A Structuralist Theory of Belief Revision

Holger Andreas

April 1, 2011

Abstract

The present paper aims at a synthesis of belief revision theory with the Sneed formalism known as the structuralist theory of science. This synthesis is brought about by a dynamisation of classical structuralism, with an abductive inference rule and base generated revisions in the style of Rott (2001). The formalism of *prioritised default logic* (PDL) serves as the medium of the synthesis. Why seek to integrate the Sneed formalism into belief revision theory? With the hybrid system of the present investigation, a substantial simplification of the ranking information that is necessary to define revisions and contractions uniquely is achieved. This system is, furthermore, expressive enough to capture complex and non-trivial scientific examples. It is thus closely related to a novel research area within belief revision theory which addresses the dynamics of scientific knowledge.

Keywords: Abduction, Belief Bases, Belief Revision, Default Logic, Defeasible Reasoning, Epistemic Ranking, Structuralist Theory of Science.

1 Introduction

The present paper aims to integrate the Sneed formalism known as the structuralist theory of science into belief revision theory. Among other things, this integration allows for a substantial simplification of the ranking information that is necessary to define revisions and contractions in a unique manner. In classical belief revision theory, some form of ranking is needed that orders any item of the belief set. Standard concepts to introduce this ranking are the relation of epistemic entrenchment (Gärdenfors 1988) and Spohn's ordinal conditional ranking functions (Spohn 1988). In the hybrid system of the present paper, by contrast, it is only *theory-elements*, i.e., pieces of background theories, that

need to have a ranking. A clear statistical explanation in terms of successful applications can be given for this ranking. Our thesis is that epistemic ranking is an effect of theorising and hence requires, for it to be investigated, an analysis of how theorising governs our beliefs.

We will give an overview of how the Sneed formalism can be integrated into belief revision theory. In structuralism, one distinguishes between intended applications of theory-elements and theoretical extensions thereof. Of particular interest are those theoretical extensions that are models of their corresponding theory-element. Classical structuralism, because of its semantic orientation, does not have the resources to account for deductive and other forms of reasoning with the help of a scientific theory. It has been shown, however, how the structuralist representation scheme can be transformed into an axiomatic representation of scientific theories (Andreas 2010b).

A further step towards a dynamisation of classical structuralism is to introduce an abductive inference rule that allows one to infer **T**-theoretical extensions from intended applications of **T**. The defeasible nature of abductive reasoning is taken into account through using the formalism of default logic. This proves convenient to express that **T**-theoretical extensions are required to satisfy internal and external links to other theory-elements. The next step is to introduce a ranking of theory-elements relevant to the domain of logical reconstruction. The resulting system resembles the paraconsistent and nonmonotonic formalism that Brewka (1991, 1989) developed for inferences from a belief base and which is used in Rott's (2001) general definition of base generated revisions in the context of only defeasibly valid background theories. Hence, we can adopt, with only very minor modifications, Rott's definition of revisions and contractions for our system. This is the final step of the synthesis.

2 Base Revisions

Let us review briefly the concept of base revisions. A *belief base* H is a set of sentences being accepted, where H is not necessarily closed under the operation of logical consequence. For a *belief set* A, by contrast, logical closure is required, which means $A = Cn(A)$, where Cn is the operator of logical consequence in classical logic. Belief bases are intended to represent belief states insofar as all the accepted sentences and only these be inferable from the belief base plus some (optional) background theory.

There is no consensus in the literature about the logical form of a belief base. In fact, few attempts have been made to characterise belief bases by formal,

logical means. Rott (2001) and Hansson (1999) introduce the notion of a belief base such that it contains only non-derived beliefs. A sentence ϕ is thus an element of the belief base H if and only if ϕ is non-derived and accepted. The investigation of base revisions with the intent of an axiomatic characterisation by postulates has been originated by Sven Ove Hansson.

A belief base H may be joined with a set E of axioms belonging to some background theory. This strategy has been studied by Rott (2001) and Brewka (1991) in a systematic way. There, the axioms of background theories need only be defeasibly valid. They are simply called *expectations*, which gives rise to the use of E as the symbol for the set of axioms of potentially relevant background theories. Hans Rott's investigation of base revisions in the context of expectations will prove highly useful for the present attempt at an integration of the structuralist framework into belief revision theory.

Why prefer the study of base generated belief changes to investigating changes of belief sets? One important reason for this is that belief sets will always be infinite, whereas belief bases are usually finite. Therefore, the model of a belief base does respect the finiteness of human beings and computers. No human being is able to be aware of an infinite set of sentences. Nor does a computer have the storage for such an infinite set. Of equal importance seems to be that base revisions do respect justifications more properly than existing formalisms for belief set revisions (Brewka 1991, p. 208).

3 Reliability and Epistemic Entrenchment

Readers of Gärdenfors (1988) will remember the introducing story about belief revisions:

> Oscar used to believe that he had given Victoria a gold ring at their wedding. He had bought their two rings at a jeweller's shop in Casablanca. He thought it was a bargain. The merchant had claimed that the rings were made of 24 carat gold. They certainly looked like gold, but to be on the safe side Oscar had taken the rings to the jeweller next door who has testified to their gold content. However, some time after the wedding, Oscar was repairing his boat and he noticed that the sulphuric acid he was using stained his ring. He remembered from his school chemistry that the only acid that affected gold was aqua regia. Somewhat surprised, he verified that

> Victoria's ring was also stained by the acid. So Oscar had to revise his beliefs because they entailed an inconsistency.

In the formal theory, the symbolic notation for a revision is $A * \phi$, where A is the body of presently accepted beliefs and ϕ a new incoming information that is potentially inconsistent with A. More precisely, $A * \phi$ denotes the result of a revision of A by ϕ, that is, the set of sentences accepted after the revision. $A * \phi$ is required to be consistent unless ϕ is a logical falsehood.

In the example, it should be obvious which propositions are the elements of A. The proposition that the ring is stained by exposure to sulphuric acid is the new incoming information ϕ.[1] Obviously, ϕ is inconsistent with A in the example. The question thus arises: which sentences of A should be retained? This question is answered in the formal theory of belief revision through invoking an order $\leq_E$ of *epistemic entrenchment* among the propositions of A. Given that there is such an order, it can be determined in a straightforward manner which sentences remain accepted in the case of a contraction $A \div \alpha$ (Gärdenfors and Makinson 1988, p. 89):

$$\beta \in A \div \alpha \text{ if and only if } \beta \in A \text{ and either } \alpha <_E (\alpha \vee \beta) \text{ or } \alpha \in Cn(\emptyset) \quad (1)$$

In words: A sentence β remains to be accepted in A contracted by α if and only if β is accepted in the original belief set A and either α is strictly less entrenched than $\alpha \vee \beta$ or α is a logical truth. ($\alpha <_E \beta$ if and only if $\alpha \leq_E \beta$ but not $\beta \leq_E \alpha$.) By means of the Levi-identity, which says that $A * \phi = (A \div \neg\phi) + \phi$, (1) can be used to define $A * \phi$ in a unique way, where an expansion $+$ is defined as $A + \phi = Cn(A \cup \{\phi\})$.

That a proposition α is more entrenched than another proposition β implies that we are more reluctant to give up α than to give up β in the case of inconsistencies. This "explanation", however, has been found insufficient for a conceptual understanding of epistemic entrenchment since the view is rather that our hesitation to give up a proposition derives from its standing in the epistemic entrenchment order (Gärdenfors 1988, p. 87). Attempts at a conceptual explanation refer to the utility of a proposition in explaining other ones, its information-theoretic content, paradigms in the sense of Kuhn, and probabilistic

[1] Strictly speaking, it is sentences and not propositions that are the elements of A. As long as we are not dealing with a formalised example, however, it is hardly possible to avoid reference to propositions. If the sentences in a belief set are associated with an interpretation, no harm arises from saying that a proposition is the element of a belief set.

considerations. These efforts, however, did not result in a widely accepted, systematic, and quantitative account of epistemic entrenchment so that this notion remained a primitive one in belief revision theory.[2] Likewise, the firmness of a belief remained a primitive notion in Spohn's ordinal conditional ranking functions (1988), which serve the same purpose as epistemic entrenchment orderings but are more powerful when it comes to iterated revisions.

Now, here is the informal discussion of the example by Gärdenfors (1988):

> He could not deny that the rings were stained. He toyed with the idea that, by accident, he had bought aqua regia rather than sulphuric acid, but he soon gave up this idea. So, because he had greater confidence in what he was taught in chemistry than in his own smartness, Oscar somewhat downheartedly accepted that the rings were not made of gold after all.

The discussion is focused on whether or not Oscar shall retain the belief that the ring is made of gold. It appears that this belief has no independent standing in the epistemic entrenchment ordering. Oscar's readiness to give it up rather derives from the justifications and refutations he has for the proposition that the ring is made of gold. He gives the result of the simple, accidental chemical test, which clearly refutes this proposition, more credit than the justification based on the testimony of the jewellers. In other words, the chemical test is considered more reliable than the testimony of the jewellers. Why is this? Presumably, there is no case known to Oscar where gold was stained by exposure to sulphuric acid, whereas Oscar knows that it happens from time to time that imitations of gold are sold without proper declaration. Therefore, the reliability order used by Oscar can, at least partially, be explained by the ratio of the number of successful applications to the total number of successful and unsuccessful applications of the inferential patterns he is considering to use as justification or refutation of a proposition. An application of an inferential pattern is successful if and only if its conclusion has not been withdrawn until the present time. Otherwise, it is unsuccessful.

Goldman (1979) defines the concept of *reliability* as the tendency of a belief forming process to produce beliefs that are true rather than false. The term *tendency* could, Goldman explains, either refer to actual long run frequency of truth versus error or to a propensity being determined through the outcomes that would occur in merely possible realisations of the process. These ideas are

[2] With the notable exception of Rott (2001), who gives an explanation of the relation of epistemic entrenchment in terms of rational choice theory.

adopted here, with the qualification that reliability is not evaluated in terms of truth but in terms of acceptance. Those inferential patterns are reliable whose conclusions remain accepted. Only actual applications of inferential patterns are considered for the estimation of reliability but no potential ones.

