A Non-classical Logical Foundation for Naturalised Realism

EMMA RUTTKAMP-BLOEM, GIOVANNI CASINI, AND THOMAS MEYER

Abstract: In this paper, by suggesting a formal representation of science based on recent advances in logic-based Artificial Intelligence (AI), we show how three serious concerns around the realisation of traditional scientific realism (the theory/observation distinction, over-determination of theories by data, and theory revision) can be overcome such that traditional realism is given a new guise as 'naturalised'. We contend that such issues can be dealt with (in the context of scientific realism) by developing a formal representation of science based on the application of the following tools from Knowledge Representation: the family of Description Logics, an enrichment of classical logics via defeasible statements, and an application of the preferential interpretation of the approach to Belief Revision.

Keywords: scientific realism; theory-observation distinction; over-determination of theories by data; theory revision; scientific progress; description logics; preferential logics; belief revision

1 Introduction

In this paper, by suggesting a formal representation of science based on recent advances in logic-based AI, we show how three serious concerns around the realisation of traditional scientific realism can be overcome such that traditional realism is given a new guise as 'naturalised'. In the account of naturalised realism proposed by Ruttkamp-Bloem (2013), science does not advance in a teleological fashion towards 'the truth'. Instead, she suggests an epistemic definition of truth implying that the criterion for determining (realist) stances on the status of the content of science is the quality of evolutionary progressive interaction between the experimental and theoretical levels of science. The success of this process depends on the quality of available evidence. This (epistemological) analysis of science suggests a new interpretation of three well-known concerns about traditional scientific realism, which, in its turn, contributes to a firmer definition of naturalised realism as an epistemic realism vs. traditional metaphysical realism. The problems are: 1. Scientific realism should be able to deal with the messy issue of 'disentangling' empirical information and general statements and laws throughout the history of scientific investigations of a given real system, and with the added complication of the non-uniqueness of such disentanglements; 2. Theories should be robust in terms of the absorption of new empirical information in order to deal with the 'over-determination' of theories by data, akin to the definition of the equivalent empirical models of van Fraassen (1980); 3. Based on new empirical evidence, science is not prescriptive; it is not the case that there is a single 'correct' method for modifying an existing theory to incorporate new evidence. But science is also constraining; there ought to be limitations on what constitutes a possible rational modification to an existing theory.

We contend that these issues can be dealt with in a formal representation of science which is the result of applying the following tools from subareas of Knowledge Representation to the above philosophical issues in the context of scientific realism: The family of knowledge representation languages known as Description Logics (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2007), an enrichment of the languages of classical logics with defeasible statements in the style of the preferential framework suggested by Kraus, Lehmann, and Magidor (1990) for propositional logic, and an application of the preferential interpretation of the approach to Belief Revision proposed by Alchourrón, Gärdenfors, and Makinson (1985), and extended to account for Iterated Belief Revision. In what follows, we first give a brief description of 'naturalised realism' as the version of realism supported by our suggested treatment of the three concerns around the possibility of actualising traditional scientific realism identified above. We then briefly discuss each of these concerns in turn, showing as we move along how in each case either Description Logics, Non-monotonic Reasoning, or Belief Revision can positively address each of these concerns. To conclude, we argue that a single coherent framework for realism could be provided which incorporates all three aspects discussed above.

2 Naturalised realism

The version of realism presented here is called 'naturalised realism' and it is based on the following twelve tenets:

1. Science is about an *independently externally existing reality* in William James's sense of experience "boiling over" (*viz.* Chang, 2012), thus in the sense of the 'resis-

tance', or perhaps rather 'reaction', of nature to the application of theories as proof of the existence of the outside world; 2. Realist claims are compatible with *the nature and the history of science* in the full (naturalised) sense of reflecting the processes of science in its evaluations of the status of the content of theories; 3. The driver of scientific progress is *revision* (not retention), where retention is a special case of revision); 4. Scientific progress is *not linear or convergent* in the traditional sense; 5. The unit of appraisal for naturalised realism is a *network of systems of theories* (not single theories), where the notion of a 'system of theories' is close to the Kuhnian notion of a 'disciplinary matrix'; 6. Continuity in science is a meta-issue and is *methodological continuity*; 7. Naturalised realism is an *epistemology of science* (not a metaphysics); 8. The epistemological framework for naturalised realism is *fallibilism*; 9. The criterion for determining realist stances towards the status of the content of science is the quality of *'evolutionary progressive' interaction* between the experimental and theoretical levels of science; 10. An *epistemic account of truth as correspondence* is the best account for realism and here the account is unpacked in terms of *truth-as-method* ('method' here in the sense of the generally acceptable way in which scientists do research based on taking empirical evidence seriously, in short the Peircean 'experiental method'); 11. Relations of *'historied' reference* are epistemological trackers of what is revealed about nature as 'true' through the course of science (rather than existential claims with regards to the ontology of reality); 12. The dichotomy of the traditional scientific realism debate is collapsed into a *continuum of realist stances* towards the status of scientific theories.

