Reviewing Reduction in a Preferential Model-Theoretic Context

Emma Ruttkamp and Johannes Heidema

In this article, we redefine classical notions of theory reduction in such a way that model-theoretic preferential semantics becomes part of a realist depiction of this aspect of science. We offer a model-theoretic reconstruction of science in which theory *succession or reduction is often better—or at a finer level of analysis—interpreted as the result of* model *succession or reduction. This analysis leads to 'defeasible reduction', defined as follows: The conjunction of the assumptions of a reducing theory T with the definitions translating the vocabulary of a reduced theory T′ to the vocabulary of T, defeasibly entails the assumptions of reduced T′. This relation of defeasible reduction offers, in the context of additional knowledge becoming available, articulation of a more flexible kind of reduction in theory development than in the classical case. Also, defeasible reduction is shown to solve the problems of entailment that classical homogeneous reduction encounters. Reduction in the defeasible sense is a practical device for studying the processes of science, since it is about highlighting different aspects of the same theory at different times of application, rather than about naive dreams concerning a metaphysical unity of science.*

1. Introduction

Philosophers of science have attempted in many different ways to offer accounts of the process and progress of science in terms of the relation(s) of theory succession. One well-known way to consider this issue is to focus on theory reduction (others are emergence and supervenience). Reduction has long been associated with the quest for the unity of science, and since it has become fashionable for some to entertain thoughts of a disunity of science (Cartwright 1999; Dupré 1993), reduction has gone out of fashion. (For an opposing view, see for example Taylor 2001 and Maxwell 1998.) In this article, we want to show, among other things, that reduction is much more viable than one

Emma Ruttkamp is Professor of Philosophy at University of Johannesburg, South Africa; Johannes Heidema is Professor of Mathematics at the University of South Africa in Pretoria, South Africa.

might have thought, given that it need not be bound to a metaphysical unity of science nor to naive correspondence theories of reference at all.

Ernst Nagel (1949, 1961) offered his derivational definition of theory reduction as providing a method of explanation within the accepted positivist covering law model for scientific explanation. Given that Nagel's model of theory reduction was presented within the positivist paradigm, keep in mind that only theories having sets of core theoretical laws and vocabularies separable into two mutually exclusive classes of theoretical and observational (or empirical) terms, and which may be formalized into classical first-order languages, were eligible for Nagel's kind of reduction. Taking into account the problematic nature of the definition of correspondence rules, and especially the fact that the theoretical/observational distinction turned out to be non-unique and very much more heuristic than could really be fitted into the strict verificationist model of positivist methodology (Ruttkamp 2002), this positivist definition of theory reduction has understandably come in for considerable criticism over the past 40–45 years.

Critical analyses of Nagel's account of theory reduction have been presented by well-known philosophers such as Feyerabend (1998), Nickles (1998), Sklar (1967), Schaffner (1967), Kitcher (1998), and the defenders of the structuralist school led by Sneed (Balzer, Moulines, and Sneed 1987) and Stegmüller (1976). Feyerabend outright rejected Nagel's 'correspondence analysis' (Feyerabend 1998, 1039), while Nickles (1998, 953) comes quite close to our view in so far as he recognizes not only that reduction is about explanation but that a certain kind of reduction also has 'justificatory' and 'heuristic' functions. Sklar (1967) stresses with Schaffner (1967) the role of so-called 'theoretic identification' in reduction, which is a problematic view, not least because of the difficulties related to the identification of properties. Kitcher (1998) is well known for his denial of the possibility of reducing classical genetics to molecular biology. Schaffner, in turn, refines reduction from merely being a logical-consequence relation to being a 'product' (Curd and Cover 1998, 961) of this relation with an analogy relation between the reduced theory (the 'secondary' (ibid.) theory) and a 'corrected secondary theory' (ibid.).

This article also offers a critical reconsideration resulting in a generalization of the classical definition of reduction. Our account of reduction broadly fits Nickles's description below, but strengthens it, and also, in a somewhat more relational way than in Nickles's case, addresses a misunderstanding concerning so-called 'domain-preserving' or 'historical reduction'. Nickles writes:

> Reduction is considered by philosophers chiefly to involve an increase in the over-all efficiency of a conceptual scheme, either by outright elimination and replacement of portions that do not work or do their work badly, or (preferably) by consolidation of those parts of the scheme whose work is discovered to overlap. 'Reduction' *means* 'elimination', 'trimming down', 'consolidation'. This consolidation or general increase in the efficiency of the conceptual scheme may be conceptual—a reduction in the number of independent theoretical assumptions—or ontological, or both at once, as in the reduction of optics to electromagnetic theory. … The reduction of theories by their historical successors constitutes a second great tradition of theoretical reduction, which, however, does not fit the above characterisation well. (Nickles 1998, 951–952)

Our account of reduction is presented within a model-theoretic reconstruction of science. We use a certain kind of 'default rule' determining rankings of models and of theories which relate to defeasible entailment, and which allows what we shall term 'defeasible reduction'. We define this kind of reduction as follows: the conjunction of the assumptions of a reducing theory T and certain 'linking assumptions' translating the vocabulary of a reduced theory T' to the vocabulary of a reducing theory T, defeasibly entails the assumptions of the reduced T'.

Why a model-theoretic interpretation of science? And why a definition of *defeasible* reduction? We view science as an historical process—or even a set of historical processes—which continuously, in many different complex contextual ways, revises itself. Although we think it is very important to accept that only science can answer ultimate questions concerning reference ('when do we "touch" reality?'), science, because of its contextual and defeasible nature, should not be viewed as holding a mirror with Nature reflected in it. We have all been influenced one way or another by the failure of naive correspondence theories of truth, and most of us are aware of the role that idealization plays in science.

In our context, it is simply incorrect to conclude from the 'no miracles' argument for realism that if we are realists, we should be realists about a world exactly like science describes. Rather, our claims in this article show how arguments for realism can say that we should believe in a world rather like that which science describes, but that it is no simple matter to say what aspects of science refer to the world, what aspects refer to bits of our models, what aspects are and are not equivalent, and so on. Also, most importantly, these 'bits' of science do not fit together without serious work being done. Therefore, the 'no miracle' argument should not—cannot—be interpreted in an ahistorical or naive way.

Our views regarding the nature of scientific knowledge fit well with the following characterization of science and its method(s) offered by Ilkka Niiniluoto (1999, 5). He depicts science as a 'local belief system' (ibid.), describing it as 'a source of cognitive attitudes about the world, characterized by its reliance on the self-corrective scientific method' (Niiniluoto 1999, 4). He claims that

> for the most part, scientific activities do not involve belief in the sense of holding-to-be-true: rather … [based on certain assumptions or so-called background knowledge] scientists propose hypotheses and pursue research programmes in investigating the limits of the correctness of their theories. (ibid.)

He concludes that science, if successful, 'will … have tentative results, in principle always open to further challenge by later investigations, which constitute what is usually referred to as the "scientific knowledge" of the day' (ibid.). Scientific knowledge, in these terms, is defeasible and may be represented in many different ways, but it is also continuous, sometimes cumulative, and always rational.

We believe that a model-theoretic representation of science, with preferential relations where appropriate, offers a mechanism with which we can magnify, as it were, the infinitely many different pathways existing among theories, their models, and real systems. It is interesting to note here that Margaret Morrison has also offered an approach to theoretical reduction and, more explicitly, of 'unifying scientific theories'

by making use of mathematical structures, in her book, *Unifying Scientific Theories: Physical Concepts and Mathematical Structures* (Morrison 2000). Also, of course, the highly complex analysis of the structure of theories offered by the structuralists (including Stegmüller 1976, and Balzer, Moulines, and Sneed 1987) offers very refined arguments of theory reduction with the help of mathematical structures.

In what follows, we shall very briefly set out the core of the classical definition of theory reduction. Some of the main critiques or alternatives available at present may be mentioned but cannot be fully discussed for reasons of space (Section 2). We then explain our model-theoretic analysis of science (Sections 3 and 4), after which we shall set out the main points of our defeasible reduction (Section 5), followed by a discussion of how defeasible reduction addresses some of the main problems of classical reduction (Section 6). To conclude, we reiterate our realist stance in Section 7.