The story about Oscar's wedding ring suggests a formalisation in terms of base generated revisions rather than in terms of belief set revisions. There seems to be no difficulty in applying the division of beliefs into derived and non-derived ones and expectations. For example, that he was told by the jeweller that the ring is made of gold is a non-derived belief held by Oscar.[3] By contrast, that the ring is made of gold belonged to his derived beliefs. Finally, that gold is not stained by sulphuric acid is an expectation, or a general inferential pattern in use for justification and refutation. As a conceptual clarification of the division of beliefs into derived and non-derived ones and expectations requires no less than a whole research paper, let us be content with noting that the division is applicable in the present example.

A further observation concerning the example can now be made. It appears that no ranking information concerning the non-derived beliefs is needed to account for Oscar's revision considerations. That the ring is stained seems to be accepted as firmly as the sentence saying that Oscar was told that the ring is made of gold by the jeweller.

Here is a formalisation of the inferential patterns being considered in the story:

(e1) $\forall x(LOOKS\_GOLDEN(x) \rightarrow MADE\_OF\_GOLD(x))$

(e2) $\forall x \forall y(JEWELLER(x) \wedge SELLS(x,y) \wedge TESTIFIES\_GOLD(x,y) \rightarrow MADE\_OF\_GOLD(y))$

(e3) $\forall x(\neg EXPENSIVE(x) \rightarrow \neg MADE\_OF\_GOLD(x))$

(e4) $\forall x \forall y(JEWELLER(x) \wedge \neg SELLS(x,y) \wedge TESTIFIES\_GOLD(x,y) \rightarrow MADE\_OF\_GOLD(y))$

(e5) $\forall x \forall y \forall z(RING(x) \wedge SULPHURIC\_ACID(y) \wedge TIME\_POINT(z) \wedge EXPOSED(x,y,z) \wedge STAINED(x,z) \rightarrow \neg MADE\_OF\_GOLD(x))$

[3]Of course, when Oscar starts theorising about the reliability of his memory, this belief may cease to be non-derived. The division of beliefs into derived and non-derived ones need not be static.

The meanings of the symbols used should be self-explanatory so that the meaning of the axioms should be obvious from the story.

Now, the only ranking information needed to account for Oscar's revision in the story is an order of reliability of the general axioms. For example, the following strict total order among the axioms may be assumed for the story:

$$e5 > e4 > e3 > e2 > e1$$

Assuming a strict total order for the reliability of the axioms seems neither adequate nor necessary, however. It appears more appropriate to say that the following information about the order of axioms is actually used in the story:

$$e5 > e4, e3, e2, e1$$

$$e4 > e3, e2, e1$$

Here, only a partial order among the axioms is assumed. The axiom saying that gold is not stained by exposure to sulphuric acid (e5) has the highest priority because it has not yet been found to have exceptions.

What lessons can be drawn from this analysis? Let us sum up some, still very conjectural results for our discussion of base generated revisions:

(i) No epistemic ranking of non-derived beliefs is needed to account for base generated revisions.

(ii) No explicit epistemic ranking of derived beliefs needs to be given either to account for revisions of such beliefs. Rather, conflicts at the level of derived beliefs can be resolved by considering the inferential patterns that were effective for the acceptance of conflicting derived beliefs. Only these inferential patterns need to have a ranking. This ranking represents their reliability.

(iii) The ranking of inferential patterns, that is, their reliability, can, at least in part, be explained in terms of statistical concepts, viz., by the frequency of their successful application.

These conjectures underlie the following systematic account of revisions by means of the structuralist framework. It should be noticed that the conjectures could also be incorporated into Rott's (2001, pp. 127–136) formalism for the direct mode of base generated revisions, which goes back to Brewka (1991, 1989).[4]

[4]Simply drop the priority information among the elements of the belief base and establish the ranking of expectations according to iii).

Why, then, do we aim to integrate the structuralist framework into belief revision theory? First, the use of set-theoretical predicates in structuralism proved more suited to represent complex examples from science than first-order systems without set-theory. The expressivity of structuralism has been appreciated even by researchers who thought formal means quite inappropriate for the analysis of science (Kuhn 1976). Second, in structuralism it is feasible to keep track of the justifications of derived beliefs since such beliefs are represented as valuations of **T**-theoretical extensions of intended applications. Elsewhere we will show that this feature allows for more efficient revision algorithms than those commonly used for first-order belief sets. Third, substantial commonalities between *frames* in the sense of Minsky (1974) and the types of set-theoretical predicates used in structuralism reveal that the structuralist representation scheme is more accurate than a non-set-theoretical, first-order representation of beliefs from a cognitive science perspective. The point is that the structuralist framework answers Minsky's request for *inter-propositional* knowledge representation, i.e., the association of propositions with information about how to use them.

4 Minimal Structuralism

A core idea of structuralism is to model the application of scientific theories to empirical systems through the application of set-theoretical predicates to sequences of sets that represent such empirical systems. The systematic use of set-theoretical predicates, therefore, distinguishes the structuralist representation scheme for scientific knowledge from other formal accounts in the philosophy of science. Classical structuralism, as expounded in Balzer et al. (1987), may further be seen as the most thorough elaboration of the *semantic view* of scientific theories, i.e., the view that the models of a scientific theory are essential for its identity. This semantic orientation does not necessarily preclude the representation of deductive and other forms of reasoning because it has been shown how the structuralist representation scheme can be transformed into an axiomatic representation of scientific theories in the style of a more Carnapian and hence more syntactic conception of scientific theories (Andreas 2010b). The surplus of this transformation is the formal representation of deductive reasoning, which was not available in classical structuralism.

An exposition of the whole theoretical apparatus of structuralist theory of science is well beyond the scope of the present paper. For this reason, we have developed a simplified version, which is intended to capture the most elementary ideas and which is expounded here for the first time. *Minimal Structuralism*

seemed a fitting description and a suitable label for this simplified version. For a wide range of studies, including also scientific examples, minimal structuralism will be sufficiently expressive. Unlike classical structuralism it allows for the formal representation of deductive and abductive reasoning and is in this respect even more expressive than classical structuralism.

Let us begin with some explanations concerning the types of set-theoretical structures and set-theoretical predicates that are introduced in a structuralist representation of scientific knowledge. The sequences of sets being used to represent scientific knowledge consist of a sub-sequence of base sets and another sub-sequence of relations:

$$\langle D_1, \ldots, D_k, R_1, \ldots, R_n \rangle \tag{2}$$

where $D_1, \ldots, D_k$ are base sets and $R_1, \ldots, R_n$ relations. Sequences of this type are also called set-theoretical *structures*. Set-theoretical predicates in structuralism lay down certain restrictions on the sets that are admitted as values in a sequence of sets. More precisely, one distinguishes between *typifications*, *characterisations*, and *laws* concerning the relations $R_1, \ldots, R_n$. A typification is a statement of the form $R_i \in \sigma(D_1, \ldots, D_k)$, where $\sigma(D_1, \ldots, D_k)$ stands for a sequence of concatenated operations on the base sets $D_1, \ldots, D_k$. The types of operations are selection, Cartesian product, and power set. A typification thus indicates that a relation R_i is of a determined set-theoretical type over the sets $D_1, \ldots, D_k$. For details, see Balzer et al. (1987, pp. 6-14).

A further specification of set-theoretical structures is brought about by formulas applying to them. Such formulas must be built up from the symbols $D_1, \ldots, D_k, R_1, \ldots, R_n$ as well as logical and set-theoretical symbols. One distinguishes between characterisations and laws. A *characterisation* is a formula which contains besides set-theoretical and logical symbols only a symbol for precisely one relation R_i, $1 \leq i \leq n$. A *law*, by contrast, is a formula that establishes some universal, non-trivial connection between at least two relations of $R_1, \ldots, R_n$. Therefore, it must have occurrences of at least two symbols for relations. To indicate that the law is non-trivial, we also speak of the *substantial law* of the theory-element $\mathbf{T}$. A substantial law in this sense may consist of more than one formal axiom.

It has been observed for a long time that theory formation goes hand in hand with concept formation. That means, the advancement of a scientific theory comes with the introduction of concepts specific to that theory. In structuralism, such concepts are called $\mathbf{T}$-theoretical, where $\mathbf{T}$ stands for the theory through

which the concepts are introduced. Paradigmatic examples of **T**-theoretical concepts are the concepts of mass and of force in classical particle mechanics. Those concepts, by contrast, which are used to describe the empirical systems that are the subject of the application of **T**, are called **T**-non-theoretical.

The distinction between **T**-theoretical and **T**-non-theoretical concepts gives rise to the following distinction between two kinds of set-theoretical entities:

$$\langle D_1, \ldots, D_k, N_1, \ldots, N_p \rangle \tag{3}$$

$$\langle D_1, \ldots, D_k, N_1, \ldots, N_p, T_1, \ldots, T_q \rangle \tag{4}$$

Structures of type (3) are intended to represent empirical systems that are the subject of the application of **T**, whereas structures of type (4) represent **T**-theoretical *extensions* of structures of type (3). The extension simply consists in a valuation of the **T**-theoretical relation symbols. Thus, the symbols $N_1, \ldots, N_p$ designate **T**-non-theoretical relations, whereas $T_1, \ldots, T_q$ designate **T**-theoretical ones. The symbols $D_1, \ldots, D_k$ designate sets of empirical objects that make up the empirical system to which the theory is applied.

Why are there different sets of empirical objects and not just one set D? This allows for a more fine-grained characterisation of what the empirical subjects of theory application are like. The argument types of the **T**-non-theoretical and the **T**-theoretical concepts thus can be characterised more accurately. Specifications of the latter kind are introduced through *typifications* and *characterisations* of relations in the above indicated manner.

If the theory involves some mathematical apparatus, as functions to natural, rational, or real numbers, and operations on such functions, then symbols for sets of mathematical objects need to be introduced. This results in structures of the following types:

$$\langle D_1, \ldots, D_k, A_1, \ldots, A_m, N_1, \ldots, N_p \rangle \tag{5}$$

$$\langle D_1, \ldots, D_k, A_1, \ldots, A_m, N_1, \ldots, N_p, T_1, \ldots, T_q \rangle \tag{6}$$

$A_1, \ldots, A_m$ are sets of mathematical objects. Some or all of the **T**-non-theoretical and **T**-theoretical relations may be functions, i.e., binary many-to-one relations. In the theories of mathematical physics, most quantities are introduced as functions taking empirical objects as arguments and having mathematical objects as values. Think of the concepts of temperature, pressure, mass, force, electromagnetic field etc. In allowing the sub-sequence of sets of mathematical objects to be empty, structures of type (5) and (6) have structures of type (3) and (4) as special cases.