The naturalised realist contends that whether realism is warranted with regards to particular scientific content depends most heavily on the kind and quality of evidence available for that content; she suggests that for the naturalised realist it makes sense to be a realist only about those aspects of scientific investigations that demonstrate actual science-world interaction, as such interaction is taken to be the source of establishing evidence for science's claims and the quality of such interaction is taken to determine the quality of evidence available for specific scientific claims at specific times.

This interaction is visible most importantly in the revisions science effects in its theories and experimental work as the result either of experimental feedback leading to theoretical adjustment or of theoretical adjustment necessitating revision at the level of experimental processes. Such revision is measured in terms of the 'evolutionary progressiveness' of theories.

A theory T is 'evolutionary progressive' at time t_n iff:

1. it is 'empirically (experimentally) adequate' according to experimental practices in the area of investigation at time t_n *in such a way that previous versions of*

theory T at time t_{n-1} have been adapted in significant ways in order to effect this adequacy.

2. it is 'theoretically adequate' in the sense that *theoretical descriptions made at t_{n-1} have been adapted* such that they describe or refer to (read: 'offer knowledge concerning') properties of observable and unobservable entities in the scope of the theory at time t_n.

The *naturalised realist criterion is the degree to which the definitions of reference and referential stability below are satisfied in a network of theories*, as 'refer' in their context means 'have evolutionary progressive knowledge of'. Evolutionary progressive science-world interaction is captured by *epistemic* relations of reference which define the context within which a theory could be 'true' at a given time, and which track the development of knowledge with regards to a particular event or phenomenon or target system as follows:

A term t 'refers' to a posited entity iff it satisfies a 'core causal description' (CCD) of 'identifying' properties associated with term t such that

1. the CCD in question has been adapted to fit the current experimental situation and thus describes properties currently thought to belong to the postulated entity and

2. the properties in question are such that the posited entity plays its putative causal role in virtue of these properties (i.e. these properties are the causal origin of claims associated with the putative entity) (*cf.* Psillos, 1999).

Reference for the naturalised realist is an epistemic issue and is thus viewed to be about tracking the development of knowledge claims concerning a particular target system, phenomenon or event, rather than about establishing the metaphysical existence of a real system, phenomenon or event. If it is true that the metaphysical import of successful theories would, in an ideal world, consist in their giving correct descriptions of the structure of the world, (*viz.* J. Ladyman, 2007), then the epistemological import of successful theories consists in their being the crystallisation of a process or method of continuous revision and sifting claims. As implied by the history of science, (*viz.* Chang, 2012; Laudan, 1981), the former option is not viable given the nature of science and therefore the naturalised realist suggests a new interpretation of 'reference' as the epistemic mechanism through which the revision and sifting of claims referred to above can be made concrete and can be tracked through the history of scientific investigation.

Returning to the definition above, 'identifying' properties are properties that can be described according to the current experimental situation – so 'core' properties are properties that have been revised, and are determined by current evidence. The noteworthy point here is that naturalised realists thus imply that what is 'core' can change, which is not a suggestion that would sit comfortably with traditional scientific realists' depiction of science. Let us now consider how the naturalised realist proposes to interpret and apply the mechanism of 'reference' in times of theory change:

Terms t and t' denote 'the same' posited entity within the same theoretical system iff

1. both t and t' each respectively satisfies a CCD of properties associated with them that has been adapted to fit the experimental situations in which the theories containing t and t' respectively have been formulated; and

2. the description of the properties in the CCD of t' has been adapted from the CCD of t; and

3. the referents of t' and t play the same causal role w.r.t. a certain set of phenomena in virtue of the properties described in their respective CCD's.

The issue is not one of identifying limiting cases of successor theories or the parts of theories that 'persevere' through theory change, but rather the possibility of finding *heuristic continuity* via revisions culminating in evolutionary progressive theories, which, *in virtue of their revision*, carry on in (heuristic/epistemic/pragmatic) continuity with their predecessor theories. In this sense, truth is assembled as science progresses through revisions and confirmations, and the content of what is *assembled* is captured or revealed by relations of reference supervening on the progressiveness of theories.