2. Positivist Theory Reduction

Ernst Nagel defines theory reduction in *The Structure of Science* as follows: '… reduction … is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain' (Nagel 1961, 338). 'Explanation' here means that T reduces T' just in case the laws of T' are logically derivable from those of T. Depending on whether or not the vocabulary of the reduced theory T' is a subset of that of the reducing theory T, Nagel distinguishes two types of reduction.

If all the descriptive terms of the reduced theory T' are contained in the reducing theory T, and the terms of T' are understood to have approximately the same meaning as they have in T, then the reduction of T' by T is called 'homogeneous' (ibid.). In other words, the formal structure of homogeneous reductions is in general the same as the structure of deductive explanations (Nagel 1998, 908). The second type of reduction is called 'heterogeneous reduction' (or 'inhomogeneous reduction', Nagel 1961) and is the type of reduction present in cases where 'at least one descriptive term in the conclusion neither appears in the premises nor is definable by those [terms] that do occur in them' (Nagel 1998, 908). Nagel adds that in such cases, '… it is clear that additional assumptions must be introduced as to how the concepts characteristically employed in the reduced laws, but not present in the reducing theory, are connected with the concepts that do occur in the latter' (ibid., 911). This 'supplementation' (ibid., 913) of T must be done by 'rules of correspondence or bridge laws' (ibid.). Nagel explains that since c-rules cannot be restricted to explicit definitions (given the condition of non-creativity for definitions), they must be a certain kind of empirical hypothesis or statement which states certain relations of dependence between things mentioned in the reduced and reducing theories (ibid.). We use much the same kind of reasoning regarding c-rules, or 'linking assumptions', in the defeasible reduction context (see Section 5).

The usual criticisms of classical reduction are given in the contexts of meaning change, consistency between the reduced and reducing systems, and the nature of the 'linking' assumptions. Defeasible reduction can offer new perspectives on each of these

classes of critique, as will be discussed in Section 6. There is also a fourth point of critique against classical reduction which concerns us particularly, given that it implies that the use of entailment in the definition of homogeneous reduction is highly problematic and, in simple terms, simply wrong. We show, in Section 6, how it might be possible to solve this problem, and still work with generalized entailment in a definition of reduction.

3. A Model-Theoretic Account of Science

A model-theoretic realism retains the notion of a scientific theory as a (deductively closed) set of sentences expressed in some appropriate classical first-order language, while simultaneously emphasizing the interpretative and referential role of the conceptual (i.a. mathematical) models of these theories. There are many reasons for retaining this syntactic notion of theories typical of the 'received' or 'statement' view of science, just as there are many good reasons for linking such a view of theories to typical 'non-statement' or 'semantic' analyses of scientific knowledge in terms of (mathematical) structures (or models). These reasons are explored in depth in Ruttkamp (2002). Briefly, we need the 'theory' so that we are not accused of empty relativism. The theory is the one specification which all its models and their substructures have in common, and we need to retain this relationship between theories and models, given that our account is *relational* rather than *relativist*. Also, the definition of defeasible entailment, which we find so suitable for our purposes, is a relation on theories.

At the same time, though, we need models or mathematical structures, because we view the process of science as manifesting itself in a differentiated way. This means that we believe that the actual growth and internal turmoil theories undergo before they may be applied, revised or, ultimately, succeeded, take place at the level of mathematical structures interpreting the theory. We identify three different such structures (interpretations or models, empirical substructures of these models, and so-called 'empirical models'), to be explained below. Moreover, contextual factors—so-called 'themata' (Holton 1995)—and preferences are not syntactically represented in the theory. In an article entitled 'When scientific models represent', Daniela Bailer-Jones writes:

> Moving towards a viable concept of when a scientific model represents will require going beyond determining a set of propositions that is entailed by the model, without denying that scientific models which represent entail at least some true propositions about the empirical world. (Bailer-Jones 2003, 61)

We agree in principle, if not entirely in exact meaning, in the sense that 'beyond' propositions, we see themata which are crucial factors in the construction of models and empirical substructures (reducts) of theories. We need somehow to find a mechanism with which to capture the role of these factors during the 'life' of a theory, and we find it in mathematical model theory, accompanied by logic(s) with a preferential semantics.

Think also, in this context, of the literature surrounding the so-called class of intended applications, I, in the structuralist programme. Many problems are created by

the insistence, on the one hand, that the class I of intended applications is part of and internal to the theory concept itself, and not somehow external to the theory; and on the other hand, the insistence that the intersection between the class I of intended applications and the class M of actual models of the theory must be viewed in terms of 'pragmatic–diachronic considerations' (Moulines 1991, 321). An impasse is created because Balzer, Moulines, and Sneed (1987, 38 ff.) claim that without I, it is impossible to know the empirical content of the theory, while on the other hand they acknowledge the 'genidentical' (Moulines 1991, 322) identity of the class I, i.e. '… I is dependent on the scientific community within which the theory under consideration has been constructed or will be applied …' (Ruttkamp 2002, 102).

In our context, we do not face this impasse, first, because the empirical content of a theory is given by its empirical models which are linked to the theory via the explicit rules of mathematical model theory and specific relations of isomorphism between empirical substructures and empirical models (explained below); and second, the context-specific nature of the motivations for the formulation (construction) of models of theories is fully taken care of by the suppleness of the model-theoretic relation of satisfaction.

Most important, perhaps, is the fact that we acknowledge that the factors causing us to 'prefer' one application or model of a theory above another are also changeable and context-specific and, above all, not part of the syntax of the theory. These factors are better captured in a preferential framework, where models are ranked according to preference determined by so-called 'default rules' which are (also) external to the language in which the theory is formulated. In other words, 'default rules' cannot be expressed as sentences of the logic language whose valuations and interpretations (models) are intended to 'relate to' a given real system. Neither do themata form part of the syntax of a theory, and so we offer our model-theoretic account of science within the context of a non-classical logic, based not on syntactic inference rules, but on preferential semantics. More specifically, as Labuschagne and Heidema (2005) claim:

> Because every default rule is represented by a different preference ordering on states of the world, there is no way, even in principle, to simulate the defeasible consequence relation by an algorithm based on syntactic inference rules. The shape of a symbol has simply no connection with the notion of normality or preference that may apply.

We can thus much more explicitly capture the external factors that help determine the ebb and flow of the life of theories by our preferential model-theoretic mechanism. More in-depth discussion of the preferential aspect of our account follows, in Section 5, after the model-theoretic aspect has been discussed in what remains of this section.

Let us return now to the main outlines of our model-theoretic account. We know from the structure of mathematical model theory and its definitions of interpretations of (sentences in) formal languages of the possibility of *many different* models of a given theory T (in language L). These models are interpretations of T's language such that a model of a theory sees to it that every predicate of the language of the theory has a definitive extension in the underlying domain of the model which satisfies T in the Tarskian sense. The choice of model(s) is determined by—first, as already

mentioned—the research intentions and thematic preferences of the scientists applying T within some accepted Kuhnian disciplinary matrix, or 'against' some background meta-theory. The choice of model(s) is also determined by the syntax of the theory—i.e. the models are interpretations of the *language* of the theory such that the theory is 'true' or 'satisfied' by them.

Then, focusing on a particular real system at issue in the context of applying a theory, which in its turn implies a specific empirical set-up in terms of the situationally observable properties and measurable quantities of that particular real system, the chosen model is trimmed ('reduced') in a specific way. In the context of theory application, and considering relations of reference, it makes sense to concentrate only on the predicates in the mathematical model of the theory under consideration that may be termed 'empirical' predicates (in the particular context of application). This is how an empirical substructure, or 'reduct' is formulated. Recall that a 'reduct' in model-theoretic terms is created by leaving out of the language and its interpretations some of the relations and functions originally contained in these entities. This kind of structure thus has the same domain as the model in question but contains only the extensions of the empirical predicates of the model. Notice that these extensions may be infinite since they still are the full extensions of the predicates in question.