Now, here are the essentials of structuralist theory representation. The application of a theory **T** to a potentially complex empirical phenomenon is not modelled by first-order order statements. Rather, we need to start with a **T**-non-theoretical description of a system of empirical entities, which are given by structures of type (5). The application of a theory **T** is then represented as the claim that the **T**-non-theoretical description can be extended to a **T**-theoretical description such that **T**'s *substantial law* is satisfied. That means a formula applying to structures of type (6) and having occurrences of at least two of the relation symbols $N_1, \ldots, N_p, T_1, \ldots, T_q$ must be satisfied by a **T**-theoretical description of the empirical phenomenon, which is represented by a structure of type (5). Of course, a **T**-theoretical description of an empirical system is a structure of type (6). The set-theoretical representation of an empirical phenomenon to which a theory-element **T** is intended to be applied is called an *intended application* of **T**.

It is a further implication of the structuralist schema of theory application that certain *links* between the various theoretical descriptions of empirical systems must be satisfied. In general, links constrain the admissible theoretical descriptions of certain tuples of intended applications. If all elements of the tuple are intended applications of one and the same theory-element, such links are called *internal*.[5] If, by contrast, more than one theory-element is involved, one speaks of an *external* link. Links are introduced particularly if one and the same empirical object is involved in different intended applications. For reasons of simplicity, the present exposition will be confined to binary links.

As just indicated, the application of a theory **T** to empirical phenomena is seen in structuralism as the search for **T**-theoretical descriptions of those phenomena, where these descriptions must satisfy first **T**'s substantial law and second certain links to $\mathbf{T}'$-theoretical descriptions of other empirical phenomena. Hence, we need to characterise by formal means those **T**-theoretical structures that satisfy these two conditions. The strategy we adopt in the following is to introduce

[5] In classical structuralism such links are called *constraints*.

a higher-order relation $\mathbf{AE}(\mathbf{T})(x, y)$ having the intended meaning that x is an admissible extension of y. A $\mathbf{T}$-theoretical extension x is admissible if and only if it satisfies $\mathbf{T}$'s substantial law and all the links of $\mathbf{T}$ are satisfied.

If $\mathbf{AE}(\mathbf{T})(x, y)$, then x needs to be a $\mathbf{T}$-theoretical structure that extends the representation y of an empirical phenomenon by a $\mathbf{T}$-theoretical description. To express this formally, we need to have a restriction function $\mathbf{r}(\mathbf{T})(x)$ that "cuts" the $\mathbf{T}$-theoretical relations:

Definition 1. If $x = \langle D_1, \ldots, D_k, A_1, \ldots, A_m, N_1, \ldots, N_p, T_1, \ldots, T_q \rangle$, then $y = \mathbf{r}(\mathbf{T})(x)$ if and only if $y = \langle D_1, \ldots, D_k, A_1, \ldots, A_m, N_1, \ldots, N_p \rangle$, i.e., $x_i = y_i$ for all $i, 1 \leq i \leq k + m + p$.

Now, we are in a position to state the first postulate characterising $\mathbf{AE}(\mathbf{T})(x, y)$:

$$\mathbf{P}1(\mathbf{T}) \quad \forall x \forall y(\mathbf{AE}(\mathbf{T})(x, y) \rightarrow y = \mathbf{r}(\mathbf{T})(x))$$

Second, an admissible $\mathbf{T}$-theoretical extension needs to satisfy $\mathbf{T}$'s substantial law. In classical structuralism, the set of structures satisfying $\mathbf{T}$'s substantial law is designated by $\mathbf{M}(\mathbf{T})$, the set of models of $\mathbf{T}$, where $\mathbf{M}(\mathbf{T})$ is introduced by an explicit definition. (The logical form of such a definition will be indicated with the example below.) Having an explicit definition of $\mathbf{M}(\mathbf{T})$ available, we can advance another postulate:

$$\mathbf{P}2(\mathbf{T}) \quad \forall x \forall y(\mathbf{AE}(\mathbf{T})(x, y) \rightarrow x \in \mathbf{M}(\mathbf{T}))$$

One word on *potential models* is necessary here. Potential models are those structures of which it is sensible to ask whether or not $\mathbf{T}$'s substantial law is satisfied. In more formal terms, that means, that, if x is a $\mathbf{T}$-theoretical structure and A the open formula expressing $\mathbf{T}$'s substantial law, then $A(x)$ will have a well defined truth-value if x is a potential model of $\mathbf{T}$. The definition of $\mathbf{M}(\mathbf{T})$ is such that it follows from $x \in \mathbf{M}(\mathbf{T})$ that x is a potential model. For this reason, $\mathbf{P}2(\mathbf{T})$ implies that x is a potential model if $\mathbf{AE}(\mathbf{T})(x, y)$.

Let us come to deal with links. Even in minimal structuralism, we shall distinguish between *internal* and *external* links. Both, internal and external (binary) links can be characterised through an open formula $\phi(\mathbf{L}_i)(x, y)$. The following considerations will be confined to binary links. An extension to links having higher arity is not difficult to accomplish. If $\mathbf{L_i}$ is an internal link, then $\phi(\mathbf{L}_i)(x, y)$ is required to hold only for x and y being theoretical structures of one and the same theory-element. If, by contrast, $\mathbf{L_i}$ is an external link between

$\mathbf{T}$ and $\mathbf{T}'$, $\phi(\mathbf{L}_i)(x, y)$ is required to hold for ordered pairs consisting of a $\mathbf{T}$-theoretical and a $\mathbf{T}'$-theoretical structure.

In view of what we have said about links, we can indicate the general form of postulates for binary links:

$$\mathbf{P}3(\mathbf{L}_i, \mathbf{T}, \mathbf{T}') \quad \forall x_1 \forall y_1 \forall x_2 \forall y_2 (\mathbf{AE}(\mathbf{T})(x_1, y_1) \ \wedge \ \mathbf{AE}(\mathbf{T}')(x_2, y_2) \rightarrow \phi(\mathbf{L}_i)(x_1, x_2))$$

where $\mathbf{T}$ and $\mathbf{T}'$ need not be different.

At this point, we arrived at a formal characterisation of the admissible theoretical extensions of a $\mathbf{T}$-non-theoretical structure. What can we say about the empirical phenomena to which $\mathbf{T}$ is applied, i.e., the intended applications of $\mathbf{T}$? We know that the members of $\mathbf{I}(\mathbf{T})$, the set of intended applications, are structures of type (5), which is ensured by postulate $\mathbf{P}1(\mathbf{T})$. It is an assumption inherent in structuralism that the set $\mathbf{I}(\mathbf{T})$ cannot be characterised completely by formal means. Rather the notion of an intended application is to be taken in the literal sense that there are scientists who think that a theory is applicable to certain empirical systems. Among other things, paradigmatic examples play an important role in determining the range of such empirical systems. Henceforth, we assume that the extension of $\ulcorner \mathbf{I}(\mathbf{T}) \urcorner$ is given somehow. The information encoded by the set $\mathbf{I}(\mathbf{T})$ is comparable to a complete interpretation of the observation language in non-set-theoretical reconstructions of scientific theories.

What we can express formally about $\mathbf{I}(\mathbf{T})$ is the requirement that any intended application has an admissible $\mathbf{T}$-theoretical extension:

$$\mathbf{P}4(\mathbf{T}) \quad \forall y (y \in \mathbf{I}(\mathbf{T}) \rightarrow \exists x \mathbf{AE}(\mathbf{T})(x, y))$$

For reasons that will become obvious in the next section, we require the relation $\mathbf{AE}(\mathbf{T})$ to be one-to-one:[6]

$$\mathbf{P}5(\mathbf{T}) \quad \forall x_1 \forall x_2 \forall y (\mathbf{AE}(\mathbf{T})(x_1, y) \wedge \mathbf{AE}(\mathbf{T})(x_2, y) \rightarrow x_1 = x_2)$$

This postulate says that the relation $\mathbf{AE}(\mathbf{T})$ is one-to-many. That $\mathbf{AE}(\mathbf{T})$ is also many-to-one follows from postulate $\mathbf{P}1(\mathbf{T})$ and the definition of $\mathbf{r}(\mathbf{T})$. Hence, $\mathbf{P}1(\mathbf{T})$, $\mathbf{P}5(\mathbf{T})$, and the definition of $\mathbf{r}(\mathbf{T})$ together imply that $\mathbf{AE}(\mathbf{T})$ is one-to-one.

[6] This postulate is adopted from Hofer (2009, p. 53).

The postulates $\mathbf{P}1(\mathbf{T}) - \mathbf{P}5(\mathbf{T})$ do, we claim, capture the intended meaning of $\mathbf{AE}(\mathbf{T})$ completely. Henceforth, the set of these postulates is designated by $\Phi_p(\mathbf{T})$. This set is called a system of postulates for the theory-element $\mathbf{T}$. $\Phi_d(\mathbf{T})$ may designate the set of the definitions of $\mathbf{M}(\mathbf{T})$, $\mathbf{M_p}(\mathbf{T})$ and $\mathbf{r}(\mathbf{T})$.

In speaking of introducing the relation $\mathbf{AE}(\mathbf{T})$ through postulates, we implicitly assumed $\mathbf{AE}(\mathbf{T})$ to be a (higher-order) theoretical term. The semantics of theoretical terms and theoretical sentences has been expounded in Andreas (2010a). As we want to keep the presentation of belief revisions with theory-elements as concise as possible, we spare the reader the original semantics of the present system of postulates and corresponding equivalence theorems. In place of this, we will be directly concerned with the introduction of an abductive inference rule that allows for inferences from intended applications to theoretical extensions thereof. The formalism of *default logic* will be used for the formulation of such an inference rule because of the defeasible nature of abduction. Then, the semantics of default logic falls into place as the proper semantics for the axiomatic system we are developing.

5 Examples of Theory-Elements

A few examples of how models of a theory-element $\mathbf{T}$ are defined may ease the understanding of the abstract concepts before we move on to develop further the formalism of minimal structuralism. A simple example is the lever rule in classical mechanics (Sneed 1979, pp. 65–73). In the following semiformal exposition, n has the intended meaning of the distance function from the centre of rotation of the lever, and t the intended meaning of the mass function. The theory-element $\mathbf{LR}$ covers the case where the weights on either side are in equilibrium.