Considering the technical aspects of the definition above, it is necessary to specify that reference is reference to the 'same' posited entity *within the same theoretical system*, as the unit for naturalised realist appraisal is a network of theories, i.e. the collection of all investigations of a particular target system, phenomenon or event over time. And, in this (broader) context, obviously it may be the case that not all descriptions of the same posited entity have been adapted from previous descriptions, as there is the possibility of incompatible descriptions of the same postulated entity – e.g. Thomson, Lorentz, Bohr, Millikan, *et al.* on the properties of electrons – given that the network of investigations being evaluated may include more than one (in/compatible) theoretical system focused on the particular phenomenon or event at issue (e.g. the phlogiston system of theories vs. the oxygen system of theories). (For our purposes here, think of a system of theories as broadly a Kuhnian paradigm in the sense of a disciplinary matrix.) Thus the reference relations of every separate theoretical '*genre*' or system of theories must all be taken into account when a realist decision is made regarding the appropriate epistemic stance to adopt towards the content of knowledge concerning a particular real system, phenomenon or event at a given time.

In conclusion, note that naturalised realism is an epistemic realism which means at the most simple level simply that naturalised realists believe scientific knowledge is knowledge about the external world. (This may or may not include the belief that the entities implied to exist according to the sum of current scientific knowledge, really exist. The epistemic realist can remain agnostic to a huge degree about such metaphysical issues.) Thus *the outcome of a naturalised realist evaluation of the content of a network of theories is a particular epistemic stance* towards the content of such a network, *on a continuum of realist stances* varying from epistemic equivalents to traditional instrumentalism to such equivalents to traditional scientific realism (e.g. think of the development of knowledge concerning certain viruses). This notion of a continuum captures the sense in which naturalised realism is *naturalised*, i.e. realism traces the movements of science, as science endeavours to understand the 'movements' of real phenomena and events.

3 The empirical-theoretical distinction re-visited

Ernst Nagel (1961) offers one of the most well-known distinctions between so-called 'experimental laws' and 'proper theories'. In his sense 'experimental laws' contain only so-called observational terms, while the purpose of the formulation of 'proper theories' is to explain 'experimental laws' by the theoretical terms they introduce. This is part of the old so-called 'two-language' depiction of the positivist syntax of science of which Rudolf Carnap's (1956) *The Methodological Character of Theoretical Concepts* is the best exposition. This view, according to which science is interpreted by "relating it to an observation language (a postulated part of natural language which is devoid of theoretical terms)" (van Fraassen,1980, pp. 13–14), has now totally run its course. The debate over the distinction between theoretical and observational (or empirical in contemporary terms) rages on though, albeit in different format, and one of its most interesting playing fields is the area of scientific realism.

To see why this is so, let us briefly consider one other major contribution to the debate about the viability of the two-language distinction, namely Grover Maxwell's (1962) *The Ontological Status of Theoretical Entities* which is directly in opposition to Carnap's views and meant to be an attack on the instrumentalism of the positivist account of science. Maxwell (1962, p. 1052) starts off by quoting a statement made by the other stalwart of positivism, Ernst Nagel (1961), that the distinction between realism and instrumentalism is simply a "conflict over preferred modes of speech" (Chapter 6). Most of what Maxwell offers in support of realism turns on this remark of Nagel's, as Maxwell implies that what should be taken into account in terms of the dichotomy postulated by the positivists is that while it must be acknowledged that the 'life' of a theoretical term often starts out as a mere "heuristic crutch"(Maxwell, 1962, p. 1053), very often its life ends as an 'observation' term – e.g. the history of the discovery of 'microbes', (*viz.* Maxwell, 1962, pp. 1053–1055). Thus, against van Fraassen's definition of 'obervable' as "an unaided act of perception"(van Fraassen, 1980, p. 15), Maxwell (1962, p. 1056) argues for a view that acknowledges that "...there is, in principle, *a continuous series* [our italics] beginning with looking through a vacuum and containing these as members: looking through a windowpane, looking through glasses, looking through binoculars, looking through a low-power microscope, looking through a high-power microscope, etc., in the order given". He (ibid.) concludes from this that accepting such a 'continuous series' of candidates for 'being observable', implies that there is no "non-arbitrary line between 'observation' and 'theory' ".

