scientific revolution. Finally, ‘non-adhocness’ in the *fifth* category, is generally valued positively because adhocness is disvalued, probably because of similar reasons as the previous type of incoherence. The saving of a theory by an ad hoc repair frequently turns out to have its empirical price. However, in this case it is important to note that it may also turn out to be a smooth kind of truth approximation. This applies *mutatis mutandis* to parsimony, for again non-parsimoniousness may show to have its empirical, notably explanatory, price, but it may also be a step in the good direction, the truth.

Let us finally confront the proposed way of dealing with EE-theories in accordance with CA3 with the provoking paper of Roger Jones [4], entitled ‘Realism about what?’ (see also, [12, pp.252-257], from which I will borrow crucial formulations). In this paper, Jones reminds us of four alternative formulations of classical mechanics, as for example applied to planetary motion. Though empirically equivalent, they suggest nevertheless a quite different ontology and metaphysics. Newton’s own approach, based on the three laws of motion and force laws, suggests an ontology of point particles and forces acting at a distance. The gravitational field approach assumes an ontology of action by contact in conformity with a principle of local causality, but postulates a new type of entity, the field. In the curved space approach, the gravitational field is absorbed into the structure of space (analogous to the way in which this field is absorbed into the structure of spacetime in general relativity). In this approach the field is

represented by the local degree of curvature and space itself is made part of the fundamental ontology having a causal efficacy of its own. Finally, the analytical mechanics approach, in which the force laws and the laws of motion can be derived from so-called minimum principles, seems to be independent of causal thinking, but requires instead a kind of teleology for particle behavior according to which it is the properties of complete paths between points of space that determine the motion of a body.

The question is of course which ontology (cum metaphysics) is suggested by a realist attitude to these EE-versions of classical mechanics. The suggestion of our analysis is that this depends on which formulations score the best according to the prevailing relevant aesthetic canon, dealing with ontological, epistemological, methodological and pragmatic aspects. However, apart from the difficult question as to which canon is prevailing and relevant, given so long a period of time in developing these formulations, the preliminary question is whether the suggested ontologies are really as different as they seem at first sight. The first approach (with forces acting at a distance) and the field approach may well be seen as compatible ontologies, provided the forces are not assumed to act at a distance instantaneously. With this (non-Newtonian) assumption they are related as the manifest global or macro-ontology to the theoretical local or micro-ontology as seems, for example, the standard view in response to Eddington's non-existence claim of his famous table. The relation between the field approach and the (non-relativistic) curved space approach does not seem to be a matter of compatible ontologies. However, looking back from general relativity theory, and its empirical success, the aesthetic appreciation of a curved spacetime and hence presumably of a curved space (both of which can be distributed properties of a theory), seem to have become (much) higher than that of the (relevant) field approach. Hence, the curved space version of the classical theory of gravitation seems to deliver the most plausible realist ontology, and hence would seem so far to be the best theory in accordance with CA3, and hence for inferring, by IBE-E3.2, that it is the closest to the theoretical as well as to the referential truth. This judgment will not be changed by the last, minimum principles formulation of mechanics, for that delivers from a realist point of view at most an as if metaphysics. However, as such it does no theoretical work, let alone observational work, for in the present deterministic setting this kind of 'as if metaphysics' is purely a matter of logico-mathematical equivalences.

This concludes our articulation and defense of the claim that CA3 can be satisfied, at least to some extent.

As we stated already, CA1 and CA2 and our main claim have primarily analytical justification. Of course, this does not exclude empirical support. On the contrary, their justification asks for real life illustrations. In particular for CA2 one may think of the kind of (implicit or explicit) testing proposed by Douven on the basis of OUD-shifts: do previous 'empirical progress' and 'referential truth preservation' conclusions survive after the shift or even, when *all* theoretical terms have become observable, can we now conclude to 'referential truth approximation'?<sup>14</sup> Unfortunately, the demands for stratified basic truth approximation are very high, so that we will have to extend the analysis to 'refined qualitative'

---

<sup>14</sup>Of course, OUD-shifts will also lead to reconsideration of previous aesthetic inductions: do they survive or has the canon (to be) changed. According to McAllister, a change of the canon is a characteristic for scientific revolutions.

or ‘quantitative versions’. This brings us to CA4.