Now, looking at theory application from the 'other' side, i.e. the side of real systems, from the experimental activities carried out in relation to the real system we are focusing on, a conceptualization of the results of these activities, i.e. of the data resulting from certain interactions with this system (such as performing certain observations and experimental tests), may be formulated. This (mathematical) conceptualization of data is represented as an 'empirical model' and is determined by data gained as a result of interaction with some aspects of the real system in question, not by the models or empirical reducts already constructed. Then, should we find that there is a one-to-one isomorphic embedding function from the empirical model into a certain empirical reduct, this would imply that there exists some relation of reference from our original theory via at least one model and at least one of its reducts to the real system (represented as at least one empirical model) we are considering. The empirical model contains finite extensions of the empirical predicates at issue in the empirical reduct, since only a finite number of observations can be made at a certain time. In other words, an empirical model usually involves only finitely many elements from the domain of the model (or reduct) and from the extensions of the predicates.

To summarize: a model interprets all terms in the appropriate relevant language and satisfies the theory at issue. In the empirical reduct only the terms called 'empirical' in the particular relevant context of application or empirical situation are interpreted, implying that in the model-theoretic context, a theory/observation distinction is non-unique. Think of this reduct (substructure) of the interpretation (model) as representing the set of all atomic sentences expressible in the particular empirical terminology true in the model. An empirical model—still a mathematical structure—can be represented as a finite subset of these sentences, and contains empirical data formulated in the relevant language of the theory.

Empirical adequacy of theory T can then be defined as follows: for every empirical interaction i with reality (thus far), the results of which T is supposed to describe, the resulting empirical model E_i can be isomorphically embedded into an empirical reduct M_i of a model M of T. Here, the empirical terms occurring in E_i and M_i depend on interaction i.

Because one language may be interpreted in many ways, i.e. because the relations of satisfaction between a given theory and its models are non-unique, i.e. domain- (or context-) specific, already the tracing (or reconstruction) of such relations is complex. It gets worse, though. From the empirical side of 'reference', we are faced also with potentially many different empirical models representing (aspects of) the same real system—referred to as 'over-determination' by Ruttkamp (2002, 45–58). This adds to the complexities of the process of science in two ways: First, finding a relation of isomorphic embeddedness between a certain empirical model and some empirical reduct of a model (of a given theory) is not necessarily a simple thing, and decidedly not something that can be manipulated, in the kind of context sketched by Putnam's (1983) 'paradox' (Ruttkamp 2002, 77–90). Second, the choice of model at a given time is also context-, application-, and goal-specific, and thus revisable. Choosing, or 'preferring', one model 'above' another is a temporary action, which does not deny the existence of other models, maybe accommodating different (or the same) empirical models of the same system, all of which we must still be able to 'trace' or 'reconstruct' somehow. This is where defeasible preferential semantics comes in.

4. Non-Classical Logics with Preferential Semantics

Contrary to traditional characterizations of logic as the archetype of rigidity and absolute truth, tying logical analyses to exact—in the sense of 'unique'—determinations of the meaning of linguistic expressions cannot succeed, given the fact that these determinations are contingent on the nature of the very models they help define. This does not cut out logic from our depiction of science and its philosophy at all, though, as we shall see below (see also Ruttkamp 2003).

In this section, we shall explore developments in formal semantics and knowledge representation which will briefly show that relationships between logic and philosophy in the context of reflections on science are alive and well. We shall argue that a model-theoretic preferential analysis of science offers the possibility of a rational discussion of science and its processes. This is possible even if science is analysed in terms of non-unique interpretations and complex clusters of model-specific theory applications.

Now, let us briefly consider the role of logic in studies of knowledge representation of real systems. The notion of a formal language has one of its main foundations in Frege's 1879 *Begriffsschrift*, in which he developed Leibniz's notion of a 'calculus of signs, an artificially constructed language having a precisely defined grammar and unambiguous sentences' (Heidema and Labuschagne 1999). Frege used his notion of a formal language to construct a foundation for mathematics, and to show that the truth of mathematics follows from universal logical principles. The objective of logicism was thus the construction of 'one large and complex formal language, the universally valid sentences

of which would represent the basic truths of mathematics' (Heidema and Labuschagne 1999)—much as the universal meta-language as proposed by philosophy in its traditional foundationalist sense (and subsequently targeted by Lyotard) was by some supposed to represent the basic truths of natural science. At the beginning of the twentieth century, knowledge representation was thus still caught up in Russell's rigid logical atomist paradigm, according to which the denotation of constants and the consequent meaning of sentences were taken to be uniquely assigned. In their continuance of Frege's programme, Russell and Whitehead, however, came up against 'legitimate mathematical assumptions that were not universal logical principles, most notably the axioms of infinity and choice' (ibid.).

This eventually contributed to a shift in the focus of logic studies towards sentences that could be regarded as representations of knowledge of *particular* systems without being universally valid. These sentences were not taken to belong to *one universal super* language, but rather to *different calculi*, 'each having a vocabulary designed to suit the representation of knowledge about the components of [some relevant] ... system' (ibid.). See also, in this regard, Hintikka's (1989, 1997) extremely important distinction between language as universal medium and language as calculus. The most important consequence of these developments was the acknowledgement that each formal language 'admits a large collection of possible interpretations' (ibid.). By the 1950s, Alfred Tarski's model theory, and his views in particular on truth and logical consequence, had matured into a definition of truth in terms of relations between sentences and interpretations. His studies of the properties between sets of sentences and classes of interpretations opened up new horizons for studies in formal logic in general, and knowledge representation in particular.

This new development in studies in knowledge representation and its application in non-classical logics undermine the connotation of 'absoluteness' traditionally given to the word 'knowledge' that used to rule reference to 'laws of nature' (Heidema and Labuschagne 2001). Given our view of science as a body of defeasible knowledge claims standing in certain time-bound relationships to reality, we advocate applications of contemporary non-classical artificial intelligence logics (such as non-monotonic logic, epistemic modal logic, belief revision, and temporal logic) to defeasible scientific knowledge, without giving up on rationally accounting for either the processes of science in general or, in particular, for the motivations behind particular choices for certain representations of real systems above others at certain times. We view the application of these non-classical logics to the problem of partial or defeasible knowledge, not in terms of the knowledge of individual agents as is the case in artificial intelligence and cognitive science applications, but rather in terms of representing the knowledge of the various contexts in which the processes of science take place—from disciplinary matrices as background to a certain theory or set of theories, to much more particular empirical models representing aspects of real systems.

Now, against this background, let us consider the 'preferential' aspect of model-theoretic realism. In particular, for our purposes here, Yoav Shoham's (1988) model-theoretic non-monotonic logic is preferable, since it offers a fairly simple way of ranking models, which perhaps is not as adequately possible in other versions of

non-monotonic logic. Note that this kind of non-monotonic logic is a very small part of a very complex set of non-monotonic logics (see Delgrande *et al.* 2004, for a detailed discussion of the field). Non-monotonic logics with a preferential semantics originated with McCarthy's notion of circumscription (McCarthy 1980, 1986). Shoham (1988) generalized McCarthy's work in his book, *Reasoning about change*. His work was in turn refined by Lehmann and Magidor (Kraus, Lehmann, and Magidor 1990; Lehmann and Magidor 1992; Lehmann 2001). Note that in our current application of Shoham's logic below, our application starts at a finer level of analysis than is usually the case in non-monotonic contexts (where we simply look at rankings of the states—models—of the system in question). The model-theoretic structuring of relations between models, empirical reducts, and empirical models makes possible the refined type of 'carrying over' of rankings that we shall set out below.

A non-monotonic logic consists (for our present expository purposes) of a propositional or predicate language together with a preferential semantics. The main idea is that an agent (a community of scientists working in some disciplinary matrix) may have two kinds of knowledge (Heidema and Labuschagne 2001): sentential information about the aspects of the real system at issue, and which may be expressed in the 'designer-built vocabulary' of the relevant formal language (or calculus; ibid.); and meta-information depicted in terms of so-called 'default rules', and motivating certain choices the agent/scientist makes at a given time. Notice that there is no grand scheme of absolute knowledge 'serving' these agents, as it were, but rather that meta-information here is a local changeable concept. The standard representation of meta-information is as a binary comparative relation on the set of states of a system.