The naturalised realist, while taking that 'observable' is not equated with 'existence' as Maxwell (1962, p. 1057) insists, accepts this arbitrariness given her focus on the revisable and shifting nature of evidence for current scientific theories. She proposes that Maxwell's continuous series of observability suggests an epistemic (not a metaphysical) series reflecting a continuum of epistemic (realist) stances towards the content of theories assembled at a given time. More specifically, as noted above, the outcome of a naturalised realist evaluation of the content of a network of theories is a particular epistemic stance towards the content of such a network, *on a continuum of realist stances* varying from epistemic equivalents to traditional instrumentalism to such equivalents to traditional scientific realism.

The real challenge for realism, rather than focusing on the separate parts/ stances making up the continuum, actually has to do with explaining how such a continuum might work *as a continuum*, hence proposing a workable suggestion to deal with the arbitrariness of theory-observation distinctions. This is where a representation of scientific knowledge based on Description Logics can assist.

Description Logics (or DLs) are a family of knowledge representation formalisms (Baader et al., 2007), corresponding to decidable fragments of first-order logic. They are based on the notions of concepts (unary predicates in first-order logic terms) and roles (binary relations in first-order logic terms), and are mainly characterised by constructors that allow complex concepts and roles to be built from atomic ones. The DLs of interest to us are all based on an extension of the well-known DL $\mathcal{ALC}$. Concept descriptions are built from concept names using the constructors disjunction ($C \sqcup D$), conjunction ($C \sqcap D$), negation ($\neg C$), existential restriction ($\exists r.C$) and value restriction ($\forall r.C$), where C, D represent concepts and r represents a role name.

A DL knowledge base κ consists of two finite and mutually disjoint sets. A *terminology box* (or *TBox*) which introduces the *terminology*, and an *assertion box* (or *ABox*) which contains facts about particular objects in the application domain. TBox statements have the form $C \sqsubseteq D$ (*inclusions*) where C and D are (possibly complex) concept descriptions. Objects in the ABox are referred to by a finite number of *individual names* and these names may be used in two types of assertional statements: *concept assertions* of the type $C(a)$ and *role assertions* of the type $r(a, b)$, where C is a concept description, r is a role name, and a and b are individual names.

The semantics of $\mathcal{ALC}$ is the standard set-theoretic Tarskian semantics. Individual names are interpreted as elements of a domain of interest, concepts as subsets of this domain, and roles as binary relations over this domain. An interpretation I consists of a non-empty set Δ^I (the domain of I), an injective *denotation function* d which maps every individual name a to an element a^I of the domain Δ^I, a function $\cdot^I$ (the *interpretation function* of I) which maps every concept name A to a subset A^I of Δ^I, and every role name r to a subset r^I of $\Delta^I \times \Delta^I$. The interpretation function is extended to arbitrary concept descriptions as follows. Let C, D be concept descriptions and r a role name, and assume that C^I and D^I are already defined. Then

$(\neg C)^I = \Delta^I \setminus C^I, (C \sqcup D)^I = C^I \cup D^I, (C \sqcap D)^I = C^I \cap D^I;$
$(\exists r.C)^I = \{x \mid \exists y \text{ s.t. } (x, y) \in r^I \text{ and } y \in C^I\};$
$(\forall r.C)^I = \{x \mid \forall y, (x, y) \in r^I \text{ implies } y \in C^I\}.$

An interpretation I *satisfies* $C \sqsubseteq D$ iff $C^I \subseteq D^I$. Similarly for ABox assertions, an interpretation I *satisfies* the assertion $C(a)$ iff $a^I \in C^I$, and it *satisfies* $r(a, b)$ iff $(a^I, b^I) \in r^I$. I is said to be a *model* of a DL (TBox or ABox) statement ϕ iff it satisfies the statement. I is said to be a model of a DL knowledge base κ iff it satisfies every statement in κ. A DL knowledge base κ *entails* a DL statement ϕ, written as $\kappa \vDash \phi$, iff every model of κ is a model of ϕ. With the formal apparatus in place, our proposal is to represent scientific theories as Description Logic knowledge bases, with the TBox of a knowledge base containing a description of the theory, and the ABox containing the accumulated empirical facts. The most relevant aspect of the structure of DL knowledge bases from the perspective of the naturalised realist is the explicit distinction drawn between statements in the TBox and the ABox. This dichotomy provides a mechanism for a formal distinction between theoretical and observational terms. The intended use of the ABox is the provision of formal empirical facts about the world, so within a DL knowledge base, the distinction between theoretical and observational terms can be formally enforced by requiring that only observational terms may be present in the ABox. The distinction between theoretical and observational terms is not *determined* by the use of DLs: it provides a framework within which the arbitrariness of the distinction between theoretical and observational terms can be formalised.