Definition 2. x is a system of the lever rule ($x \in \mathbf{M}(\mathbf{LR})$) if and only if there exist $D, \mathbb{R}, n, t$ such that

(i) $x = \langle D, \mathbb{R}, n, t \rangle$;

(ii) $|D| > |\emptyset|$;

(iii) $n : D \to \mathbb{R}$;

(iv) $t : D \to \mathbb{R}$;

(v) $\sum\limits_{y \in D} n(y) \cdot t(y) = 0$.

In this case it is obvious why the formalism of higher-order logic or set-theory is needed for the reconstruction. As the application of the theory concerns systems of empirical objects with variable cardinality, a non-set-theoretical first-order representation would not do the job.

Since we started the discussion of revisions with Gärdenfors's story about the purchase of a golden ring, we shall give a formalisation of at least two axioms being applied there. Admittedly, the formalisation looks contrived as the example does not require the set-theoretical formalism of structuralist theory of science. In other words, using that formalism for representing Gärdenfors's story is like using a sledge hammer to crack a nut. Yet, as is the case with many examples in applications of mathematical logic, particularly with those of Artificial Intelligence, they may well serve the purpose of conveying an understanding of the formalism. The meanings of the symbols will be obvious from the above formalisation, where the symbols of that formalisation are abbreviated now by their initials as follows: $J(x)$ - x is a jeweller; $S(x, y)$ - x sells y; $T\_G(x, y)$ - x testifies that y is made of gold; $M\_G(x)$ - x is made of gold.

Definition 3. x is a system of a theory about a jeweller's testimony ($x \in \mathbf{M}(\mathbf{JT})$) if and only if there exist D, J, S, T_G, M_G such that

(i) $x = \langle D, J, S, T\_G, M\_G \rangle$;

(ii) $|D| = 2$;

(iii) $J \subseteq D$;

(iv) $S \subseteq D \times D$;

(v) $T\_G \subseteq D \times D$;

(vi) $M\_G \subseteq D$;

(vii) $\forall y_1 \forall y_2 (J(y_1) \wedge \neg S(y_1, y_2) \wedge T\_G(y_1, y_2) \rightarrow M\_G(y_2))$.

The one-place relation M_G, which stands for being made of gold, is **JT**-theoretical in a weak sense. It is possible to determine the extension of that relation with the help of **JT** but it is not necessary to use this theory-element for that purpose, which would be required for M_G being **JT**-theoretical in the strong sense.

Another theory through which we can determine the extension of M_G, at least via negation, captures axiom (e5) of the story (symbols: $S\_A(x)$ - x is sulphuric acid; $E(x, y, z)$ - x is exposed to y at time z; $ST(x, y)$ - x is stained at time y):

Definition 4. x is a system of a theory about the sulphuric-acid-test of gold ($x \in \mathbf{M}(\mathbf{SAT})$) if and only if there exist $D_1, D_2, S\_A, E, ST, M\_G$ such that

(i) $x = \langle D_1, D_2, S\_A, E, ST, M\_G \rangle$;

(ii) $|D_1| = 2$;

(iii) $|D_2| \leq \aleph_0$;

(iv) $S\_A \subseteq D_1$;

(v) $E \subseteq D_1 \times D_1 \times D_2$;

(vi) $ST \subseteq D_1 \times D_2$;

(vii) $M\_G \subseteq D_1$;

(viii) $\forall y_1 \forall y_2 \forall y_3 (S\_A(y_2) \wedge E(y_1, y_2, y_3) \wedge ST(y_1, y_3) \rightarrow \neg M\_G(y_1))$.

The intended interpretation of D_2 is a set of time points, where it is assumed that this set has not more than countably many members.

As it stands, one and the same object may be classified as being made of gold according to an application of **JT** and, at the same time, classified as not being made of gold according to an application of **SAT**. This result is desired to some extent. It makes it possible to represent situations where two pieces of evidence conflict with one another such that the conflict at the theoretical level can be traced back to a conflict at the empirical level. On the other hand, it is desirable that the conflict becomes recognisable in the formal system. To achieve this, an *external link* must be established:

$\mathbf{P}3(\mathbf{L}_1, \mathbf{JT}, \mathbf{SAT})$ $\forall x \forall y \forall z (x \in \mathbf{M}(\mathbf{JT}) \wedge y \in \mathbf{M}(\mathbf{SAT}) \wedge z \in (D_1)_x \wedge z \in (D_1)_y \rightarrow (R_4)_x = (R_4)_y)$

This formulation uses the convention that, if $x = \langle D_1, \ldots, D_k, R_1, \ldots, R_n \rangle$, $(D_m)_x$, $1 \leq m \leq k$, denotes D_m and $(R_p)_x$, $1 \leq p \leq n$, R_p.

For readers being familiar with Minsky's seminal (1974) paper and with object oriented programming languages such as C++ and Java, whose development was deeply inspired by ideas of Minsky, the connection of the structuralist framework to *frames* will be worth consideration. According to the cognitive characterisation, a frame is a data structure that we re-collect from memory when we encounter a typical situation. According to the formal characterisation, a frame is a set of slots that allow only certain types of values as fillers. There are 'simple' type conditions that express these constraints on admissible values. In addition, there are 'complex' conditions that relate the values of

one slot to the values of another. In structuralism, simple type conditions are called *typifications* and complex ones *laws*. For a frame in the sense of Minsky, it is essential to be related to other frames in various ways. In structuralism, we have at least three types of relations among frames: internal and external links, and the relation of specialisation. (The latter relation was not covered by the present account.) Of course, our view is that theory-elements are frame concepts.

A frame in the sense of Minsky as well as a class in C++ and Java come with their own descriptive, i.e., non-logical vocabulary. This allows for a modularisation of data manipulation, which makes programming code more intuitive, less error-prone and easier to maintain. In precisely the same way it holds for the structuralist framework that the descriptive vocabulary of the global language of science is divided into sub-vocabularies, where a theory-element has its own non-logical vocabulary. Intended applications and **T**-theoretical extensions thereof are individuated by a particular interpretation of that sub-vocabulary. Scientific reasoning, then, occurs at different levels: within an intended application of a theory-element, among the intended applications of one and the same theory-element (by means of internal links), among different intended applications of different theory-elements (by means of external links). Arguably, this allows for a more fine-grained and more accurate representation of scientific reasoning.

6 Abductive Inferences in Minimal Structuralism

Abduction is a mixed bag. Informally speaking, an abductive inference is an inference from phenomena to hypotheses explaining such phenomena. A more formal account is given by the following inference rule (Flach 2000, p. 93):

$$\frac{C \qquad A \models C}{A} \tag{7}$$

Underlying this inference rule is the view that abductions are inverted deductions, which is common to most AI approaches to abduction. Hence, the rule allows one to infer from a proposition C another proposition A which has C among its semantic consequences.

In analysing reasoning from phenomena to potential explanations, however, we should distinguish more carefully between the background theory and the hy-

pothesis, which together explain the phenomenon. The kind of inference we are particularly interested in here is one that leads to an *interpretation* of phenomena in light of an antecedently given theory to the effect of extending the report of phenomena by a theoretical description. (7) would only make sense if C were to include some piece of a theory, but the process of theory formation seems too chaotic and too complex to be captured by a simple inference rule.

This being said, we may improve upon (7) as follows:

$$\frac{C \qquad A, T \models C}{A} \tag{8}$$

where T stands for a theory being given in some axiomatic formulation, C represents an observed phenomenon and A a specific, theoretical antecedent condition that entails C in the context of T. This inference rule is inspired by and intended to capture the ingenious definition of a *conjectural consequence relation* by Flach (2000, p. 96). If one thinks of explanations in terms of the old DN approach, inference rule (8) can be said to allow for inferences from empirical phenomena to hypotheses potentially explaining those phenomena, together with a non-empty set of general laws: A, T represents the explanans, C the explanandum.

The DN approach to explanation encountered some rivals in the second half of the twentieth century. Among them, the unification approach to scientific explanation does seem to be of particular relevance to our investigation (Friedman 1974; Kitcher 1976). A formal equivalent of the unification approach has been seen in the idea that a **T**-theoretical structure being a model of **T** and extending an intended application of y explains y with the help of **T** since y becomes embedded into a theoretical model once such a **T**-theoretical structure is found. Intended applications in the structuralist sense are thus seen as empirical phenomena waiting to be explained. This model-theoretic notion of explanation has been advanced by Stegmüller (1985, p. 113) and was systematically investigated by Bartelborth (2002). Taking it up, we can give a first, approximative explanation of what a hypothesis explaining a phenomenon is in structuralism:

Explanation 1. Let y be a **T**-non-theoretical structure that represents some phenomenon. Then, a **T**-theoretical structure x explains y with the help of a background theory **T** if and only if i) $\mathbf{r}(\mathbf{T})(x) = y$ and ii) $x \in \mathbf{M}(\mathbf{T})$.

Using proposition (1) for an abductive inference results in the following rule:

$$\frac{y \in \mathbf{I}(\mathbf{T}), y = \mathbf{r}(\mathbf{T})(x), x \in \mathbf{M}(\mathbf{T})}{(x, y) \in \mathbf{AE}(\mathbf{T})} \tag{9}$$

This inference rule schema represents corresponding inference rules for every theory-element $\mathbf{T}$ under consideration in the logical reconstruction. Unlike (8), (9) does not require the empirical phenomenon, that is, in this case, the intended application, to be derivable from the explanans.[7] Hence, the abductive inference rule (9) is more liberal than (8).[8]

That y is an intended application of $\mathbf{T}$, x a $\mathbf{T}$-theoretical extension of y, and x satisfies $\mathbf{T}$'s substantial law are minimal requirements for x to be an admissible extension of y. Moreover, postulates of type $\mathbf{P}3(\mathbf{L}_i, \mathbf{T}, \mathbf{T}')$ must hold, i.e., any link between $(x, y) \in \mathbf{AE}(\mathbf{T})$ and $(x', y') \in \mathbf{AE}(\mathbf{T}')$ must be satisfied. It appears sensible, therefore, to include the satisfaction of links to other, potentially admissible theoretical extensions in the antecedence of the inference rule (9). The problem arising then is that we would need to know the extension of $\mathbf{AE}(\mathbf{T}')$ for any $\mathbf{T}'$ being potentially linked to $(x, y) \in \mathbf{AE}(\mathbf{T})$. However, any proposition of the form $(x', y') \in \mathbf{AE}(\mathbf{T}')$ can only be inferred by means of the abductive inference rule (9) in the first place. For this reason, it is not possible to include the satisfaction of links among the premisses of this rule.