Regarding CA4, the situation is not difficult. Recall:

**CA4** In a probabilistic version of an adequate explication of IBE ‘EE-updating’ of the relevant belief should remain within the confines of Bayesian conditionalisation.

The strictures of the ‘argument from Bayesianism’, that is, resulting from non-Bayesian conditionalization, can be avoided in at least one way, viz. by taking aesthetic features only into account in the (expert) prior distribution. As far as I know there is no other way<sup>15</sup>. As I already remarked in Section 3.3, a merely probabilistic version will not do in general, for probabilistic updating makes only differential sense for unfalsified theories. However, if we leave behind the strict conditions of stratified basic truth approximation, probabilistic estimation of the distance from the truth [15, 16], using Bayesian updating, is generally applicable.

## 4.2 Review and extension of the comparison with other realist responses to antirealism

Some preliminary remarks are in order before we can take up the comparison of the comparative approach with other realist responses. First, there are, of course, no *compelling* reasons neither for comparative realism nor for antirealism, whether the latter concerns theoretical, inductive or even experiential skepticism. However, Van Fraassen’s adagium “what is rational to believe includes anything that one is not rationally compelled to disbelieve” [22, pp.171-172] is a license for all kinds of wishful thinking, blocking a ‘good reasons’ debate. Although there is not much hope to convert Van Fraassen, a paradigm believer in antirealism, more specifically, constructive empiricism, into a believer in comparative realism, I have some hope that analytical philosophers with realist inclinations but doubts based on the antirealist charges, may become comparative realists instead of remaining adherent of one of the other retreats from full blown realism.

Let me first review the realist responses already dealt with so far. The retreat to claims of ‘approximately true theories’ instead of ‘true theories’ was put aside at the start as of no help, for that remains basically non-comparative. Although we certainly prefer giant leaps to the truth, any step in the direction of the truth, from whatever to whatever distance to the truth should be welcomed by the realist as a kind of progress. Moreover, the retreat to ‘approximate truth’ requires a necessarily arbitrary threshold. Further we came across restricting realism to ‘mature’ theories, or to theories with ‘novel’ predictive success or to the ‘essential’ parts and aspects of theories. The general tenet of our criticism of these moves was a kind of redundancy. From the comparative perspective all theories may remain in the game, no ban of certain kinds or aspects of theories are required. This is not to say that some of the suggested distinctions may not be useful for other purposes, related or not. The same holds for the one opposite kind of response we have met, viz. stretching realism by stretching the (causal) theory of reference underlying many conceptions of realism: it is not necessary

<sup>15</sup>However, Niiniluoto reminded me of the fact that Hintikka’s  $\alpha$ - $\lambda$ -system of inductive probabilities converges to the simplest generalization that is still compatible with the evidence. Hence, it need not be necessary to have to take explicitly care of aesthetic considerations in order to be respected.

for defending realism. As we have seen, such a defense is possible by using a non-standard criterion of reference, which may well be compatible with the causal theory of reference, or some stretched refined version.

Let me now turn to the extension of the comparison, in particular, with other realist responses by restricting realism. To begin with, ‘entity realism’, that is, realism with respect to the referential claim attached to theoretical entity terms, has been motivated, at least partially, by the idealizational, hence truth-abusing, nature of scientific practice, in particular model construction. From the comparative perspective this motive for restricting realism in order to save the important idealization practice is not relevant, for the question is whether of two successive theories, the second, being a concretization of the first, idealized theory, is closer to the truth than the first. In [6, chap.10], I have shown and illustrated that the refined version of the truth approximation theory leaves perfect room for a positive answer to this question, roughly due to the fact that the second theory has models that are more similar to correct models than the models of the first. Of course, the same argument can be applied to the less restrictive version of entity realism, called ‘referential realism’, that is, realism with respect to the referential claim of all theoretical terms, entity as well as attribute terms. Recall that I have remarked in passing that the transition from entity to referential realism is particularly plausible in view of the non-standard criterion of reference.