4 Over-determination of theories by data

Imagine a typical semantic account of theories, (e.g. Suppes, 1967; van Fraassen, 1980), according to which theories are depicted in terms of sets of mathematical structures that are the models of the theory in question. Some commentators on this school of thought, (e.g. Kuipers, 2001; Ruttkamp, 2002), further identify empirical reducts of the models in question as follows. Consider what it really means to formulate a model of a particular theory. 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. Now, focusing on a particular real system at issue in the context of applying a theory, which in turn implies a specific empirical set-up in terms of the measurable quantities of that particular real system, the predicates in the mathematical model of the theory under consideration that may be termed ‘empirical’ predicates may be selected. This is how an empirical 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.

The method of (‘empirical’) verification of each of the set of models depicting the theory (i.e. how well do each of them reflect/represent/link to the relevant system in the real world?), is decided by the specific nature of the specific model in question, *as well as* by the nature of the specific real system in question. Hence if the phenomena in some real system and the experimental data concerned with those phenomena are logically reconstructed in terms of a mathematical structure – call it an 'empirical' or 'data' model – then the relation of empirical adequacy becomes a relation which is an isomorphism from the empirical model into an empirical reduct of a relevant model of the theory in question.[1] To summarise: In the empirical reduct are interpreted only the terms called 'empirical' in the particular relevant context of application or empirical situation. Think of this substructure of the model at issue 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.

But what is the status of the models or empirical reducts or empirical models not chosen at a specific time, then? The knowledge or information about the particular empirical model(s) in question that they carry, certainly still *is* knowledge, is it not? Well, yes and no. What is needed is a formal mechanism by which to depict choices, the motivations for choices, and the change of both of these, should there be a change of context within which a theory is being applied. A decision is taken to work with a certain model or empirical reduct at a certain time, but, on the one hand, a different choice/decision based on new information can be made at any time, or on the other hand, what are deemed to be 'empirical predicates' can change according to new knowledge and thus the formulation of empirical reducts can change, (*cf.* Ruttkamp, 2005). For instance, the general theory of relativity was formulated by Einstein (and Hilbert) in 1915. For almost a century now physicists have been constructing literally dozens of different types of models – all models of precisely the same theory – to fit both experimental and observational data about the space-time structure of the real universe *and* certain paradigmatic preferences. This illustrates (perhaps in a different way from what he (Kuhn, 1996) intended), Kuhn's point that neither the content of science nor any system in reality should be claimed to be "uniquely exemplified" by scientific theories from the viewpoint of studies of "finished scientific achievements". In this sense the 'open-endedness' of science can be taken to imply that the terms of an already established the-

[1] Recall that according to van Fraassen (1976, p. 631) a theory is empirically adequate if "all appearances are isomorphic to empirical substructures in at least one of its models".

ory are 'about' an on-going potential of entities in some system of reality to relate to some objects and relations in *any* model of that theory. The actualization of this potential requires human action in the sense of finding and finally articulating 'satisfying' relations between aspects of systems in reality and certain empirical aspects (reducts) of models of the theory at issue. It is in the process of establishing such relations that one becomes aware of 'empirical proliferation' in the sense of 'over-determination of theories by data'. In a sense this is the reverse of the traditional scenario of the underdetermination of theories by data. [2]

Scientific knowledge is amendable and even defeasible, because of its contingent and particularized links with the reality it describes (and explains). In general scientific theories, depicted as syntactic (linguistic) entities that need to be interpreted to be given semantic meaning and reference, are not able to uniquely capture their semantic content. In terms of theory application, taking now both the syntactic and semantic accounts of theories into account, within a model-theoretic context, two sets of relations are conducive to empirical proliferation: the set of relations between the terms in some theory and their extensions in its various models; and the set of relations between the terms of models (or of only one model) – via an empirical reduct (or empirical reducts) of that (those) model(s) – and the objects and relations of some real system (or systems) conceptualized in one or many empirical models. Retaining the notion of scientific theories as linguistic expressions at the 'top' level of science solves the problems regarding the justification of the existence of many (conceptual) models as interpretations of any one theory by the simple (formal) fact of the incompleteness of formal languages. Thus the possibility of a given scientific theory being interpreted in more than one mathematical model (structure) is natural in a very basic sense in model-theoretic terms. The second proliferation of relations between models and their empirical reducts and between these and empirical models may also turn out to be less counterintuitive than might be supposed at first glance, if it can be shown that the possibility of articulating rational reasons for choosing one relation among the others at any given time is *not* jeopardized under such circumstances.