Defeasible reasoning is the process of making informed 'choices' on the basis of a mixture of definite knowledge and heuristic considerations determining our processing of this knowledge, informally expressed as so-called default rules, and formally as relations called 'total pre-orders'. In the case of the preferential semantics related to non-monotonic logics, a total pre-order—which determines defeasible entailment in a given context—is a preference relation on states (worlds, models) and is a reflexive, transitive relation capable of effecting comparisons between arbitrary states.[1] Intuitively, such relations are thought of as allocating models or states (of some real system) to levels of normality, or preference.

The meaning of a formula in classical logic is the set of interpretations that satisfies it, or its set of models. In the context of a non-monotonic logic we focus on a subset of those models, that is, those that are 'preferable' or 'most preferred' in a certain respect (these preferred models are for historical reasons sometimes called 'minimal models'). The semantic consequence relation, namely defeasible entailment, defined by total pre-orders, represents the 'key distinction between defeasible and indefeasible inferences' (Ginsberg 1987, 9), since it makes explicit the difference between monotonic and non-monotonic reasoning in the following way. In classical logic, $A \models C$ if C is true in *all* the models of A, however 'unwanted' or 'inapplicable'. Moreover, since all the models of $A \wedge B$ are also models of A, it follows that $A \wedge B \models C$, and hence that an increase in the knowledge represented by the antecedent of an

entailment relation in classical logic does not invalidate the knowledge represented by the consequent of the relation, and so classical logic is 'monotonic'. In line with the fact that 'defeasible conclusions may need to be retracted in the presence of additional information' (ibid.), in a non-monotonic framework we define $A \mid\sim C$ (read: A defeasibly entails C) to mean that C is true in all *preferred* models of A, which implies that we choose only a subset of the models of A, according to some preference we have for them at a given time. Furthermore, in terms of an increase perhaps of our knowledge of the system at issue, $A \wedge B$ may have preferred models that are not preferred models of A, and so the consequences of A are not necessarily included among those of $A \wedge B$ in the non-monotonic context.

The word 'defeasible' reflects the fact that our preference may change, in other words that the default rule may be 'defeated' by exceptional circumstances, or a change in circumstances caused by a change in the content of our knowledge. Defeasible inferences are inherently non-monotonic, since enhancing our system of assumptions might change our conclusions. 'Thus ... [preferential] model semantics provides one way to make precise the notion of a defeasible belief: a sentence supported by the agent's knowledge in the sense of being true in the ... [preferential] models of that knowledge' (Heidema and Labuschagne 1999). Raymond Reiter (1980, 81) comments that:

> The need to make default assumptions is frequently encountered in reasoning about incompletely specified worlds. Inferences sanctioned by default are best viewed as beliefs which may well be modified or rejected by subsequent observations. It is this property that leads to the non-monotonicity of any logic of defaults.

In terms of philosophy of science, the above offers a mechanism for showing and expressing the fact that we do sometimes reflect on knowledge, and its acquisition, communication, and representation from some 'meta'-stance, but that such stances are still *local*; moreover, they are not local in the sense that they can merely be reduced to context but rather in the sense that they represent amendable, tentative, or defeasible viewpoints.

5. Preferential Model-Theoretic Reduction

'Defeasible reduction' of T' by T may be defined as follows: the conjunction of the assumptions of a reducing theory T and the linking assumptions of terms in T', supplementing the vocabulary of the language of T, *defeasibly entails* the assumptions of the reduced theory T' (expressed as $T \wedge D \mid\sim T'$, where D stands for the set of linking assumptions translating T' into the language of T). Notice that here it is not simply the reducing theory T which defeasibly entails the reduced theory T' but the reducing theory T in the presence of some set D of linking assumptions. (Recall that stating that A defeasibly entails B means that the set of preferred models of A, written as $\mathrm{Pmod}(A)$, is a subset of the set of models of B, written as $\mathrm{Mod}(B)$.)

We shall now unpack the notions implied by the above definition in more detail. Our discussion will include the following: the merged language, the default rule determining $\mathrm{Pmod}(T \wedge D)$, a discussion of the philosophical aspects of defeasible reduction, and a

toy example. In terms of the merged language, it seems obvious that both the conjunction of T and D, and theory T' must be formulated in the same language. This 'merged' language will usually be the union of the languages of T and T'. We assume that $T \wedge D \wedge T'$ (for simplicity taken as a sentence of the union of the languages of T and T') is consistent. D consists of one or more sentences of the merged language, linking the two sets of vocabularies. Usually, D will contain definitions of the T'-terms not occurring in the T-language in terms of the T-language. Occasionally, D may contain definitions of new terms which help to smooth other definitions. An example: the notion of the average kinetic energy of gas molecules may facilitate the definition of temperature (from T' = thermodynamics) in terms of mass, speed, etc. (from T = mechanics). In principle, D may sometimes even contain linking axioms which transgress the usual eliminability and non-creativity of genuine definitions. In these cases, T is effectually strengthened logically.

To check whether defeasible reduction holds logically, the models of $T \wedge D$ (or, usually, all interpretations of the language of $T \wedge D$) must be ranked according to a particular default rule (determined by meta-information). This default rule is a preference order (reflexive, transitive, and a total relation) so that Pmod($T \wedge D$) can be defined as the set of (most) preferred models of $T \wedge D$. Ruttkamp (forthcoming) has proposed a specific default rule in terms of two conditions—more empirical terms and more precision. Given both that the default rule she discusses is just one possible default rule (i.e. the feasibility of the rankings presented by a default rule cannot be established in general, once and for all) and the fact that the motivations behind default rules are context-specific, in what follows we offer six examples of information at the meta-level which co-determine six possible preference orders applicable in the context of theory reduction in science.

(i) Historically, the first preference criterion for worlds, considered by McCarthy (1980) as a special case of 'circumscription', was to have as few abnormal individuals as possible. He introduced a unary predicate *Ab* for 'abnormal' and considered a world to be preferable the smaller in number the extension of *Ab*, i.e. the number of abnormal individuals, in that world is. As an example, suppose that from 'Tweety is a bird', we want to infer defeasibly that 'Tweety can fly', since we feel that normally birds can fly. Then, in any world, the set of non-flying birds will constitute the extension of *Ab*. In all most-preferred models of 'Tweety is a bird', Tweety will lie outside the extension of *Ab*, i.e. will be able to fly, and we can then draw the nonmonotonic conclusion 'Tweety can fly'. (Of course we assume that no contradictory evidence, such that Tweety is a penguin, or is dead, is present.) Note how the preference order is bound to the specific context, namely the need to express the meta-information that normally birds can fly.

(ii) A second possibility is to use information regarding precision or accuracy to define a preference order. In such a case, the interpretations are ordered by the precision of sentences true in them which express data from empirical models. Technological developments these days make possible empirical statements of almost unbelievable precision compared with previous times. For example, in physics we have extraordinarily accurate theories. Penrose reminds us that:

> In quantum field theory, which is the combination of quantum mechanics with Maxwell's electrodynamics and Einstein's Special Theory of relativity, there are effects which can be computed to be accurate to about one part in 10^{11}. Specifically, in a set of units known as 'Dirac units', the magnetic moment of the electron is predicted to be 1.001159652(46), compared with the experimentally determined value of 1.0011596521(93). (Penrose 1997, 51)

Or think of a sentences such as 'The rest mass of an electron is $9.1093897(54) \times 10^{-31}$ kg'.

