Possible World Semantics for Analogous Reasoning

J.-J.Ch. Meyer and J.C. van Leeuwen

UU-CS-1995-38
December 1995

Utrecht University

Department of Computer Science

Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

Possible World Semantics for Analogous Reasoning

J.-J.Ch. Meyer and J.C. van Leeuwen

Technical Report UU-CS-1995-38
December 1995

Department of Computer Science
Utrecht University
P.O.Box 80.089
3508 TB Utrecht
The Netherlands

ISSN: 0924-3275

Possible World Semantics for Analogous Reasoning

J.-J.Ch. Meyer
Utrecht University

J.C. van Leeuwen
Eindhoven University of Technology

Abstract

Analogous reasoning is a form of reasoning that is often used in daily life situations. It is also a form of reasoning that appears in certain AI applications such as learning and knowledge acquisition. Mostly a kind of quantitative notion of (dis)similarity is employed. In this paper we present a modal model for a qualitative notion of similarity and thus we obtain a basis for qualitative analogous reasoning.

1 Introduction

Analogies are ubiquitous in common-sense situations. Often we reason by analogy to predict the outcome of a situation at hand on the basis of a *similar* case we have encountered in the past. In fact, reasoning by analogy is the very hart of learning by intelligent agents. By matching new cases to familiar ones we may extend our knowledge by transposing (assuming or investigating) what we know to hold in the old case with what it would correspond to in the new case. Analogies are used in a wide spectrum of cases ranging from poetry where metaphors are employed to explain things in other (more familiar or expressible) terms via science where models of familiar notions may guide the exploration of new concepts to even such a rigourous discipline as mathematics, where in proofs often a phrase is used such as "we have proven the case in detail for such and such; the case so and so is analogous".

Not surprisingly, also in AI reasoning by analogy has an important role. For instance, it is employed for automatic (machine) learning and for classifying newly obtained knowledge into the scheme of knowledge so far in knowledge acquisition. In the literature we find many classifications and variants of reasoning by analogy (e.g. transformational and derivational analogy, bottom up analogy (Evans 68, Winston 80), top down analogy (Burstein 86)) as well as a number of programs that are available (cf. Hall 89) and perform some form of analogical reasoning in a concrete context (e.g. CARL in the context of learning

assignment statements in BASIC (Burstein 86, 88), a full treatment of which is beyond the scope of this paper. We will focus on the following dichotomy proposed by Indurkhya. (Actually, Indurkhya also distinguishes a third form, proportional analogy, which is of a slightly different nature and which we shall ignore here. In (van Leeuwen 95) it is indicated how also this form of analogous reasoning can be fitted into our model framework.)

Indurkhya distinguishes the following types of analogy: *analogy by rendition* and *predictive analogy* (Indurkhya 89).

Analogy by rendition (translation) views a situation (target) or object as if it were another (source). Elements are mapped and translated. It is a way of interpreting a target situation in the light of a known source situation in order to gain more or different information about the target or to get a better understanding of it. This is accomplished by projecting 'framework' and terminology from source to target. The source domain may be an artificial one, a model in which one may focus on the relevant level of abstraction. It is thus closely related to the use of models in problem solving and metaphors. Poets use this kind of analogy all the time and designers obtain creative ideas through analogy by rendition.

Predictive analogy is also based on rendition, but goes further than just stating renditions or similarities between two domains that are both known completely: when a rendition is possible between two domains predictions are made about more similarities by considering relations on the already mapped elements. In this context rendition must agree on the ontology of elements in both domains. The emphasis here is on making an *inference* (prediction) about the target domain on the basis of what is known about a usually more familiar source domain. This form of analogical reasoning is mostly considered in AI applications.

However, the distinction between analogy by rendition and predictive analogy is not entirely clear-cut, since it depends on what is known exactly about the (source and particularly the) target domain whether a conclusion is a prediction or rather a mere rendition. Below we shall consider rendition on the basic elements of the logic at hand (atomic propositions in the propositional case), while we view prediction as the result following from this rendition (translation) to more complex formulas. Of course, if both domains are completely known, these 'predictions' are then not much more than simple translations / renditions. Later in the paper we will see how we can extend this idea to a first-order language (and logic), where we can express more refined notions of rendition and prediction. Also the idea that some predictions are not completely certain and are some more or less 'educated guesses' will be discussed and treated formally in the paper.