In this situation, we take recourse to the formalism of *default logic*. This logic comes with a new type of inference rule that has the syntactic form

$$\frac{\phi : \psi_1, \ldots, \psi_n}{\chi} \tag{10}$$

where ϕ, $\psi_1, \ldots, \psi_n$, and χ are closed formulas of predicate logic. Its meaning is: If ϕ and it is consistent to assume that $\psi_1, \ldots, \psi_n$, then χ. Consistency is understood as being relative to a set W of presumed facts, or accepted sentences, joined with the set of sentences that have been derived from W so far. $\psi_1, \ldots, \psi_n$

[7] The intended application is only derivable in the trivial sense that it can be obtained by the restriction function $\mathbf{r}(\mathbf{T})$ from the theoretical extension. Yet, the intended application need not be derivable from the theoretical part of the theoretical extension x, that is, the $\mathbf{T}$-theoretical relations of x. In a more extended discussion it could be shown that (9) allows us to capture inferences licensed by (8) as well, if the theory T in (8) is formalised in the structuralist style.

[8] Even though inference rule (9) is more liberal than (8), it does not capture every kind of inference being associated with the term "abduction". In particular, (9) does not capture the process of theory formation. The term "abduction" is used here merely to make the exposition more vivid. Nothing depends on how appropriate this use is. For an excellent systematic survey concerning abduction see Schurz (2008).

are called *justifications* or *consistency conditions*. A *normal default* is one where the consistency conditions are identical with the consequent χ:

$$\frac{\phi : \chi}{\chi} \tag{11}$$

Open defaults, i.e., default rules with occurrences of free variables, are, in standard default logic, taken to represent all of their ground instances.

Using default logic, the satisfaction of links can be taken into account through the following two steps. First, every postulate of type $\mathbf{P}3(\mathbf{L}_i, \mathbf{T}, \mathbf{T}')$ is included in the set W of accepted sentences. Second, inference rule (9) is turned into a default rule as follows:

$$\delta 1(\mathbf{T}) \qquad \frac{y \in \mathbf{I}(\mathbf{T}), y = \mathbf{r}(\mathbf{T})(x), x \in \mathbf{M}(\mathbf{T}) : (x, y) \in \mathbf{AE}(\mathbf{T})}{(x, y) \in \mathbf{AE}(\mathbf{T})}$$

In words: If y is an intended application of $\mathbf{T}$, x a $\mathbf{T}$-theoretical extension of y, x a model of $\mathbf{T}$, and it is consistent to assume that $(x, y) \in \mathbf{AE}(\mathbf{T})$, then $(x, y) \in \mathbf{AE}(\mathbf{T})$. Suppose there is an ordered pair $(x', y') \in \mathbf{AE}(\mathbf{T}')$ such that a link postulate $\mathbf{P}3(\mathbf{L}_i, \mathbf{T}, \mathbf{T}')$ is falsified for the quadruple (x, y, x', y'). In this case, $(x, y) \in \mathbf{AE}(\mathbf{T})$ cannot consistently be added to the stock of accepted sentences.

Let us be more precise about the default theory capturing abductive inferences to theoretical extensions of intended applications. As is well known, a default theory is a pair (W, D), where W is a set of sentences and D a set of default rules. In our case, W contains any proposition of the form $b \in \mathbf{I}(\mathbf{T})$, where b is a $\mathbf{T}$-non-theoretical structure and $\mathbf{T}$ a theory-element. Further, W contains the postulates $\mathbf{P}1(\mathbf{T}) - \mathbf{P}3(\mathbf{T})$ and $\mathbf{P}5(\mathbf{T})$ for any theory-element $\mathbf{T}$. Moreover, the set of definitions given by $\Phi_d(\mathbf{T})$ is contained in W for any theory-element $\mathbf{T}$.[9]

The general form of the defaults in our default theory is given by the schematic inference rule $\delta 1(\mathbf{T})$. Since the number of theory-elements to be considered and

[9] One may wonder whether $\mathbf{P}1(\mathbf{T})$ and $\mathbf{P}2(\mathbf{T})$ are not dispensable once the inference rule $\delta 1(\mathbf{T})$ is introduced. This is not the case. According to this rule, $y \in \mathbf{I}(\mathbf{T}), y = \mathbf{r}(\mathbf{T})(x), x \in \mathbf{M}(\mathbf{T})$ are sufficient conditions for inferring $(x, y) \in \mathbf{AE}(\mathbf{T})$ as long as $(x, y) \in \mathbf{AE}(\mathbf{T})$ can consistently be added. By contrast, $\mathbf{P}1(\mathbf{T})$ and $\mathbf{P}2(\mathbf{T})$ state that $y = \mathbf{r}(\mathbf{T})(x), x \in \mathbf{M}(\mathbf{T})$ are necessary conditions for $(x, y) \in \mathbf{AE}(\mathbf{T})$.

In not including $\mathbf{P}4(\mathbf{T})$ among the axioms of W, we obtain a system that is more liberal than $\Phi_p(\mathbf{T}) \cup \Phi_d(\mathbf{T})$ with deductive logic. The motivation for not including $\mathbf{P}4(\mathbf{T})$ will become apparent in the next section when priorities for defaults are introduced to account for the reliability of applying theory-elements to empirical systems.

the number of intended applications of any theory-element are always finite, the schematic letter $\mathbf{T}$ and the free variable y may be replaced by all of its instances, which yields a set of open defaults of the form:

$$\delta 2(\mathbf{T}) \qquad \frac{b_j \in \mathbf{I}(\mathbf{T}), b_j = \mathbf{r}(\mathbf{T})(x), x \in \mathbf{M}(\mathbf{T}) : (x, b_j) \in \mathbf{AE}(\mathbf{T})}{(x, b_j) \in \mathbf{AE}(\mathbf{T})}$$

Replacing the remaining free variable in $\delta 2(\mathbf{T})$ would lead to a cumbersome formulation of the set D of defaults. We shall therefore "postpone" the replacement of open defaults with their ground instances to the definition of the inference notions in the next section. This will allow us to make an important observation there concerning the applicability of open defaults.

7 The Semantics of Default Logic

Assuming the reader to be roughly familiar with the elements of default logic, here we will review briefly the notion of an extension and the corresponding inference notions, with consideration of explanation (6). In the standard semantics for default logic it is assumed that open defaults are replaced by all of their ground instances such that every default $\delta \in D$ is closed (cf. Brewka et al. (2008) and Antoniou (1997), p. 25). The following definitions, however, are adjusted to default theories with open normal defaults having exactly one occurrence of a free variable, as is the case with our default theory. Let us begin with the fixed-point semantics. There, the notion of an extension is defined as follows:

Definition 5. Let (W, D) be a default theory. The operator Γ assigns to every set S of formulas the smallest set U of formulas such that:

(i) $W \subseteq U$,

(ii) $Cn(U) = U$,

(iii) For all substitutions (u/x) with u being a tuple of ground terms: if $\phi(x) : \psi(x)/\psi(x) \in D, U \models \phi(u/x), S \not\models \neg\psi(u/x)$, then $\psi(u/x) \in U$.

A set E of formulas is an extension of (W, D) if and only if $E = \Gamma(E)$, that is, E is a fixed-point of Γ.

The operational definition of an extension by Antoniou (1997) may in some respect seem more intuitive. The underlying idea is to define, first, what it is for a sequence of closed defaults to be applicable to the set of currently accepted sentences. The latter set grows with each application of a default unless the consequence of the default was already believed in. Second, the idea is to define what it is for such a sequence to be maximal in the sense that no further default of D can be applied. Sequences of defaults satisfying both conditions are called *closed and successful processes*. Let us begin with the definition of the concept of a default being applicable to a deductively closed set of formulas:

Definition 6. A closed normal default $\delta = \frac{\phi:\psi}{\psi}$ is applicable to a deductively closed set of formulas E if and only if $\phi \in E$ and $\neg\psi \notin E$.

Now, let Π be a sequence of closed defaults such that, for every element δ_1 of Π, there is a substitution (u/x) and some $\delta_2 \in D$ with $\delta_1 = \delta_2(u/x)$. This sequence is a *successful process* if and only if every default can in the order given by the sequence be applied to the deductive closure of W joined with consequences of previously applied defaults. Π is a *closed process* if and only if there is no default $\delta_2 \in D$ and no substitution (u/x) such that $\delta_2(u/x) \notin \Pi$ and $\delta_2(u/x)$ is applicable to $E = Cn(W \cup \{cons(\delta) \mid \delta \text{ occurs in } \Pi\})$, where $cons(\delta)$ denotes the consequence of the default δ. Then, the notion of an extension of the default theory can be defined as follows:

Definition 7. A set E of formulas is an extension of the default theory (W, D) if and only if there is some closed and successful process Π of (W, D) such that $E = Cn(W \cup \{cons(\delta) \mid \delta \text{ occurs in } \Pi\})$.

Note that the notion of an extension in the sense of default logic must by no means be confounded with the notion of a **T**-theoretical extension in the sense of structuralism. Both notions are in play in our system, however.

Note that, in our default theory, an open default cannot be applied with two different substitutions in one and the same successful process. If an open default of the form $\delta 2(\mathbf{T})$ has been applied for a substitution (u_1/x), we receive a conclusion of the form $(u_1, b_j) \in \mathbf{AE}(\mathbf{T})$. Because of this and because $\mathbf{P}5(\mathbf{T})$ requires the relation $\mathbf{AE}(\mathbf{T})$ to be one-to-many, $\delta 2(\mathbf{T})$ cannot be applied with another substitution $(u_2/x), u_2 \neq u_1$ in the same process. However, several instantiations of such an open default may occur in different processes being closed and successful.[10]

[10]For the present structuralist theory of belief revision this has the consequence that $\mathbf{P}5(\mathbf{T})$ does not imply a unique valuation of the **T**-theoretical relations.

Upon the notion of an extension of a default theory, two inference relations are introduced:

Definition 8. A sentence ϕ can sceptically be inferred from a default theory (W, D) - in symbols: $(W, D) \mid\sim_s \phi$ - if and only if ϕ is a member in any extension of (W, D).

Definition 9. A sentence ϕ can credulously be inferred from a default theory (W, D) - in symbols: $(W, D) \mid\sim_c \phi$ - if and only if ϕ is a member of at least one extension of (W, D).

In Andreas (2010a), we argued at length for a semantics in which a theoretical sentence ϕ is true only then if it is true in any admissible interpretation of the language, where the set of admissible interpretations is determined by the postulates and the intended interpretation of the non-theoretical symbols. Following these considerations, we should only trust the sceptically valid inferences of (W, D).