Finally, another retreat, which has become much debated in the last decade, is so-called structural realism, according to which typical realist features of successful theories consist of their structural features as expressed by the formalism specifying the formal relations. This retreat is mainly motivated by frequent referential failures, as highlighted by the pessimistic meta-induction. Again, it is an unnecessary retreat from the comparative perspective, now in view of the possibility of referential truth approximation.

## 5 Concluding remarks

I would like to conclude this paper with a general remark and a survey of epistemological positions.

All realist responses to the antirealist charges are retreats of realism of a non-comparative nature. As already suggested by my puzzling question in the introduction, it seems that all responses are non-comparative because of the lack of awareness that in the present debate relations, in particular, comparative ones, may be crucial and should be taken into account. Of course, comparative realism is also a kind of retreat. However, since it is a kind of piecemeal realism, it fits perfectly well into what may be seen as the core idea of realism, namely that our theories succeed more and more in giving ‘a literally true story of what the world is like’, to use Van Fraassen’s well-known characterization of realism [21, p.8].

In my *From Instrumentalism to Constructive Realism* [6], I have presented a hierarchical survey of epistemological positions. In the course of time I have refined and revised it at some places, but the core is still the same. I would like to present here the latest version, for it captures the main lines of thought in this paper rather well. It is structured by a number of epistemological questions, however, starting with an ontological one (as Tab.I). A non-standard kind of summary of the main lines in this paper, some of it implicit, is as follows,

Q0:	independent natural world? ↓ * Yes: ontological realism	No ⇒	ontological idealism
Q1:	true claims about it possible? ↓ * Yes: epistemological realism	No ⇒	epistemological skepticism - experiential skepticism - inductive skepticism
Q2:	beyond the observable? ↓ * Yes: scientific realism	No ⇒	theoretical skepticism instrumentalism - constructive empiricism
Q3.1	beyond reference?	No ⇒	referential realism ⇒ entity realism
Q3.2	beyond structures? ↓ * 2 × Yes: theory realism	No ⇒	structural realism
Q4:	ideal conceptualization? ↓ Yes: essentialistic realism	No ⇒ *	constructive realism

Table I: Hierarchy of epistemological positions

in the scheme indicated by the starred arrows. There is a human-independent natural world, about which true claims may be (non-compellingly) justified, not merely restricted to what is observable, but also with respect to theoretical terms and statements, provided we take the comparative piecemeal perspective, being a core feature of comparative ‘theory realism’. However, as to the question whether there is some ideal vocabulary fitting the natural world, and hence leading to a kind of realism, my answer is negative, for which reason I speak of ‘constructive’ realism, e.g. in the title of my book. Finally, I would like to stress that constructive realism is a kind of nomic realism. As became clear in the present paper, according to my view, theory formation and revision is not directed at truth approximation with respect to the actual world but at the nomic world, that is, the realm of what is physically possible.

## Acknowledgements

I like to thank a number of people for various useful comments received at the try-out in Groningen and in Beijing: David Atkinson, Luis-Alberto Cordero-Lecca, Richard Dawid, Lars-Göran Johansson, James Joyce, Barteld Kooi, Ulianov Montano, Ilkka Niiniluoto, Robert Nola, Jeanne Peijnenburg, Isabelle Peschard, Gerhard Schurz, Jan-Willem Romeijn, Allard Tamminga, Sjoerd Zwart. Finally, I like to thank the Netherlands Institute for Advanced Study (NIAS) in Wassenaar for the opportunity to spend two weeks to complete this and another paper.