In this paper we shall give a semantical treatment of the above two forms of reasoning, put into a possible world framework. Proofs of propositions and theorems are omitted here; those of sections 2 and 3 can be found in (van Leeuwen 95).

2 A Propositional Modal Logic of Analogy

2.1 Defining the concept of similarity

In this section we will try to identify some key concepts of analogical reasoning within the context of propositional logic. The most important of these concepts is that of *similarity*. It is important to note that we will develop a *semantical* theory of these concepts. We consider this semantics-based treatment of analogical reasoning one of the main contributions of our paper, which is lacking in most approaches in the literature (a notable exception is (Thiele 86)). Furthermore, we embed our theory into a possible world semantics. For this modal approach we were influenced by (Morgan 79), who was in his turn inspired by early work by C.S. Peirce. As compared to the work of Morgan, we make the model much more explicit, both with respect to the modal aspect and the aspect of the rendition mapping.

From a practical stand-point propositional logic is clearly too 'poor' as to expressive power to enable one to represent 'real-life' examples of analogical reasoning: viewing propositional logic as predicate logic without function, relation, and constant symbols, obviously there is little room for similarity between models left, because of the sheer simplicity of these models. However, we believe that the simple setting of propositional logic enables one to concentrate on some important issues without having the to deal with the complexity of richer logics. The results of attempting to describe analogical reasoning in this setting will also serve as a natural basis for our further development of the theory in the sequel of this paper.

In commonsense use, analogical reasoning manifests itself between two domains of knowledge, which may be represented by formal theories. The very essence of analogical reasoning indicates that these two domains between which it takes place should display some form of correspondence or *similarity*. When looked upon semantically this means that the models of these theories should show some form of similarity as well. In our simple propositional setting the only way to express this similarity is to consider these on the level of propositional atoms. We will do this in the most simple way conceivable: as a (*similarity*) mapping $\mathbf{T}$ from propositional atoms from the one domain (which we shall call the *source domain*) to the other (the *target domain*), representing that the propositional atom $\mathbf{T}(p)$ in the target domain is similar to the propositional atom p in the source domain (as far as the context of reasoning at hand is concerned). In fact, this function $\mathbf{T}$ may be viewed as the (formal counterpart of the) translation mapping (*rendition*) in the rendition type of analogy mentioned in the introduction. Of course, one might also consider more general similarity mappings $\mathbf{T}$ from *formulas* in $\mathcal{L}(\mathcal{P}_1)$ to *formulas* in $\mathcal{L}(\mathcal{P}_2)$ as primitive, but in our view this yields a rather non-compositional theory in the sense that it is then not clear at all how to determine the similarity mapping for complex formulas of which the function $\mathbf{T}$ is not given.

To make a start with our formal treatment, we assume two sets of propositional atoms $\mathcal{P}_1$ and $\mathcal{P}_2$, which, for convenience, we assume to be disjoint. On the basis of a set $\mathcal{P}$ of propositional atoms we construct a propositional language $\mathcal{L}(\mathcal{P})$ as usual: the smallest set containing $\mathcal{P}$ and closed under the propositional connectives $\neg$, $\wedge$ and $\vee$ (and other connectives which may be introduced as abbreviations in terms of these, such as $\rightarrow$). We furthermore use $\top$ and $\bot$ to denote the constants for 'truth' and 'falsehood', respectively. The symbols φ and ψ are used as metavariables for formulas in a propositional language. Unless stated otherwise, we shall use the language $\mathcal{L}(\mathcal{P}_1)$ for the description of the source domain and $\mathcal{L}(\mathcal{P}_2)$ for that of the target domain. As usual in propositional logic we describe the semantics of formulas by means of *valuations*. We use tt for the truth value true and ff for the truth value false.

A propositional model over the set $\mathcal{P}$ of propositional atoms is a valuation function v with $v : \mathcal{L}(\mathcal{P}) \rightarrow \{tt, ff\}$ (induced by a function $v : \mathcal{P} \rightarrow \{tt, ff\}$). In this propositional context we may call $\mathcal{P}$ the *signature* of model v. The class of valuations (propositional models) over $\mathcal{P}$ is denoted $VAL(\mathcal{P})$.