In the exposition of the inference notions of default logic, the replacement of open defaults with their ground instances has been moved to the definition of the notion of an extension of a default theory (W, D). Both the standard and the present definitions work properly only when there is, for every object in the domain of interpretation, a ground term naming that object. A proper semantics for open defaults has been developed, among others, by Lifschitz (1990). This semantics could be used for our default theory as an alternative to the present modification of the standard definitions. Lifschitz, however, though working with structures, finally uses substitutions of free variables too by introducing extensions of the set of object constants.[11]

In yet another respect, the present system differs from standard expositions of default logic. The present default theory comes with semiformal axioms couched in naive set theory. Its complete formalisation, therefore, would require either use of formal set theory or higher-order logic. Standard expositions of default logic, by contrast, are confined to first-order default theories. Since, however, the syntax and semantics of default logic do not essentially rely on a first-order setting and since the present use of that logic is more philosophically than computationally motivated, no objection shall arise from that difference.

[11] Working with substitutions of free variables in open defaults is objectionable in similar veins as the substitutional reading of the quantifiers is in predicate logic. In the case of unnamed objects in the domain of interpretation, we may lose important solutions. Particularly critical is the treatment of real numbers which form an uncountable domain. For a defence of the substitutional quantification against common objections, see Wallace (1971) and Gottlieb and McCarthy (1979).

8 Theory-Elements Prioritised

The key motivation for the study of belief change is the observation that incoming information happens to be inconsistent with our presently accepted body of beliefs quite frequently. In applying theory-elements to empirical systems, we are facing inconsistency problems as well. Contrary to the postulates $\mathbf{P}2(\mathbf{T})$ and $\mathbf{P}4(\mathbf{T})$, an intended application of $\mathbf{T}$ may not have a theoretical extension being a model of $\mathbf{T}$. More frequent is the case where two intended applications $b_1 \in \mathbf{I}(\mathbf{T}_i)$ and $b_2 \in \mathbf{I}(\mathbf{T}_j)$ have theoretical extensions being models of $\mathbf{I}(\mathbf{T}_i)$ and $\mathbf{I}(\mathbf{T}_j)$ respectively, but there are no corresponding theoretical extensions x_1 and x_2 such that $x_1 \in \mathbf{M}(\mathbf{T}_i)$, $x_2 \in \mathbf{M}(\mathbf{T}_j)$, and x_1 and x_2 are satisfying every link between $(\mathbf{T}_i)$ and $(\mathbf{T}_j)$.

In the case of our simple case study, there is no *process* Π that contains both an instance of $\delta 1(\mathbf{JT})$ and one of $\delta 1(\mathbf{SAT})$. There is an extension, however, that contains a set-theoretical representation of the proposition that the ring is made of gold and another extension that contains a set-theoretical representation of the proposition that the ring is not made of gold. Hence, neither proposition can sceptically be inferred from the default theory. Reference to credulous inferences would not make sense here because we could infer then both a sentence ϕ and its negation.

Our knowledge of chemistry and our everyday experience say that the chemical test with sulphuric acid is more reliable than the testimony of the jewellers. Hence, we think that the proposition that the ring is not made of gold should be inferable. How shall we formally represent this? Experts on default logic will certainly anticipate the answer: We introduce priorities among defaults! These priorities represent reliability information in the sense that $\delta 1(\mathbf{T}_i) <_R \delta 1(\mathbf{T}_j)$ means that the application of any instance of $\delta 1(\mathbf{T}_j)$ is more reliable than the application of any instance of $\delta 1(\mathbf{T}_i)$. Note that the priority order concerns open defaults in the form of $\delta 1(\mathbf{T})$. Any two instances of such an open default occupy the same position in the ranking of defaults.

Even though we may have some intuitions about the reliability of theory application, it is desirable to have a precise explanation of this notion. In Section 3 we indicated that a finite frequency interpretation of reliability may do the job. Let us assume that the theory-elements have a history of applications that is not represented by the sets $\mathbf{I}(\mathbf{T})$. Let a *successful application* be one from which a $\mathbf{T}$-theoretical proposition was inferred that remained accepted. An application of a theory-element is unsuccessful, by contrast, if there is no such $\mathbf{T}$-theoretical proposition. For example, an experimental test whose result became accepted without revision is a successful application of a corresponding theory-element.

If, however, the experimental result could not gain acceptance or was revised at a later time, this is an instance of an unsuccessful application.

Let $C_s(\mathbf{T})$ denote the cardinality of the set of successful applications of $\mathbf{T}$ and $C_u(\mathbf{T})$ denote the cardinality of the set of unsuccessful applications of $\mathbf{T}$. Then, our priority order should satisfy the following minimal requirement:

$$\text{If } \delta 1(\mathbf{T}_i) <_R \delta 1(\mathbf{T}_j), \text{ then } \frac{C_s(\mathbf{T}_i)}{C_s(\mathbf{T}_i) + C_u(\mathbf{T}_i)} < \frac{C_s(\mathbf{T}_j)}{C_s(\mathbf{T}_j) + C_u(\mathbf{T}_j)} \quad (12)$$

In short, this requirement says that if the application of $\mathbf{T}_j$ is considered more reliable than the application of $\mathbf{T}_i$, then the ratio of successful applications of $\mathbf{T}_j$ must be higher than that ratio of $\mathbf{T}_i$.

We may also introduce the following stronger condition:

$$\delta 1(\mathbf{T}_i) <_R \delta 1(\mathbf{T}_j) \text{ if and only if } \frac{C_s(\mathbf{T}_i)}{C_s(\mathbf{T}_i) + C_u(\mathbf{T}_i)} < \frac{C_s(\mathbf{T}_j)}{C_s(\mathbf{T}_j) + C_u(\mathbf{T}_j)} \quad (13)$$

Introducing the weaker condition allows one to take other aspects into account because it allows for exceptions from a strict finite frequency interpretation of the reliability of theory-elements. Further research is needed to show whether additional rules must be considered for the epistemic ranking of theory-elements. Even so, this alone would have no bearing on the thesis that the epistemic ranking of propositions can be replaced with a ranking of inferential patterns being effective for derived beliefs. The central claim of the present investigation is not that the finite frequency interpretation of the epistemic ranking of theory-elements is more than an approximation and thus comprehensive.

So much for the material conditions that the priority order must satisfy. The formal conditions that $<_R$ must satisfy should not be chosen too restrictively. It is not sensible to require that the priority information for theory-elements is given in the form of a strict well order. For example, if

$$\frac{C_s(\mathbf{T}_i)}{C_s(\mathbf{T}_i) + C_u(\mathbf{T}_i)} = \frac{C_s(\mathbf{T}_j)}{C_s(\mathbf{T}_j) + C_u(\mathbf{T}_j)}$$

then neither $\delta 1(\mathbf{T}_i) <_R \delta 1(\mathbf{T}_j)$ nor $\delta 1(\mathbf{T}_i) >_R \delta 1(\mathbf{T}_j)$ should be assumed. Consider further the case where a new theory comes into play. In that case, no statistics about the ratio of successful applications of its theory-elements will be available. Hence, it is hardly achievable, if not impossible, to give precise priority

information for these theory-elements unless the applications of some other well established theory can be reduced to the applications of the new theory. To require $<_R$ to be a strict partial order leaves enough room for cases where no precise priority order among two or more theory-elements can be determined. Another important reason for not requiring $<_R$ to be a strict well order is that any two instances of an open default in the form of $\delta 1(\mathbf{T})$ must occupy the same position in the ranking.

Still, it may be argued that in the case where the ratio of successful applications of $\mathbf{T}_j$ is only marginally higher than the ratio of successful applications of $\mathbf{T}_i$, one should not assume that $\delta 1(\mathbf{T}_i) <_R \delta 1(\mathbf{T}_j)$. To account for this point, one shall require the difference in the ratio of successful applications to be higher than or equal to a certain margin $m, m < 1$:

$$\delta 1(\mathbf{T}_i) <_R \delta 1(\mathbf{T}_j) \text{if and only if} \frac{C_s(\mathbf{T}_i)}{C_s(\mathbf{T}_i) + C_u(\mathbf{T}_i)} < \frac{C_s(\mathbf{T}_j)}{C_s(\mathbf{T}_j) + C_u(\mathbf{T}_j)} \text{ and } \frac{C_s(\mathbf{T}_j)}{C_s(\mathbf{T}_j) + C_u(\mathbf{T}_j)} - \frac{C_s(\mathbf{T}_i)}{C_s(\mathbf{T}_i) + C_u(\mathbf{T}_i)} \geq m \tag{14}$$

Let us now come back to the formalism of prioritised default logic. The effect of introducing priorities among defaults is simply that defaults cannot be applied in an arbitrary order any more. Rather, the order of defaults in a successful process must respect the priority information of defaults. Here are the formal details of prioritised default logic (PDL) (Antoniou 1997, p. 93):

Definition 10. (PDL-Extension) $T = (W, D, <)$ is a prioritised default theory if (W, D) is a normal default theory and $<$ a strict partial order on D. E is a PDL-extension of T if and only if there is a strict well order $\ll$ on D which contains $<$ and generates E.

A strict well order $\ll$ on D is said to generate an extension E if and only if there is a closed and successful process Π such that the elements of Π are ordered according to $\ll$ so that the first elements of Π are those of highest priority, and $E = Cn(W \cup \{cons(\delta(u/x)) \mid \delta(u/x) \,\text{occurs in}\, \Pi\})$. Note that it is not necessary that any default of D is applied. A strict well order $\ll$ contains a strict partial order $<$ if and only if it holds for all $x, y : x < y \rightarrow x \ll y$.

Once a priority among theory-elements is introduced in our example, saying that the theory-element of the chemical test is more reliable than the testimony of the jewellers, and the formalism of PDL is adopted, it is no longer admissible

to apply an instance of the open default $\delta 1(\mathbf{JT})$ before one of $\delta 1(\mathbf{SAT})$ has been applied, given that the open default $\delta 1(\mathbf{SAT})$ is applicable at all to the set W of facts. Hence, the result is that any extension of the default theory represents the proposition that the ring is not made of gold and no extension represents the opposite result. Hence, the proposition that the ring is not made of gold can be inferred sceptically from the default formalisation of Gärdenfors's example.