In what follows we claim in particular that an application of *defeasible* reasoning to situations of over-determination of theories by models and data may enable formalising this second kind of complexity in terms of a particu-

[2]The context of this discussion is empirical equivalence in van Fraassen's sense (1980, p. 67).

lar kind of preferential ranking of these models (*Cf.* Ruttkamp, 2005). In this context, let us consider the possibility of introducing into the wide empirical equivalence debate, concentrated on issues concerning over-determination of theories by data, the non-monotonic mechanism of defeasible reasoning, refined into a model-theoretic non-monotonic logic based on the preferential logics of Lehmann and Magidor (1992) offering a formal method to rank models. To enable defeasible reasoning, the language of classical propositional logic is enriched with a notion of *defeasible implication*, represented with the connective $\mid\!\sim$, with statements of the form $\alpha \mid\!\sim \beta$ informally intended to represent the information that β normally follows (defeasibly) from α. The intention is to exploit the ability to represent defeasible information in order to deal with the over-determination of theories.

For example suppose that our theory contains the information that mammalian and avian red blood cells are vertebrate red blood cells ($\mathsf{MRBC} \rightarrow \mathsf{VRBC}$, $\mathsf{ARBC} \rightarrow \mathsf{VRBC}$), that vertebrate red blood cells have a nucleus ($\mathsf{VRBC} \rightarrow \mathsf{hasN}$), but that mammalian red blood cells don't ($\mathsf{MRBC} \rightarrow \neg\mathsf{hasN}$). Notice that mammalian red blood cells, by virtue of being vertebrate red blood cells, are also said to have a nucleus ($\mathsf{MRBC} \rightarrow \mathsf{hasN}$). So a consequence of the theory is that mammalian red blood cells have, and don't have a nucleus – a classical case of over-determination.

The proposal we are advancing here is that a language with defeasible implication allows for a more appropriate and accurate expression of scientific theories. In the case of the example above a more accurate theory would be one in which it is stated that mammalian and avian red blood cells are vertebrate red blood cells ($\mathsf{MRBC} \rightarrow \mathsf{VRBC}$, $\mathsf{ARBC} \rightarrow \mathsf{VRBC}$), as before, but that vertebrate red blood cells *normally*, but not necessarily, have a nucleus ($\mathsf{VRBC} \mid\!\sim \mathsf{hasN}$), and that mammalian red blood cells normally don't ($\mathsf{MRBC} \mid\!\sim \neg\mathsf{hasN}$). Furthermore, an appropriate notion of entailment for such a richer language would correctly identify mammalian red blood cells as *exceptional* vertebrate red blood cells, therefore not conforming to the default property of vertebrate red blood cells having a nucleus, and in doing so, avoid any over-determination.

It turns out that suitable model-theoretic notions of entailment can be obtained by employing a form of preferential semantics. Recall that the semantics of propositional logic is based on the set of propositional valuations V, where a valuation v is a function which assigns a truth value (true or false) to every propositional atom from which the language is generated. The preferential semantics for defeasible reasoning proposed by Lehmann and colleagues (Kraus et al., 1990; Lehmann & Magidor, 1992) involves the establishment of an ordering $\prec$ over valuations,[3] with valuations lower down in the ordering viewed as more plausible. That is, $v \prec w$ means that we consider the situation described by the valuation v as more plausible than the situation described by the valuation w; a *preferential model* $\mathcal{M}$ consists of a set of valuations and an ordering, on which we can impose the satisfaction of specific properties. Preferential models can be used to define the semantics of classical propositional formulas, as well as defeasible implications: let $\hat{\alpha}$ be the set of the valuations in $\mathcal{M}$ that satisfy α, $\mathcal{M}$ satisfies $\alpha \mathrel{|\!\sim} \beta$ iff $\min_{\prec}(\hat{\alpha}) \subseteq \hat{\beta}$, that is, all the most plausible valuations satisfying α satisfy also β. Lehmann and Magidor (1992) propose that *rational* forms of entailment for defeasible reasoning be captured in terms of a specific class of models. Given a knowledge base κ expressed in a propositional language enriched with defeasible implication, their proposal is that the set of formulas entailed by κ ought to be exactly those that are satisfied in one of the ranked models satisfying all elements of κ. Since, for most knowledge bases κ, there would be more than one ranked model satisfying it, there may well be more than one appropriate form of entailment as well. A detailed investigation into different forms of entailment is beyond the scope of this paper, but the point is that defeasible reasoning, and more specifically, the introduction of defeasible implication based on a preferential model-theoretic semantics, can provide a successful formal mechanism for managing the problem of over-determination in logic-based representation of scientific theories.