It may not be necessary or even desirable to map all of the propositions because some elements may be irrelevant. Therefore we choose $\mathbf{T}$ to be a *partial* function. We will further assume this function to be injective, since this will enable us to speak about its inverse later on. We do not assume surjectivity of the function $\mathbf{T}$, since it might well be that the target domain has some elements (viz. propositional atoms) that do not correspond (have no counterpart) in the source domain. First we define the concept of *rendition* as discussed above, which is similarity on the smallest elements (atomic propositions):

Assuming an injective partial function $\mathbf{T} : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ with $dom(\mathbf{T}) \neq \emptyset$, the *similarity mapping* $R_{\mathbf{T}}$ from $VAL(\mathcal{P}_1)$ to models $VAL(\mathcal{P}_2)$ induced by $\mathbf{T}$ is given by:

$$\forall\, v_1 \in VAL(\mathcal{P}_1), v_2 \in VAL(\mathcal{P}_2)\, \forall\, p \in \mathcal{P}_1 \cap dom(\mathbf{T}) : R_{\mathbf{T}}(v_1)(\mathbf{T}(\mathcal{P}_1)) = v_1(\mathcal{P}_1)$$

We can lift $\mathbf{T}$ to formulas which will create a base for the *predictive* part of analogy. We just define:

- $\mathbf{T}(\varphi \wedge \psi) = \mathbf{T}(\varphi) \wedge \mathbf{T}(\psi)$
- $\mathbf{T}(\varphi \vee \psi) = \mathbf{T}(\varphi) \vee \mathbf{T}(\psi)$
- $\mathbf{T}(\varphi \rightarrow \psi) = \mathbf{T}(\varphi) \rightarrow \mathbf{T}(\psi)$
- $\mathbf{T}(\neg\varphi) = \neg\mathbf{T}(\varphi)$

Proposition 1 *The above map* $\mathbf{T} : L(\mathcal{P}_1) \rightarrow L(\mathcal{P}_2)$ *induced by a similarity mapping* $\mathbf{T} : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ *satisfies the property: for all* $v \in VAL(\mathcal{P}_1)$ *we have:*

$$\forall\, \varphi \in L(\mathcal{P}_1) \cap dom(\mathbf{T}) : v \models \varphi \iff R_{\mathbf{T}}(v) \models \mathbf{T}(\varphi)$$

$\Diamond$

2.2 Kripke models and modalities for analogous reasoning

To build a modal logic of analogical reasoning based on similarity mappings, we start with a notion of Kripke model tailored for this purpose. We assume $\mathcal{P} \neq \emptyset$ to be the universe of propositional atoms. For convenience we consider partial valuations v over $\mathcal{P}$, which means that the function $v : \mathcal{P} \rightarrow \{tt, ff\}$) (and thus also the function $v : \mathcal{L}(\mathcal{P}) \rightarrow \{tt, ff\}$) is partial and may not be defined for all arguments. The set of partial valuations over $\mathcal{P}$ is denoted VAL. To simplify notation we will use these valuations directly as our set of worlds. Thus our set of worlds are exactly the valuations over $\mathcal{P}$. The use of *partial* valuations enables us to effectively vary the domain of propositional atoms defined in a world without having to bother with distinct sets of propositional atoms per world. For a valuation v we denote its domain (i.e. the set of propositional variables on which v is defined) by $dom(v)$.

For the accessibility relations we use relations induced by similarity mappings $VAL \rightarrow VAL$: given a set $\{\mathbf{T}_i | i = 1, ..., n\}$ of partial functions $\mathcal{P} \rightarrow \mathcal{P}$, we define $R_{\mathbf{T}_i} \subseteq VAL \times VAL$ (overloading notation slightly) as:

$$R_{\mathbf{T}_i}(v_1, v_2) \Leftrightarrow R_{\mathbf{T}_i}(v_1) = v_2$$

Note that since we have stipulated that $R_{\mathbf{T}_i}$ is a partial injective function we can define its inverse function $R^{-1}_{\mathbf{T}_i}$ as:

$$R^{-1}_{\mathbf{T}_i}(v_1, v_2) \Leftrightarrow R_{\mathbf{T}_i}(v_2, v_1)$$

Naturally, $R^{-1}_{\mathbf{T}_i}$ is the accessibility relation associated with the inverse $\mathbf{T}_i^{-1}$ of the similarity mapping $\mathbf{T}_i$, thus $R^{-1}_{\mathbf{T}_i} = R_{\mathbf{T}_i^{-1}}$.