9 The Final Synthesis

In the preceding sections, we have attempted to demonstrate how a dynamisation of classical structuralism can be brought about through the following operations on its formalism: (i) introduction of the relation $\mathbf{AE}(\mathbf{T})$ by postulates, (ii) introduction of an abductive inference rule that is embedded in the formalism of default logic, and (iii) introduction of priorities among the theory-elements, which represent the reliability of their application. One important surplus of this dynamisation is that potentially inconsistent information - inconsistent in the sense of classical logic - can be dealt with in a sensible way, that is, without letting reasoning break down because anything becomes inferable from the accepted body of beliefs.

Where is the connection of the dynamisation of classical structuralism to belief revision theory? Once an inferential formalism is available that is sufficiently powerful to deal with classically inconsistent information and, moreover, a division of beliefs into derived and non-derived ones and expectations available, revisions can be defined in a straightforward manner. This is an implication of Rott's (2001) investigation of base generated revisions. Thus, the style in which he defines base generated revisions can be adopted for our prioritised default theory of scientific theories in the structuralist framework, as we will show now.

Rott (2001) distinguishes between the direct and the coherence constrained mode of base revisions. According to the direct mode, the revision of belief bases is plain and straightforward insofar as incoming beliefs are simply added to the belief base. Simple removals of elements of the base are also allowed. This direct form of base change is combined with a more sophisticated formalism of deriving beliefs that should be paraconsistent and nonmonotonic. In the coherence constrained mode, by contrast, belief bases are changed by sophisticated operations that require choices where the incoming information is inconsistent with the base. This kind of base change is combined with a straight form of theory derivation, in which the inference relation of classical logic is adopted.

For the direct mode of base revisions, Hans Rott defines the set of sentences

inferable from a prioritised base $\mathcal{H} = \langle H_1, \ldots, H_n \rangle$ and a set of prioritised expectations $\mathcal{E} = \langle E_1, \ldots, E_m \rangle$ as follows (2001, p. 128):

$$A = Inf(\mathcal{H}) = Consol(\mathcal{E} \circ \mathcal{H}) \tag{15}$$

where A stands for the set of accepted propositions and $Consol$ for a nonmonotonic formalism through which a classically consistent set of consequences can be attained from a potentially classically inconsistent set of basic beliefs plus expectations. $\mathcal{E} \circ \mathcal{H}$ designates the concatenation of $\mathcal{E}$ and $\mathcal{H}$, i.e., a sequence of sets of the form $\langle E_1, \ldots, E_m, H_1, \ldots, H_n \rangle = \langle G_1, \ldots, G_k \rangle = \mathcal{G}$, where $k = m + n$. Brewka's construction of *preferred subtheories* (1991; 1989) is used to define $Consol(\mathcal{E} \circ \mathcal{H})$. Here is a condensed account of this definition: Any subset of $G_1 \cup \ldots \cup G_n$ is a subtheory of $\mathcal{G}$. A set F is a preferred subtheory of $\mathcal{G}$ if and only if (i) it is classically consistent; and (ii) there is no classically consistent subtheory F' of $\mathcal{G}$ such that there is an i with $(F \cap G_i) \subset (F' \cap G_i)$ and, for all $j > i$, $F \cap G_j = F' \cap G_j$. A sentence ϕ is an element of $Consol(\mathcal{E} \circ \mathcal{H})$ if and only if it is entailed by all preferred subtheories of $\mathcal{G}$.

Once $Inf(\mathcal{H})$, the inference operation for a prioritised base in the context of prioritised expectations, has been defined by means of $Consol$, the definition of revisions is straightforward (Rott 2001, p. 130):

$$A * \phi = Inf(\mathcal{H} \circ \langle \phi \rangle) = Consol(\mathcal{E} \circ \mathcal{H} \circ \langle \phi \rangle) \tag{16}$$

where $\circ\langle\phi\rangle$ simply stands for placing ϕ on top of the prioritised base $\mathcal{H}$. The new incoming information ϕ has thus top priority. If the belief base is not prioritised, as is the case with our system, $\mathcal{H} \circ \langle \phi \rangle$ reduces to $H \cup \{\phi\}$, where H is the belief base voided of priorities.

For our system it is natural to use what Rott describes as the direct mode of base generated revisions. Incoming information has the form $b_i \in \mathbf{I}(\mathbf{T}_j)$ and is simply added to the base. All members of the base have this logical form; no priority order for the elements of H is assumed. Contractions consist in the elimination of a proposition of the form $b_i \in \mathbf{I}(\mathbf{T})$, accordingly. Let us introduce some more symbolic notations for the definition of belief changes in our system:

(i) H is the set of accepted propositions saying that a non-$\mathbf{T}_j$-theoretical structure b_i is an intended application of $\mathbf{T}_j$.

(ii) Φ_p is the set of postulates $\mathbf{P}1(\mathbf{T})$ - $\mathbf{P}3(\mathbf{T})$, and $\mathbf{P}5(\mathbf{T})$ for any theory-element $\mathbf{T}$.

(iii) Φ_d is the set of definitions of $\mathbf{M_p}(\mathbf{T})$, $\mathbf{M}(\mathbf{T})$, and $\mathbf{r}(\mathbf{T})$ for any theory-element $\mathbf{T}$.

(iv) D is the set of defaults in the form of $\delta 2(\mathbf{T})$ for any theory-element $\mathbf{T}$ and any intended application $b \in \mathbf{I}(\mathbf{T})$.

(v) $<_R$ is a strict partial order for defaults in the form of $\delta 2(\mathbf{T})$ that represents the reliability of the application of the corresponding theory-elements.

(vi) $W = H \cup \Phi_p \cup \Phi_d$

(vii) $Inf(W, D, <_R) = \{\phi \mid (W, D, <_R) \mid\sim_s \phi\}$

According to (vii), PDL is used as nonmonotonic formalism in place of *Consol*.[12] The present definitions of revisions and contractions thus differ from Rott's original ones with regard to the inference formalism in use. (This variation is feasible since Rott's systematisation of base revisions is not bound to a particular nonmonotonic inference operation being used to define such revisions in the direct mode.) Moreover, no priority information of the belief base is needed for these definitions. This simplifies the overall ranking information needed for the system substantially.

Now, we can define revisions and contractions along the lines of Brewka (1991) and Rott (2001):

$$A = Inf(W, D, <_R) \tag{17}$$

$$A * \phi = Inf(W \cup \{\phi\}, D, <_R) \tag{18}$$

$$A \div \phi = Inf(W \setminus \{\phi\}, D, <_R) \tag{19}$$

[12] One must wonder whether the inference formalism of PDL is also paraconsistent. Strictly speaking, the answer to this question is no. If the set W of the triple $(W, D, <_R)$ is classically inconsistent, then the set $Inf(W, D, <_R)$ is also inconsistent. Yet, the inference operation of default logic is paraconsistent in the sense that the introduction of consistency conditions $\psi_1, \ldots, \psi_n$ in (10) turns a classically inconsistent set of facts plus defeasible inference rules into a pair (W, D) from which sensible information can be inferred.

where ϕ is always a proposition of the form $b_i \in \mathbf{I}(\mathbf{T}_j)$.

The use of prioritised default logic in place of Brewka's construction of preferred subtheories deserves a brief consideration, finally. There are no compelling reasons for this choice. The defeasible nature of the abductive inference rules of the present system could also be accounted for by means of the preferred subtheories approach to defeasible reasoning. It is not difficult to transform the present system into one that works with that approach. To our mind, however, the default presentation of the system has the merit of being intuitively very well accessible. Its syntactic format allows for a very direct and, hence, intuitive representation of the defeasible nature of abductive inference rules of type $\delta 2(\mathbf{T})$ and their priority ordering. As for the semantics, it must be observed that the operational semantics for prioritised default theories with only normal defaults strongly resembles the construction of preferred subtheories.

10 Further Examples

The discussion of a few more examples may serve to illuminate some features of the present system. It is reasonable to think that the succession of scientific theories in the history of science can be represented as a process of revisions, in which the advancement of novel theories leads to a retraction of older ones. According to this view, Newtonian physics was retracted at the time when relativistic physics could be shown to have a wider range of successful applications and to resolve certain anomalies of Newtonian physics. The picture coming with the present system, however, is not so strict as to assume retraction of and revision by whole theories. Rather, it is *intended applications* of certain axioms of scientific theories as opposed to whole scientific theories that are the subject of belief changes in science. Scientific theories - in the sense of formal or informal axiomatic systems - are only indirectly revised by adding or retracting applications of certain axioms from our corpus of scientific beliefs. Therefore, the axioms of classical physics remain in place for an overwhelmingly large range of applications despite the advancement of relativistic physics and quantum mechanics. It is only those applications where the empirical phenomena resisted an explanation in terms of classical physics that had to be retracted, such as black body radiation, the perihelion of Mercury, and the speed of light in moving reference systems.

Furthermore, it is important to note that the superiority of relativistic physics or quantum mechanics over classical physics in the critical applications does not depend on the epistemic ranking of corresponding theory-elements. The

problem with classical physics here rather is that the empirical findings cannot be extended to a **T**-theoretical model. If this is the case, the antecedent of the corresponding abductive inference rule $\delta 2(\mathbf{T})$ is not satisfied so that the rule cannot successfully be applied to the stock accepted sentences. Therefore, the epistemic ranking among theory-elements need not be invoked to explain why classical physics cannot be used and is in fact not used for a theoretical account of black body radiation, the perihelion of Mercury, the speed of light in moving reference systems etc.

To find scientific examples where the epistemic ranking of theory-elements is decisive, it seems more promising to study areas in which the experimental methods are less reliable and corresponding results more tentative than in physics as canonised by textbooks and university courses. For this reason, let us have a look at a theory in some other scientific discipline, viz., *sequence analysis* in molecular biology. The rationale of (computational) sequence analysis is to establish computational measurements on which judgements of homology concerning DNA sequences can be based. Two DNA sequences are homologous if and only if they have a common ancestor sequence in evolutionary history. In other words, the goal of sequence analysis is to find biologically meaningful metrics so as to make homology inferable from a high similarity value. Putative knowledge of homology is then used to assign functions to protein sequences since proteins that are expressed by homologous DNA sequences are likely to have the same function in biochemical pathways. Such assignments are particularly important if the function of one of the proteins is not accessible to direct experimental investigation. Besides this application, assignments of function being derived from judgements about homology may serve as hypotheses that will be tested experimentally.