5 Revision

Returning now to naturalised realism, is it a case of anything goes? All of science is about revision after all, so are naturalised realists, realists about all of science? The issue for the naturalised realist is not simply noting that theories are revised in the course of science, it is *about determining what scientists learn from revision*, as the naturalised realist is specifically interested in revision that is 'evolutionary progressive'. Thus the naturalised realist suggests that scientific progress is the result of the available network of knowledge claims becoming more refined as scientists learn from their 'mistakes' and *adapt* their theories such that they *show how their theories can accommodate revision* (note that maintaining the *status quo* when no revision is necessary in the face of new evidence, is recognised as a special

[3] Strictly speaking, an ordering over states, where states are mapped to valuations.

case of 'revision'). Continuity lies in the method of science – thus in revision, in the epistemic sense that the quality of knowledge claims depends for ever on the quality of evidence available at the time. In this sense perhaps theories that have been most revised (e.g. phlogiston) are theories with the most realist potential.

This is not in contradiction with the fact that no such thing as phlogiston exists, because naturalised realism expands the scope of realism from being tightly (almost exclusively) focused on metaphysical issues to allowing an emphasis on the epistemological aspects of theory content evaluation. A theory that can absorb revision (thus which is evolutionary progressive) is much stronger than one that is true in all possible worlds, because an evolutionary progressive theory contains much more knowledge of the system (or phenomenon or event) it describes than a tautology does, in the sense that it reflects knowledge of what should be left out of descriptions (based on current evidence), while a theory that is always true does not reflect such knowledge. Naturalised realism is first and foremost an epistemology of science, not a metaphysics of science.

All investigations of a phenomenon or event over time inform our current knowledge of it. The *unit of (naturalised) realist appraisal is networks of (systems of) theories*, not single theories. The more revision there is, the better movements within such a network of theories can be tracked and interpreted and the better the strength of the current realist stance towards knowledge of the event or phenomenon at issue can be evaluated. We suggest *belief revision* as a method through which to demonstrate these aspects of a naturalised realist definition of progress in science in terms of revision rather than retention.[4]

Generally, belief revision is concerned with the rational management of changes to an agent's information about the world it inhabits. More specifically, it aims to describe how an agent ought to change its mind when confronted with new, and possibly conflicting, information to be admitted into its knowledge about the world. To incorporate belief revision into the methodology put forward by the naturalised realist, it is best to consider the approach to revision advanced by Alchourrón, Gärdenfors, and Makinson, the so-called AGM approach (1985). Technically, AGM belief revision is applicable to any logic with a standard Tarskian semantics for which the

[4] A related issue is the case of theory change, e.g. Kepler to Newton, Newton to Einstein, or the 18^{th} and 19^{th} century ether theories of light to electromagnetism. Coping with such examples demands a more specific extension of vocabulary (languages) which is somewhat outside the scope of this paper, but which we address in a series of forthcoming works.

Compactness Theorem holds, but we shall make matters concrete here by focusing on propositional logic. A knowledge base in this context is represented as a set κ of propositional formulas. For the naturalised realist κ would therefore represent a scientific theory. With the knowledge base κ fixed, a revision operator $\circ$ takes as input a propositional formula to be incorporated into the current knowledge base and produces a new knowledge base (a set of propositional formulas).

The most important aspect of AGM revision is that it does not prescribe a unique method for revising a knowledge base. Instead, it provides the following set of postulates to which any *rational* revision operator $\circ$ should conform:

(K1) $\kappa \circ \alpha = Cn(\kappa \circ \alpha)$;
(K2) $\alpha \in Cn(\kappa \circ \alpha)$;
(K3) $Cn(\kappa \circ \alpha) \subseteq Cn(\kappa \cup \{\alpha\})$;
(K4) If $\kappa \cup \{\alpha\} \not\models \bot$ then $\kappa \circ \alpha = Cn(\kappa \cup \{\alpha\})$;
(K5) If $\kappa \circ \alpha \models \bot$ then $\alpha \models \bot$;
(K6) If $\alpha \equiv \beta$ then $Cn(\kappa \circ \alpha) = Cn(\kappa \circ \beta)$;
(K7) $Cn(\kappa \circ (\alpha \wedge \beta)) \subseteq Cn(\kappa \circ \alpha \cup \{\beta\})$;
(K8) If $(\kappa \circ \alpha) \cup \{\beta\} \not\models \bot$
then $Cn((\kappa \circ \alpha) \cup \{\beta\}) \subseteq Cn(\kappa \circ (\alpha \wedge \beta))$