(iii) A third example of a preference rule concerns the set of empirical terms. The interpretations are ordered by the number of terms—the more the better—that can be considered to be empirical, given the context of the application, the available technology, etc. Example: take the term 'atom'. A hundred years ago, Ernst Mach could still consider this to be a purely theoretical term, and doubt the existence of real entities corresponding to this term. Today, with scanning tunnelling electron microscopes, we can see, even manipulate, individual atoms. 'Atom' is now an empirical term—i.e. a previously exogenous term has now become endogenous in a set of empirical reducts of the models of a language interpreted within a disciplinary matrix where the above manipulation is known. Note that we acknowledge that developments of technology and precision, on the one hand, and of the advancements of the scope of experimental science, on the other hand, might not always move in the same direction, but for our 'realist' purposes we consider that they do, so that (ii) and (iii) may often be combined sensibly.

(iv) A fourth possibility is systemic information on the system(s) intended as interpretations of $T \wedge D$, e.g. about physically impossible or unintended states, given a context of application. For example, in the toy example of an electrical system (heater, light, and meter) that follows at the end of this section, certain states are physically impossible, even though this information is not contained in $T \wedge D$. Or, another example is a statement such as 'One might not consider gases in the laboratory with a temperature higher than 1000°C'.

(v) A fifth example of meta-information determining a specific preference order is inside information, likelihood of states, or a rule of thumb, available only to a particular agent, which helps to compare interpretations. For example, in the toy example that follows later, the agent knows that both the heater and the light are more likely to be on than off.

(vi) A last example is that of merged information (e.g. of types like the above five examples) or merged comparative orders based on information (e.g. of such types). We refer here, as an example, to merging in belief revision. Labuschagne and Heidema (2005) point out that non-monotonic logics with preferential semantics are closely related to AGM belief change theory (Alchourrón, Gärdenfors and Makinson 1985). They also (Labuschagne and Heidema, 2005) refer to the fact that Meyer (1999, 75) has shown that 'every AGM belief revision operation can be carried out by simply taking the defeasible consequences according to an appropriate preference ordering'. Also note that one of the main kinds of merging used in belief revision is so-called 'lexicographical orderings', the most extreme case of which is the case where only one of the merging conditions is chosen. In particular, the criteria for comparison are prioritized

linearly. The models are then ordered according to the most important criterion. Models equivalent on the first criterion are then ordered mutually according to the second criterion, etc. for the following criteria.

Now, when the models of $T \wedge D$ have been ranked by a preferential order (or general total pre-order) such as those discussed above, the possibility of defeasible reduction between T and T' may be investigated. After the languages of T and T' have been merged as set out above, we have to formulate interpretations of the language, specifying the domain of discourse, denotations of all individual constants, function symbols, and predicate symbols. The next step is to determine which of these interpretations are models of T', and which are models of $T \wedge D$. Next we consider the ranking of models according to the default rule we chose and so identify the set of preferred models of $T \wedge D$—referred to as $\mathrm{Pmod}(T \wedge D)$. Finally, we check whether it is the case that $\mathrm{Pmod}(T \wedge D) \subseteq \mathrm{Mod}(T')$.

As far as the empirical interpretation of the definition for defeasible reduction goes, the following. Here, we have to take into consideration the model-theoretic analysis of the structure of scientific theories. Recall that an empirical model is a conceptualization of the results of experimental processes. If a theory 'refers' to some objects or relations in some real system, an empirical model (resulting from some experimental or observational interaction with some real system to which the theory may be applied) may be found to be isomorphically embedded into an empirical reduct determined by the relevant application. This reduct is a substructure of a model in which the theory is true. If $T \wedge D$ defeasibly reduces T', then translating (or preserving) the empirical adequacy of T' means that every empirical model $\mathrm{EM}(T')$ must be isomorphically embeddable into some empirical reduct $\mathrm{ER}(T')$ of T', which happens to be a reduct of a preferred model of $T \wedge D$.[2]

In conclusion of this section, before we introduce our toy example, let us briefly summarize the philosophical implications of defeasible reduction as we have encountered them so far. First, the context-specific nature of the preference ordering or default rule seems to be clear now. This particular aspect of default rules points also to their temporary or defeasible nature, which in its turn points to the next philosophical aspect of our discussion: the continuity of science. 'Continuity' here means that science is alive and changing, i.e. it might be cumulative, and sometimes it might not, but at least it is never static. Progress in science should be interpreted as not necessarily always in terms of development into a universally better or higher form of knowledge, but rather, development in the sense of adapting to circumstances or context.

In classical (Nagelian) reduction, all models of $T \wedge D$ are some of the models of T', while in the case of defeasible reduction, some models (the preferred ones) of $T \wedge D$ are some of the models of T'. This implies that the conjunction of T and D is semantically stronger in the defeasible case, since it has fewer models than in the classical case. This, in turn, implies that T might be easier to falsify (more counter-examples exist), which is compatible with a Popperian attitude to theories.

Furthermore, the fact that neither a theoretical/ observational distinction nor empirical adequacy is fixed or unique—which is related to the fact that one empirical model may be embedded into more than one empirical reduct of the same or different models

of the theory, since the formulation of an empirical reduct is relative to the application of a theory and empirical situation at a specific time—means that at least at an empirical level 'meaning change' may be accommodated by the flexibility of model-theoretic realism.

But, what about theoretical terms? Everything at issue is somehow linked to the models of T (we have defeasible entailment after all). We thus have at least a logical connection to the theory T, since it is T which is true in its models (or its preferred models), and furthermore this logical connection is sufficient for the referential intentions of a model-theoretic realist account of science. The contextually empirical terms refer directly, and the contextually theoretical terms indirectly, 'by implication', via their conceptual and logical links to the empirical terms established by the theory. Some philosophers might scorn this kind of 'weak' reference, while actually the realism containing this notion of reference is 'weak' only because 'strong' means traditional metaphysical realism. 'Weak' means '*relational*', and in that sense model-theoretic realism is much stronger, in the sense of being able to deal with very complex relations, and more flexible, than typical metaphysical scientific realism.

In terms of realism, note also the following: we can trace the various reductions of the models of theories into empirical reducts, and the various preferences for certain empirical models (isomorphically embedded into some empirical reduct(s)), in such a way that we can at least get an approximate grip on the complex branchings weaving the fabric of science. In other words, we do not deny that the links between theories and reality are extremely complex, non-unique, or temporary at best. On the contrary, we emphasize these aspects of reference, but because we can *articulate* this complexity with the help of both a model-theoretic analysis of the structure of theories, and the mechanism of a preferential semantics, we remain realists.

Furthermore, some of the examples of possible default rules we set out above (like (ii) and (iii)) imply a ranking of the 'strength' of links between theories and reality embodied in different models of the theory, and the 'empirical interpretation' of the definition of defeasible reduction discussed above implies that empirical adequacy is preserved under defeasible reduction. (The ranking of models in terms of preference does not imply at any stage denying their status as true interpretations of the language of the theory.)

We illustrate defeasible reduction with a toy example. My roommate and I, poor students, rent a room, an outbuilding with entrance and window facing away from the house, from a stingy landlord. Besides the light, we are not supposed to use any electrical appliances. To survive the cold, we have secretly bought a second-hand Kaloriter, a one-bar heater. By switching off everything in the main house and checking whether the meter is still running, the landlord can (unbeknownst to us) check whether we are using electricity. He cannot see whether our light is on, but we can, even from the outside, through the window. One cannot see from the outside whether the heater is on, though.

The language of the landlord has two interpreted propositional symbols:

- l: our light is on (theoretical term);
- m: the meter is running (observational or empirical term).

Observing that the meter is running, the landlord formulates his (soon to be defeasibly reduced) theory

$$T' = l \wedge m.$$

Walking home in the cold to our hopefully cosy room, my language has two symbols:

- k: the Kaloriter is on (theoretical term);
- l: our light is on (partially observational or empirical term).

Remembering our agreement to keep the heater on, I formulate my theory as

$$T = k.$$

We now merge the two languages to the new k, l, m language, but with m defined by

$$m :\equiv (k \vee l)$$

or

$$D = [m \leftrightarrow (k \vee l)],$$

which would of course surprise our landlord! Take into account also that the meter runs when the Kaloriter or the light or both are on.