Our notion of Kripke model now comes down to the following. A (simplified) Kripke model is an ordered tuple $\mathcal{M} = (VAL, \{R_{\mathbf{T}_i} | i = 1, ..., n\})$.

On the basis of this notion of a Kripke model, we introduce a modal language which we can evaluate with these models. This modal language consists of the propositional language over the propositional atoms $\mathcal{P}$ together with a clause for modal operators $\Box_{\mathbf{T}_i}$. The latter are interpreted on a Kripke model $\mathcal{M} = (VAL, \{R_{\mathbf{T}_i} | i = 1, ..., n\})$ as follows:

$$\mathcal{M}, v \models \Box_{\mathbf{T}_i}\varphi \Leftrightarrow \textit{for all } w \textit{ with } R_{\mathbf{T}_i}(v, w) : \mathcal{M}, w \models \varphi.$$

We can also introduce further modalities derived from these $\Box_{\mathbf{T}_i}$, viz. $\Box^{-1}_{\mathbf{T}_i}$, $\Box_i$, $\Box^*_i$, $\Box$ and $\Box^*$, based on the (derived) relations $R^{-1}_{\mathbf{T}_i}$, $R_i = R_{\mathbf{T}_i} \cup R^{-1}_{\mathbf{T}_i}$, R^*_i, the transitive, reflexive closure of the relation R_i, $R = \bigcup_{i=1}^{n} R_i$, and R^*, the transitive, reflexive closure of the relation R, respectively. So, for instance,

$$\mathcal{M}, v \models \Box^*_i\varphi \Leftrightarrow \textit{for all } w \textit{ with } R^*_i(v, w) : \mathcal{M}, w \models \varphi,$$

and similarly for the other operators. Also we may use the duals $\Diamond_{\mathbf{T}_i}$, $\Diamond_{\mathbf{T}_i}^{-1}$, $\Diamond_i$, $\Diamond_i^*$, $\Diamond$ and $\Diamond^*$ of these operators, defined as usual.

These operators enable us to express properties of analogies transcending just the relation between a source and a target domain. For instance, note that the relation R_i^* yields the equivalence class (*analogy class*) associated with similarity mapping $\mathbf{T}_i$, i.e. all models that are 'analogous with respect to $\mathbf{T}_i$'. Thus, the modality $\Box_i^*$ states something about what is common between domains that are related on the basis of $\mathbf{T}_i$. The modality $\Box^*$ states even more general properties, viz. those common to all domains that are related with respect to *some* similarity mapping. (Perhaps these properties are even too general to be useful, but this may depend on the context.)

As usual we define validity of a formula φ in a model $\mathcal{M}$, denoted $\mathcal{M} \models \varphi$, by $\mathcal{M}, v \models \varphi$ for all valuations v in $\mathcal{M}$, and validity of φ, denoted $\models \varphi$, by $\mathcal{M} \models \varphi$ for all models $\mathcal{M}$.

Since this gives us a normal modal logic in the sense of (Chellas 80), the modal operators above satisfy the K-axiom:

$$\models \Box_{\mathbf{T}_i}\varphi \wedge \Box_{\mathbf{T}_i}(\varphi \rightarrow \psi) \rightarrow \Box_{\mathbf{T}_i}\psi$$

(If in every i-similar world both φ and $\varphi \rightarrow \psi$ holds, then in all those worlds ψ holds.)

and the necessitation rule:

$$\models \varphi \Rightarrow \models \Box_{\mathbf{T}_i}\varphi$$

Moreover, we can now directly put Proposition 1 in modal terms:

Theorem 1

$$\models \varphi \leftrightarrow \Box_{\mathbf{T}_i}\mathbf{T}_i(\varphi)$$

$\Diamond$

This theorem states precisely how rendition can be obtained by considering a similar world.