The significance of these postulates have been discussed at length (e.g., by Gärdenfors (1988)), and we won't do so here. For the naturalised realist this approach to revision is important because revision need not be unique, thereby allowing for the development of systems of theories, all conforming to the eight AGM revision postulates. At the same time, revision is not naïvely open-ended either. The constraints imposed by the eight postulates are suitably restrictive, ensuring ample justification for the maintenance of a system of theories. AGM revision provides a formal framework for developing a system of theories, but since it only describes a mechanism for a single revision step, it does not address the issue of the maintenance of a *network* of systems of theories. For that it is necessary to consider the area of *iterated belief revision* which seeks to describe the dynamics of a sequence of belief revision steps. A detailed description of iterated revision is beyond the scope of this paper. It will suffice to remark that the most successful approach (Darwiche & Pearl, 1997) involves an extension to AGM revision, and allows for the construction of a network of results in line with the naturalised realist stance.

6 Conclusion

We have argued that DL can possibly shed new light on the formal representation of the empirical-theoretical distinction, that Non-monotonic Logic in the form of defeasible implication can be employed to manage the over-determination of scientific theories, and that Belief Revision can be used as a mechanism for representing the naturalised realist approach to progress in science. A single unifying framework for doing so would involve the incorporation of Non-monotonic logic and Belief Revision into Description Logics. In recent years, significant progress has been made in this regard, with Description Logics being extended to include notions of *defeasible subsumption* (Giordano, Gliozzi, Olivetti, & Pozzato, 2013) and Belief Revision (Meyer, Lee, & Booth, 2005). What is still missing for a truly integrated framework is a system in which Belief Revision can be applied to Description Logics that are enriched with defeasible subsumption. We are currently investigating this.

References

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic*, *50*, 510–530.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (2007). *The DL Handbook* (2nd ed.). Cambridge Univ. Press.

Carnap, R. (1956). The Methodological Character of Theoretical Concepts. *Minnesota Studies in the Philosophy of Science*, *1*, 38–76.

Chang, H. (2012). *Is Water H2O? Evidence, Realism and Pluralism*. Springer.

Darwiche, A., & Pearl, J. (1997). On the Logic of Iterated Belief Revision. *Art. Intell.*, *89*(1-2), 1–29.

Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press.

Giordano, L., Gliozzi, V., Olivetti, N., & Pozzato, G. L. (2013). Rational Closure for Description Logics of Typicality. In *IAF-13* (pp. 1–9).

J. Ladyman, D. R. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford University Press.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Art. Intell.*, *44*, 167-207.

Kuipers, T. A. F. (2001). Epistemological Positions in the Light of Truth Approximation. *Proc. of* 20^{th} *World Congr. of Philos.*, *2001*, 79–88.
Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science*, *48*(1), 19–49.
Lehmann, D., & Magidor, M. (1992). What does a conditional knowledge base entail? *Art. Intell.*, *55*, 1-60.
Maxwell, G. (1962). The ontological status of theoretical entities. *Minnesota Studies in the Philosophy of Science*, *3*, 181–192.
Meyer, T., Lee, K., & Booth, R. (2005). Knowledge integration for description logics. In *Proc. of AAAI-05* (p. 645-650).
Nagel, E. (1961). *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace & World.
Psillos, S. (1999). *Scientific Realism:How Science Tracks Truth*. Routledge.
Ruttkamp, E. (2002). *A Model-Theoretic Realist Interpretation of Science*. Springer.
Ruttkamp, E. (2005). Overdetermination of Theories by Empirical Models: A Realist Interpretation of Empirical Choices. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *84*(1), 409–436.
Ruttkamp-Bloem, E. (2013). Re-Enchanting Realism in Debate with Kyle Stanford. *Journal for General Phil. of Sc.*, *44*(1), 201–224.
Suppes, P. (1967). What is a scientific theory? In S. Morgenbesser (Ed.), *Philosophy of science today* (pp. 55–67). Basic Books.
van Fraassen, B. (1976). To Save the Phenomena. *J. of Phil.*, *73*, 623–632.
van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press.

Emma Ruttkamp-Bloem
University of Pretoria
South Africa
E-mail: emma.ruttkamp-bloem@up.ac.za

Giovanni Casini
University of Pretoria & CAIR (CSIR-Meraka)
South Africa
E-mail: giovanni.casini@gmail.com

Thomas Meyer
CAIR (CSIR-Meraka) & University of KwaZulu-Natal
South Africa
E-mail: tmeyer@csir.co.za