Using this information, we obtain the truth table shown in Table 1 (note that $T \wedge D = k \wedge [m \leftrightarrow (k \vee l)] \equiv (k \wedge m)$.

We see that the rows of the table marked with an s are *spurious*: they clash with definition D and represent physically impossible states of the system comprising the Kaloriter, the light, and the meter ((iv) above). We also note that $T \wedge D$ does *not* classically entail T', since, in the third row, state 101, $T \wedge D$ is true, but T' is false. For the merged k, l, m language, k is a theoretical term, while l and m are observational (at least when I and the landlord join observational forces), so that we can write the empirical reduct of a state $x\ y\ z$ as $*\ y\ z$. The Kaloriter stays unobserved from outside.

While walking to our room, I construct an order on the eight states of the electrical system, expressing their relative likelihood, what I would normally expect. The four

Table 1 Truth table

				$T \wedge D$	T'
	k	l	m	$k \wedge m$	$l \wedge m$
	1	1	1	1	1
s	1	1	0		
	1	0	1	1	0
s	1	0	0		
	0	1	1	0	1
s	0	1	0		
s	0	0	1		
	0	0	0	0	0

spurious states can be definitely ruled out. Then, given the cold and the agreement with my roommate, I expect the Kaloriter to be on rather than off ((v) above), as expressed in T. What about the light? Knowing the amount of work my roommate has to do, the chances are slightly better that she is in and working under the light than otherwise ((v) again). The state of the meter is completely determined by those of the heater and of the light. So, I come to the preference relation (shown in Figure 1) about what state I am likely to find, with 'higher up' expressing 'more likely'.

The preferred set of models of $T \wedge D$, $\mathrm{Pmod}(T \wedge D) = \{111\}$ is contained in $\mathrm{Mod}(T') = \{111, 011\}$; hence $T \wedge D$ defeasibly entails T' and we have defeasibly reduced T' to T (with D)—all of this relative to the preference order.

Now, the landlord is watching the running meter. He has the empirical (for him) model **1 of his theory $T' = l \wedge m$, which is isomorphically embeddable into *11. The latter is an empirical reduct (in the merged sense) of both a model of T' and of the preferred model 111 of $T \wedge D$. So, the empirical adequacy of the reduced theory T' (having an empirical reduct *11 into which the empirically found model **1 can be isomorphically embedded) extends to the reducing theory $T \wedge D$, in the sense that a preferred model 111 of $T \wedge D$ can be found into which **1 can be isomorphically embedded.

6. Classical vs. Defeasible Reduction

As mentioned in Section 2, the usual criticisms of classical reduction can be presented in terms of three classes: meaning change, consistency, and the nature of 'linking' assumptions. First, as far as meaning change is concerned, the fact that there is no 'neutral meaning' of any observational term is accommodated—and articulable—in a

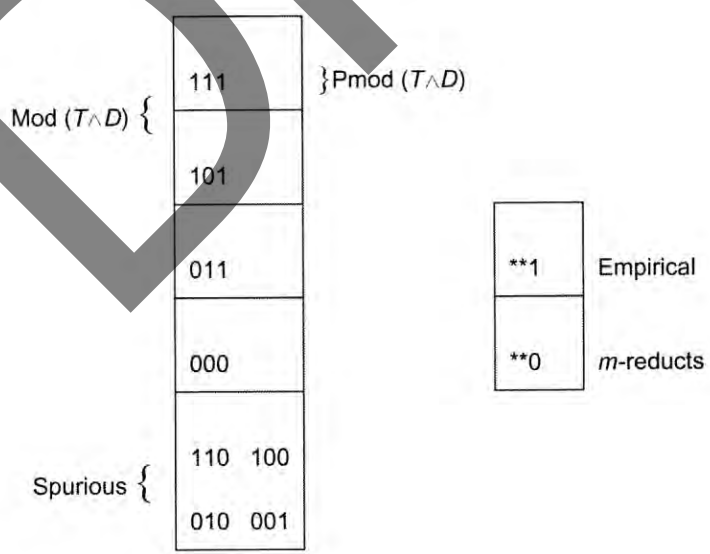

Figure 1 Preference Relation.

model-theoretic preferential framework, since the observational/theoretical divide here is non-unique and relative to particular theory applications in specific empirical contexts (Sections 3–5). Second, any form of reduction common sensically implies consistency (in the semantic sense of sharing at least one model) between the reduced and reducing theories. Moreover, any derivational version of reduction surely must accept this assumption, since classical entailment (and defeasible entailment for that matter) implies at least consistency. Third, the nature of linking assumptions can be either empirical (in cases of heterogeneous reduction) or definitional in Suppes's (1957) sense of the word (in cases of homogeneous reduction), for reasons that will become clear below.

There is a fourth point of critique against classical reduction though, in particular, there is something very seriously wrong with homogeneous reduction. Since this kind of reduction is 'domain-preserving' (Nickles's (1998, 953) term), and usually discussed in terms of relating theories T to their special cases described by T', which seems to imply that T' is made logically stronger than T, in these cases it seems that T' has fewer models than T. This, in turn, implies that claiming that T' can be classically derived from T, or from the conjunction of T and D, cannot be the case. Semantically, if $T \models T'$, this means that $\mathrm{Mod}(T) \subseteq \mathrm{Mod}(T')$, thus that T' has more models than T. Nickles (ibid., 953 ff.) offers a distinction between two kinds of reduction: reduction$_1$ (heterogeneous reduction), and reduction$_2$ (homogeneous reduction). He describes the former as applicable to concurrent theories, the result of ontological economy, and as domain-combining. Here we have logical derivation, and we say T reduces$_1$ to T' (with the help of some kind of linking assumption) if T explains T'. Reduction$_2$ is the reduction of successor theories T to their predecessors T', and is domain-preserving. This is limiting-case reduction, i.e. T reduces$_2$ to T' if T' can be approximately recovered as a limiting or special case of T. According to Nickles (ibid., 953), domain-combining reduction is 'chiefly obtained by derivational reduction in Nagel's sense', while domain-preserving reduction implies an 'invertion' (ibid., 950) of classical reduction such that successors T are said to reduce to their predecessors T' under limiting and other 'appropriate transformations' (ibid.).

Sneed (Balzer, Moulines, and Sneed 1987, 253 ff.) in turn distinguishes between historical and practical reduction. Historical reduction is at issue in cases of '… 'dramatical' epoch-making transitions, like those from Aristotelian to Galilean concepts of motion, …, from caloric theory to phenomenological thermodynamics to statistical mechanics …' (ibid. 253), and in these cases preceding theories reduce to succeeding theories. This happens '… in a way that transmits the main achievements of the preceding theory so that they can also be regarded as achievements of [the succeeding theory]'. This seems to be reduction$_1$ in Nickles's sense, since he writes that historically 'distinct domains of phenomena involve different descriptive vocabularies, almost by definition' (Nickles 1998, 954) and 'must involve the logical derivation of one theory from the other (plus connective principles) … for … [such reduction] absorbs one theory into another without eliminating the former as incorrect' (ibid.). This is obviously domain-combining reduction, and thus heterogeneous reduction in Nagel's sense.

Sneed's 'practical' reduction, though, arises from

> ... problem-solving situations in which theories and their theoretical equations are applied. It often happens that solving some theoretical problem or equation is difficult ... or very costly. In such situations ... [o]ne tries to simplify (to 'reduce') the theory used by consciously omitting some 'parts' of it which for the problem at hand do not distort the 'correct T' solution too much ... [we solve] the problem in a 'coarse' version and thus ... [obtain] an approximating solution of the original problem. (Balzer, Moulines, and Sneed 1987, 254)[3]

This is 'limiting case reduction', or domain-preserving reduction, and thus homogeneous. Sneed concludes that

> ... it appears that 'practical' reduction of a theory T^* to a theory T formally is just the converse relation of the general scheme for a 'historical' reduction of T to T^*... . From this standpoint, the term 'reduction' simply has two usages, one of which is the exact opposite of the other—at least with regard to formal structure. (Balzer, Moulines, and Sneed 1987, 254)

We agree with both Sneed and Nickles in so far as they point to some kind of 'inverted' or 'opposite' reduction in the case of homogeneous (reduction$_2$, or practical) reduction. We, however, still want to construe as entailment the reduction of T' as a special case, because we want to normalize all types of reduction by giving one formal definition applicable to the different cases (which are still distinguishable). In this context, we propose the following approach.