Yet, homology does not necessarily imply functional equivalence of the corresponding gene expression products. It is only orthologous genes where the inference to functional equivalence is considered reliable, even though not valid without exceptions. Two sequences are orthologous if and only if their divergence was caused by a speciation, that is, a divergence of lineages of organisms. Other forms of homology are gene duplication, i.e., the divergence of lineages of genes within an organismal lineage, and horizontal gene transfer, i.e., the divergence of lineages of genes by transfer across different organismal lineages (Pevsner 2003, pp. 41–86, 223–272).

The idea underlying the computational measurement of sequence similarity is to find optimal alignments of sequences and to evaluate then the number and the kind of matches and mismatches. The quality of an alignment is given by the total alignment value, which is determined by summarising over the matches

and mismatches, where mismatches usually have weights so that different types of mismatches have a different impact on the alignment value. (Matches make a positive contribution to the alignment value, whereas mismatches a negative one.) An alignment is optimal if and only if there is none with a higher total alignment value. The alignment value of the optimal alignment is taken as sequence similarity value. Different ways of weighing matches and mismatches are represented by so-called *scoring matrices* (Gusfield 1997, pp. 215-226).

Since there is a many-to-one correspondence between DNA and protein sequences, the computation of optimal alginments can also be applied to protein sequences. The latter method is usually considered more informative concerning relations of homology than the direct comparison of DNA sequences for two reasons. First, there are different triples of nucleotides that encode one and the same protein when transcribed. Second, since proteins react in biochemical pathways, they determine more directly the properties of the cells than original DNA sequences. There is a whole theory about properly aligning protein sequences, which led to so-called *PAM-matrices* as scoring schemes.

In structuralist terms, we can say that we have here several **T**-theoretical relations on the domain of sequences: homology, orthology, paralogy, functional equivalence. **T** may stand for the whole of molecular biology. Further, there is a **T**-theoretical function that assigns a similarity value to pairs of sequences. To be more precise, there are as many functions of sequence similarity as there are scoring matrices in use for DNA and protein sequences. Several theory-elements were touched upon in the preceding explanations: One that represents the inference from homology to functional equivalence ($\mathbf{T}_1$), another one representing the inference from orthology to having an equivalent function ($\mathbf{T}_2$), one saying that two DNA (protein) sequences with a sufficiently high similarity value are homologous ($\mathbf{T}_{3,d}$ ($\mathbf{T}_{3,p}$)), and, finally, theory-elements representing the computation of sequence similarity with the many different scoring schemes ($\mathbf{T}_{4,1}, \ldots, \mathbf{T}_{4,n}$). Unfortunately, there is no single theory-element that allows to ascribe the function to a DNA sequence, but rather an intricate network of inferential patterns that would and actually does take tremendous resources to unfold.

What can be said about the priority ordering of these theory-elements? Certainly, none should be assigned the highest rank in the overall ordering of molecular biology since all are only defeasibly valid. For reasons indicated above, $\mathbf{T}_2$ is better than $\mathbf{T}_1$. Further, sequence similarity of protein sequences is a more reliable indicator of homology than sequence similarity of DNA sequences, which means that $\mathbf{T}_{3,p}$ is better than $\mathbf{T}_{3,d}$. Different scoring schemes lead to differently reliable indicators of homology depending, in particular, on the temporal

distance of the presumed event of divergence in evolutionary history. Further comparative judgements about the reliability of the theory-elements considered are difficult to justify. Of course, these are only cursory remarks which may be seen as preliminary to a thorough case study.

11 Conclusion

The system outlined here has it that incoming information must be of the form $b_i \in \mathbf{I}(\mathbf{T}_j)$. How plausible is that? What does this constraint upon the logical form of incoming information mean in less formal terms? Now, it means that for an incoming piece of information we can always distinguish between some phenomenal evidence for a $\mathbf{T}$-theoretical proposition, the $\mathbf{T}$-theoretical proposition itself, and the theory $\mathbf{T}$ in light of which we interpret the phenomenal evidence.

Take, for example, the proposition ϕ saying that the degree of global warming is $1,8°$ Kelvin for the period between 1900 and 2000. Several empirical phenomena may lead us to consider adoption of the theoretical proposition ϕ. We may have read a sentence having the meaning of ϕ in a newspaper of the yellow press. Alternatively, we may have read such a sentence in a scientific journal. Finally, we may be the one who conducted the by no means trivial statistical study from which ϕ resulted. Little consideration is necessary to see that the computation of the degree of global warming is anything but straightforward. In any case, some non-empty set of universal axioms and inferential patterns is used to infer proposition ϕ from empirical data. Even in the case of the newspaper, there is an inference recognisable from the proposition that I read a sentence with the meaning that ϕ to ϕ itself. In the structuralist representation scheme, such inferential patterns are represented by theory-elements.

In the belief revision literature it has been observed that the reliability of the belief forming process through which we came to accept a proposition ϕ matters for the epistemic standing of ϕ. In light of this observation, the standard notation for belief revisions $A * \phi$ appears to be an oversimplification because it does neither represent the belief forming processes through which we came to accept the elements of A nor the process driving us to accept a new proposition ϕ that is potentially inconsistent with the set of previously accepted ones. The present system, by contrast, *forces* one to represent the belief forming processes, at least for derived beliefs. Observations concerning the relation between reliability and acceptance, which had to be made in the informal explanations in the belief revision literature so far, are thus coming into the reach of a formal

representation.

In essence, the present system attempts to account for the view that the epistemic ranking of propositions is an effect of theorising and that the ranking of theory-elements can be interpreted, at least for a certain range of cases, in statistical terms. If this view is correct, non-derived empirical data are accepted with equal firmness without qualification. The ranking of derived beliefs ϕ depends then on the ranking of the $\mathbf{T}$-non-theoretical premisses and on the ranking of the theory-element $\mathbf{T}$ that was used to derive ϕ. A few minor emendations are necessary, however, to take into account that $\mathbf{T}$-non-theoretical premisses coming with an intended application of $\mathbf{T}$ need not be pure empirical data but rather may be derived by another theory-element $\mathbf{T}'$.[13] Being careful, we should say that the present system is about having the resources to keep track of justifications that are effective for derived beliefs.

References

Andreas, H. (2010a). A Modal View of the Semantics of Theoretical Sentences. Synthese **174**(3): 367–383.

Andreas, H. (2010b). New Account of Empirical Claims in Structuralism. Synthese **176**(3): 311–332.

Antoniou, G. (1997). Nonmonotonic Reasoning. MIT Press, Cambridge, Mass.

Balzer, W., Moulines, C. U., and Sneed, J. (1987). An Architectonic for Science. The Structuralist Program. D. Reidel Publishing Company, Dordrecht.

Bartelborth, T. (2002). Explanatory Unification. Synthese **130**(1): 91–108.

Brewka, G. (1989). Preferred Subtheories: An Extended Logical Framework for Default Reasoning. In: Proceedings of Eleventh International Joint Conference on Artificial Intelligence, Detroit, Morgan Kaufmann, San Mateo, CA, 1043–1048.

Brewka, G. (1991). Belief Revision in a Framework for Default Reasoning. In: Proceedings of the Workshop on The Logic of Theory Change, Springer, London, 602–622.

[13] The problem lying behind is that, in the present exposition, it is assumed that intended applications are elements of the belief base, even though they may contain derived beliefs. Strictly speaking, we must say that the present exposition is confined to one level of theorising. To remedy the situation, *partial intended applications* must be introduced. We did not include this to avoid details being not necessary for an understanding of the basic ideas.

Brewka, G., Niemelä, I., and Truszczyński, M. (2008). Nonmonotonic Reasoning. In: F. v. Harmelen, V. Lifschitz, and B. Porter (eds.) Handbook of Knowledge Representation, Elsevier, Amsterdam.

Flach, P. A. (2000). On the Logic of Hypothesis Generation. In: P. A. Flach and A. C. Kakas (eds.) Abduction and Induction, Kluwer Academic Publishers, 89–106.

Friedman, M. (1974). Explanation and Scientific Understanding. The Journal of Philosophy **71**: 1–19.

Gärdenfors, P. (1988). Knowledge in Flux. Cambridge, Mass.

Gärdenfors, P. and Makinson, D. (1988). Revision of Knowledge Systems Using Epistemic Entrenchment. In: M. Vardi (ed.) TARK' 88 - Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, Morgan and Kaufmann, Los Altos, 83–95.

Goldman, A. (1979). What is Justified Belief? In: G. S. Pappas (ed.) Justification and Knowledge, Reidel, Dordrecht, 1–23.

Gottlieb, D. and McCarthy, T. (1979). Substitutional Quantification and Set Theory. Joumal of Philosophical Logic **8**: 315–331.

Gusfield, D. (1997). Algorithms on Strings, Trees, and Sequences. Computer Science and Computational Biology. Cambridge Press, Cambridge.

Hansson, S. O. (1999). A Textbook of Belief Dynamics. Theory Change and Database Updating. Kluwer, Dordrecht.

Hofer, L. (2009). Holistische Elemente in der strukturalistischen Wissenschaftskonzeption. M. A. Thesis, LMU Munich.

Kitcher, P. (1976). Explanation, Conjunction and Unification. Journal of Philosophy **73**: 207–212.

Kuhn, T. S. (1976). Theory-Change as Structure-Change: Comments on the Sneed Formalism. Erkenntnis **10**(2): 179–199.

Lifschitz, V. (1990). On Open Defaults. In: J. Lloyd (ed.) Proceedings of the Symposium on Computational logic, Springer, Berlin, 80–95.

Minsky, M. A. (1974). Framework for Representing Knowledge. AI Laboratory, Massachusetts Institute of Technology, http://web.media.mit.edu/ minsky/papers/Frames/frames.html.

Pevsner, J. (2003). Bioinformatics and Functional Genomics. Wiley & Sons, Hoboken, NJ.

Rott, H. (2001). Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning. Oxford University Press, Oxford.

Schurz, G. (2008). Patterns of Abduction. Synthese **164**: 201–234.

Sneed, J. (1979). The Logical Structure of Mathematical Physics. D. Reidel Publishing Company, Dordrecht, 2nd edn.

Spohn, W. (1988). Ordinal Conditional Functions: A Dynamic Theory of Epistemic States. In: Causation in Decision, Belief Change, and Statistics II, Kluwer, Dordrecht, 105–134.

Stegmüller, W. (1985). Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Bd. 2 - Theorie und Erfahrung, Teil D. Springer-Verlag, Berlin-Heidelberg-New York.

Wallace, J. (1971). Convention T and Substitutional Quantification. Noûs **5**(2): 199–211.