## Bibliography

- [1] Douven, I. *In Defence of Scientific Realism*. Ph.D. thesis, University of Leuven, Leuven, 1996.
- [2] Douven, I. Testing inference to the best explanation. *Synthese*, 120:355–377, 2002.
- [3] Douven, I. Empirical equivalence, explanatory force, and the inference to the best theory. In Festa, R. Aliseda, A., Peijnenburg, J. (Eds.), *Confirmation, Empirical Progress, and Truth Approximation (Poznan Studies in the Philosophy of the Sciences and the Humanities)*, 281–309. Amsterdam/New York, NY: Rodopi, 2005.
- [4] Jones, R. Realism about what? *Philosophy of Science*, 58:185–202, 1991.

- [5] Kayzer, W. *Het boek over de schoonheid en de troost*. Amsterdam: Contact, 2000.
- [6] Kuipers, T. *From Instrumentalism to Constructive Realism. On some relations between confirmation, empirical progress, and truth approximation (Synthese Library)*, vol. 287. Dordrecht: Kluwer Academic Publishers, 2000.
- [7] Kuipers, T. *Structures in Science (Synthese Library)*, vol. 301. Dordrecht: Kluwer Academic Publishers, 2001.
- [8] Kuipers, T. Beauty, a road to the truth. *Synthese*, 131:291–328, 2002.
- [9] Kuipers, T. Inference to the best theory, rather than inference to the best explanation. kinds of abduction and induction. In Stadler, F. (Ed.), *Induction and Deduction in the Sciences*, 25–51. Dordrecht: Dordrecht: Kluwer Academic Publishers, 2004.
- [10] Kuipers, T. Theories looking for domains. fact or fiction? In Magnani, L. (Ed.), *Model-based reasoning in science and engineering*, Studies in Logic, vol. 2, 33–55. London: College Publications, King’s College, 2004.
- [11] Kuipers, T. Introduction: Explication in philosophy of science. In Kuipers, T. (Ed.), *General Philosophy of Science: Focal Issues (Handbook of the Philosophy of Science)*, vol. 1, vii–xxiii. Amsterdam: Elsevier, 2007.
- [12] Ladyman, J. *Understanding Philosophy of Science*. London: Routledge, 2002.
- [13] Ladyman, J. Ontological, epistemological, and methodological positions. In Kuipers, T. (Ed.), *General Philosophy of Science: Focal Issues (Handbook of the Philosophy of Science)*, vol. 1, 303–376. Amsterdam: Elsevier, 2007.
- [14] McAllister, J. *Beauty and Revolution in Science*. Ithaca: Cornell UP, 1996.
- [15] Niiniluoto, I. *Truthlikeness (Synthese Library)*, vol. 185. Dordrecht: Reidel, 1987.
- [16] Niiniluoto, I. *Critical Scientific Realism*. Oxford: Oxford University Press, 1999.
- [17] Niiniluoto, I. Evaluation of theories. In Kuipers, T. (Ed.), *General Philosophy of Science: Focal Issues (Handbook of the Philosophy of Science)*, vol. 1, 175–217. Amsterdam: Elsevier, 2007.
- [18] Panofsky, W., Phillips, M. *Classical electricity and magnetism*. London: Addison-Wesley, 2nd edn., 1962.
- [19] Psillos, S. *Scientific Realism: How Science Tracks Truth*. London: Routledge, 1999.
- [20] Sober, E. *Philosophy of Biology*. Oxford: Oxford University Press, 2000.
- [21] van Fraassen, B. *The Scientific Image*. Oxford: Clarendon Press, 1980.
- [22] van Fraassen, B. *Laws and Symmetry*. Oxford: Clarendon Press, 1989.
- [23] Weinberg, S. *Dreams of a Final Theory*. London: Vintage, 1993.
- [24] Zajonc, R. Attitudinal effects of mere exposure. Monograph supplement 9 of *The Journal of Personality and Social Psychology*, 1968.
- [25] Zajonc, R. Evidence for nonconscious emotions. In Ekman, P., Davidson, R. (Eds.), *The Nature of Emotion*, 293–297. Oxford: Oxford University Press, 1994.