Recall that in classical (Nagelian) reduction, all models of $T \wedge D$ are some of the models of T', while in the case of defeasible reduction, some models (the preferred ones) of $T \wedge D$ are some of the models of T'. And we may say that T is a 'better' reducing theory for T' in the defeasible than in the classical sense, precisely because of the fact that T is interpreted as logically stronger, and thus needs to be true in fewer models in the defeasible case than in the classical reduction case. Now, this makes particular sense when T' is supposed to deal with a small subclass of the class of systems dealt with by T—as is the case when T' is thermodynamics, and T is full Newtonian dynamics. Therefore, to the other criteria (whichever ones we chose) co-determining the preferential ordering on models of $T \wedge D$, we may in this case add another: models of $T \wedge D$ which seem relevant to the limited class of structures that T' is intended to be talking about, are preferred to other models of $T \wedge D$.

In other words, the crucial point is that because in these cases, the class of structures described by T' is a subclass (i.e. a 'special case') of the class of structures described by T, we must find a way to cut back the set of models of T so that it can still fit into (be a subset of) the set of models of T', which has shrunk if it deals only with a small subclass of the class of systems dealt with by T. Now, we may want to do this syntactically by stengthening D, which would mean that we must find (a) sentence(s) in the union of the languages of T and T', expressing the characteristic features of the intended models of T. This will only be possible in very ideal cases. Rather, we turn to semantic ways of cutting back the models of $T \wedge D$ such that they still form a subset of the cutback $\mathrm{Mod}(T')$. And, the standard way of doing this in non-monotonic logic is to use a preference relation to order the models of $\mathrm{Mod}(T \wedge D)$. In other words, we use a default

rule as explained in Section 5 and incorporate the condition that models of $T \wedge D$ which seem relevant to the limited class of structures that T' is intended to be talking about are preferred to other models of $T \wedge D$. This may then result in a defeasible entailment relation between $T \wedge D$ on the one hand and (the logically stronger than T) T' on the other hand. Notice that the syntactic case where we strengthened D is the special case of this more general semantic procedure in which the elements of $\mathrm{Pmod}(T \wedge D)$ are also models of the logically stronger T'.

We know that the Special theory of Relativity yields Classical (Newtonian) Mechanics in the correct limit, yielding the preference order 'the lower the relative speed, the more preferred (i.e. the closer to being Newtonian) the system'. This implies that some (preferred) models of Einstein's theory are (approximately, i.e. within the limit of precision of empirical models) some of the models in Newtonian mechanics, which implies defeasible reduction. However, should the speed limit be set too high, we do not necessarily still have a case of theory reduction, since the reduction would have been defeated.

To summarize, in cases of domain-combining (heterogeneous) reduction, relations of defeasible reduction may be 'better' than those of classical reduction because in the former case, the models of $T \wedge D$ are restricted, and so, here the reducing theory (in the presence of some D) is made logically stronger than in the classical case. In cases of domain-preserving reduction, we may still have entailment, if it is defeasible, since here defeasible entailment offers a way to cut back the models of $T \wedge D$ to such an extent that they may still be a subset of $\mathrm{Mod}(T')$ if $\mathrm{Mod}(T')$ has been shrunk as a result of making T' logically stronger than T, i.e. presenting it as a special case of T.

7. Conclusion

The flexibility of defeasible reasoning turns it into a very effective mechanism in support of an account of scientific progress incorporating falsifiability and defeasibility. Issues of theory succession or reduction are often, at a finer level of analysis, issues of model succession or reduction, and this implies that certain aspects of our knowledge are more temporary than others. The relation of defeasible reduction, given the specification of context of reduction it offers, allows in cases of change in default rules and consequent re-rankings of models in the light of additional knowledge becoming available, articulation of a more flexible kind of reduction than in the classical case. Reduction in the defeasible sense is a practical device for studying the processes of science, since it is about highlighting different aspects of the same theory at different times of application rather than about vague dreams concerning a metaphysical unity of science.

Why is our account of science and its processes *realist*? As touched on in Section 5, we speak of realism, because we show that we can trace relations of references from terms in theories to entities and their features in real systems. Ours is a minimalist kind of realism though, because we show how complex these relations are, that they are always qualified, and that they are mostly temporary and always tentative. We show that a realist stance is not easy to have, and that it takes a lot of work to discover if and

how, at a certain time, theories reflect aspects of reality. We may add that model-theoretic realism is related to structural realism, although we define 'structures' formally as mathematical structures of the Tarskian kind.[4]

More generally, in conclusion, to explain the senses in which our account of science is realist, let us consider within a model-theoretic realist context the five theses of realism that Ilkka Niiniluoto (1999, 10 ff.) formulates as part of his excellent discussion of the various forms of realism philosophy of science has to offer in his book entitled *Critical scientific realism* (from Ruttkamp 2002, 176–78):

- Ontological realism: 'At least part of reality is ontologically independent of human minds' (Niiniluoto 1999, 10). Model-theoretic realism accepts this thesis.
- Semantic realism: 'Truth is a semantical relation between language and reality. Its meaning is given by a modern (Tarskian) version of the correspondence theory, and its best indicator is given by semantic enquiry using the methods of science' (ibid.). Model-theoretic realism also accepts this thesis.
- Theoretical realism: 'The concepts of truth and falsity are in principle applicable to all linguistic products of scientific enquiry, including observation reports, laws, and theories. In particular, claims about the existence of theoretical entities have a truth value' (ibid.). Model-theoretic realists accept this thesis but in terms of truth-in-models and empirical adequacy as defined in Section 3.
- Axiological realism: 'Truth (together with some other epistemic utilities) is an essential aim of science' (ibid.). Model-theoretic realism has a combined notion of truth, namely truth-in-a-model and empirical adequacy. That is why questions concerning truth are referentially interpreted and answered. Preferential model-theoretic truth is essential for scientific progress, but 'truth' in an absolute sense cannot be considered as an aim of science, since the notion itself is without sense in a model-theoretic depiction of science and its processes.
- Critical realism: 'Truth is not easily accessible or recognisable, and even our best theories can fail to be true. Nevertheless, it is possible to approach truth, and to make rational assessments of such cognitive progress' (ibid.). We rather focus on the amendable or defeasible character of the 'truth' of scientific theories which is model-theoretically depicted not by 'approaching truth', but rather by analyses of scientific progress in terms of preferential model semantics, especially in terms of preference orderings represented by default rules as discussed in Section 5.
- Linked to the previous thesis is the last one: 'The best explanation for the practical success of science is the assumption that scientific theories in fact are approximately true or sufficiently close to the truth in relevant aspects. Hence, it is rational to believe that the self-corrective methods of science in the long run has been, and will be, progressive in the cognitive sense' (ibid.). Model-theoretic-realistically, the self-corrective methods of science are also taken as evidence for past and future scientific progress, because this focuses on the continuously revising character of the processes of science. Science is progressive in the sense that its claims, even if defeasible, can be used, sometimes in 'adapted' form, in future processes, because past processes can be 'mapped' or 'traced'.

Model-theoretic realism thus implies ontological and semantic realism, in order to be an epistemological (in terms of scientific knowledge claims) realism, showing traces of critical realism only in so far as the self-corrective methods of science are also accepted and taken to make rational assessments of scientific progress possible.

Again, this form of realism might seem weak by traditional standards. If, though, both truth and reference are interpreted contextually (i.e. model-specifically), shown to be intelligible notions, and shown to hold without a traditional metaphysical analysis of the ontology of reality, then we are actually dealing with a much richer form of realism. Also, model-theoretic realism is much closer to the actual nature of science than metaphysical realism, since no rigid one-to-one relations between scientific language and the world are posited or needed for realizing realist ideals. Tarski's interpretation of truth as reflecting a relationship between sentences in some language and interpretations of that language, rather than a property of a sentence, is the best thing that happened to realism in the last 70 or so years.

Notes

[1] See any textbook on non-monotonic logic, such as Shoham (1988), or Reiter (1980) for formal definitions.

[2] Thus, we may try to determine whether there is an empirical model of T' which cannot be embedded into any empirical reduct of $T \wedge D$.

[3] 'In some cases, the solution thus obtained may even be equally satisfactory from the empirical point of view' (Balzer, Moulines, and Sneed 1987, 254).

[4] Structural realism is one of the most vibrant species of realism at the moment. It has been advocated by, among others, Poincaré (1900, 1902, 1905), Maxwell (1970), Worrall (1989, 1990), and also Psillos (1999). For a good overview of the field, see Psillos (1999).

References

Alchourrón, C. E., P. Gärdenfors, and D. Makinson. 1985. On the logic of theory change: Partial meet functions for contraction and revision. *Journal of Symbolic Logic* 50: 510–30.

Bailer-Jones, D. M. 2003. When scientific models represent. *International Studies in the Philosophy of Science* 17: 59–74.

Balzer, W., C. U. Moulines, and J. Sneed. 1987. *An architectonic for science: The structuralist programme.* Dordrecht: Reidel.

Cartwright, N. 1999. *The dappled world: A study of the boundaries of science.* Cambridge: Cambridge University Press.

Curd, M., and J. A. Cover, eds. 1998. *Philosophy of science: The central issues.* New York: Norton.

Delgrande, J., T. Schaub, H. Tompits, and K. Wang. 2004. A classification and survey of preference handling approaches in nonmonotonic reasoning. *Computational Intelligence* 20: 308–34.

Dupré, J. 1993. *The disorder of things.* Cambridge, MA: Harvard University Press.

Feyerabend, P. 1998. How to be a good empiricist: A plea for tolerance in matters epistemological. Reprinted in *Philosophy of science: The central issues*, edited by M. Curd and J. A. Cover. New York: Norton.

Ginsberg, M. L., ed. 1987. *Readings in nonmonotonic reasoning.* Los Altos, CA, Morgan Kaufmann.

Heidema, J., and W. A. Labuschagne. 1999. Logics for all seasons. In *Papers delivered at the symposium in honour of Professor George McGillivray on the occasion of his retirement*, edited by M. R. De Villiers *et al.* Pretoria: University of South Africa Press, 15–28.

Heidema, J., and W. A. Labuschagne. 2001. Knowledge and belief: The agent-oriented view. In *Culture in retrospect: Essays in honour of E. D. Prinsloo*, edited by A. P. J. Roux and P. H. Coetzee. Pretoria: University of South Africa Press.

Hintikka, J. 1989. Explaining possible worlds. In *Possible worlds in humanities, arts, and sciences: Proceedings of Nobel Symposium 65*, edited by S. Allèn. Berlin: De Gruyter.

Hintikka, J. 1997. *Lingua universalis vs. calculus ratiocinator: An ultimate presupposition of twentieth-century philosophy.* Dordrecht: Kluwer.

Holton, G. 1995. The role of themata in science. *Foundations of Physics* 26: 453–65.

Kitcher, P. 1998. 1953 and all that: A tale of two sciences. Reprinted in *Philosophy of science: The central issues*, edited by M. Curd and J. A. Cover. New York: Norton.

Kraus, S., D. Lehmann, and M. Magidor. 1990. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44: 167–207.

Labuschagne, W. A., and J. Heidema. 2005. Natural and artificial cognition: On the proper place of reason. *South African Journal of Philosophy* 24(2): 137–151.

Lehmann, D. 2001. Nonmonotonic logics and semantics. *Journal of Logic and Computation* 11: 229–56.

Lehmann, D., and M. Magidor. 1992. What does a conditional knowledge base entail? *Artificial Intelligence* 55: 1–60.

Maxwell, G. 1970. Theories, perception, and structural realism. In *The nature and function of scientific theories*, edited by R. Colodny. Pittsburgh, PA: University of Pittsburgh Press.

Maxwell, N. 1998. *The comprehensibility of the universe: A new conception of science.* Oxford: Oxford University Press.

McCarthy, J. 1980. Circumscription: A form of nonmonotonic reasoning. *Artificial Intelligence* 13: 27–39.

McCarthy, J. 1986. Applications of circumscription to formalising common-sense knowledge. *Artificial Intelligence* 28: 89–116.

Meyer, T. 1999. Semantic belief change. PhD dissertation, University of South Africa. Available from http://www.cse.unsw.edu.au/~tmeyer/pubs.html; INTERNET.

Morrison, M. 2000. *Unifying scientific theories: Physical concepts and mathematical structures.* Cambridge: Cambridge University Press.

Moulines, C. U. 1991. Pragmatics in the structuralist view of science. In *Advances in scientific philosophy: Essays in honour of Paul Weingartner*, edited by G. Schurz and G. J. W. Dorn. Amsterdam: Rodopi.

Nagel, E. 1949. The meaning of reduction in the natural sciences. In *Science and civilisation*, edited by R. C. Stauffer. Madison, WI: University of Wisconsin Press.

Nagel, E. 1961. *The structure of science.* London: Routledge and Kegan Paul.

Nagel, E. 1998. Issues in the logic of reductive explanations. Reprinted in *Philosophy of science: The central issues*, edited by M. Curd and J. A. Cover. New York: Norton.

Nickles, T. 1998. Two concepts of intertheoretic reduction. Reprinted in *Philosophy of science: The central issues*, edited by M. Curd and J. A. Cover. New York: Norton.

Niiniluoto, I. 1999. *Critical scientific realism.* Oxford: Oxford University Press.

Penrose, R. 1997. The mysteries of quantum physics. In *The large, the small, and the human mind*, edited by R. Penrose. Cambridge: Cambridge University Press.

Poincaré, H. 1900. Les relations entre la physique expérimentale et la physique mathématique. *Rapports présentés au Congrés International de Physique 1900.* Paris.

Poincaré, H. 1902. *La science et l'hypothèse.* Reprint. Paris: Flammarion, 1968.
Poincaré, H. 1905. *La valeur de la science.* Reprint. Paris: Flammarion, 1970.
Psillos, S. 1999. *Scientific realism: How science tracks truth.* London: Routledge.
Putnam, H. 1983. Models and reality. In *Realism and reason*, edited by H. Putnam. Cambridge: Cambridge University Press.
Reiter, R. 1980. A logic for default reasoning. *Artificial Intelligence* 13: 8–132.
Ruttkamp, E. B. 2002. *A model-theoretic realist interpretation of science.* Dordrecht: Kluwer.
Ruttkamp, E. B. 2003. On truth and reference in postmodern science. *South African Journal of Philosophy* 22: 220–235.
Ruttkamp, E. B. Forthcoming. Over determination of theories by empirical models: A realist interpretation of empirical choices. In *Logics of scientific cognition: Essays in debate with Theo Kuipers*, edited by R. Festa. Amsterdam: Rodopi.
Schaffner, K. F. 1967. Approaches to reduction. *Philosophy of Science* 34: 137–47.
Shoham, Y. 1988. *Reasoning about change: Time and causation from the standpoint of artificial intelligence.* Cambridge, MA: MIT Press.
Sklar, L. 1967. Types of intertheoretic reduction. *British Journal for the Philosophy of Science* 18: 109–24.
Stegmüller, W. 1976. *The structure and dynamics of theories.* Berlin: Springer.
Suppes, P. 1957. *Introduction to logic.* New York: Van Nostrand.
Taylor, J. C. 2001. *Hidden unity in nature's laws.* Cambridge: Cambridge University Press.
Worrall, J. 1989. Structural realism: The best of both worlds? *Dialectica* 43: 99–124.
Worrall, J. 1990. Scientific revolutions and scientific rationality: The case of the 'elderly holdout'. In *Scientific theories*, edited by C. W. Savage. Minneapolis, MN: University of Minnesota Press.