The other modal operators satisfy the following validities (Here $\Box_{(i)}^*$ stands for a modal operator in the set $\{\Box^*, \Box_i^*\}$):

- $\models \Box_i\varphi \leftrightarrow \Box_{\mathbf{T}_i}\varphi \wedge \Box_{\mathbf{T}_i}^{-1}\varphi$
- $\models \varphi \rightarrow \Box_{\mathbf{T}_i}\Diamond_{\mathbf{T}_i}^{-1}\varphi$
- $\models \varphi \rightarrow \Box_{\mathbf{T}_i}^{-1}\Diamond_{\mathbf{T}_i}\varphi$
- $\models \Box_{(i)}^*\varphi \rightarrow \varphi$

- $\models \Box^*_{(i)}\varphi \rightarrow \Box_{(i)}\Box^*_{(i)}\varphi$
- $\models \Box^*_{(i)}(\varphi \rightarrow \Box_{(i)}\varphi) \rightarrow (\varphi \rightarrow \Box^*_{(i)}\varphi)$
- $\models \Box\varphi \leftrightarrow (\Box_1\varphi \wedge \ldots \wedge \Box_n\varphi)$

This follows directly from the definition of the relations that are associated with these operators.

Moreover, since always R_i and R are symmetrical and R^* is an equivalence relation, independent of the properties of the relation $R_{\mathbf{T}_i}$ itself, we also immediately have the following validities:

1. $\models \varphi \rightarrow \Box_i\Diamond_i\varphi$
2. $\models \varphi \rightarrow \Box\Diamond\varphi$
3. $\models \varphi \rightarrow \Box^*_{(i)}\Diamond^*_{(i)}\varphi$
4. $\models \Box^*_{(i)}\varphi \rightarrow \Box^*_{(i)}\Box^*_{(i)}\varphi$
5. $\models \Diamond^*_{(i)}\varphi \rightarrow \Box^*_{(i)}\Diamond^*_{(i)}\varphi$

2.3 Analogical inferences

We have seen above how the concept of analogy by rendition can be formalized by means of our similarity mappings. We now show how we can also formulate true analogical reasoning in the sense of making inferences in our setting.

Let $\mathcal{R}$ be an inference: $\mathcal{R} = \varphi_1 \vdash \varphi_2 \vdash \ldots \vdash \varphi_m$. Now we would like to make an 'analogous' reasoning in another domain: $\mathcal{R}' = \varphi'_1 \vdash \varphi'_2 \vdash \ldots \vdash \varphi'_m$.

By using our similarity mappings and associated modal operators we can now do this in a formal way. Let us say that the similarity mapping involved is $\mathbf{T}$. Then we would expect that the 'source' inference $\mathcal{R} := \varphi_1 \vdash \varphi_2 \vdash \ldots \vdash \varphi_m$ could be transformed into: $\mathcal{R}' = \mathbf{T}(\varphi_1) \vdash \mathbf{T}(\varphi_2) \vdash \ldots \vdash \mathbf{T}(\varphi_m)$. That this is indeed the case is justified by the following derived rule:

$$\frac{\varphi \rightarrow \psi}{\Box_{\mathbf{T}}\mathbf{T}(\varphi) \rightarrow \Box_{\mathbf{T}}\mathbf{T}(\psi)}$$

Via this rule we can now reason as follows: suppose $\varphi \vdash \psi$. Then by the deduction theorem of classical logic we obtain $\vdash \varphi \rightarrow \psi$, and so by the above rule, $\vdash \Box_{\mathbf{T}}\mathbf{T}(\varphi) \rightarrow \Box_{\mathbf{T}}\mathbf{T}(\psi)$, and hence $\Box_{\mathbf{T}}\mathbf{T}(\varphi) \vdash \Box_{\mathbf{T}}\mathbf{T}(\psi)$. This is indeed the formal statement of the 'translated' inference above. Note, moreover, that our modal framework exactly pin-points the worlds where this analogous reasoning takes place. If one did not have this, one would be forced to use a 'semantically polluted' inference rule (so not really an inference rule at all) such as

$$\frac{v \models \varphi \rightarrow \psi}{w \models \mathbf{T}(\varphi) \rightarrow \mathbf{T}(\psi)}$$

where $w = R_{\mathbf{T}}(v)$.

3 Extension to First-Order Logic

If we want to extend our propositional logic to a first-order one, we need to enrich the structure of our models, and consequently redefine our notion of similarity between models. Here we build on work by (Thiele 86) but again provide for a possible world semantics. The way we consider similarity mappings in the first-order case is also reminding of work done on so-called *interpretability logics* in a completely different context, viz. in metamathematics for the proof of consistency and undecidability of mathematical theories (cf. e.g. (Tarski, Mostowski & Robinson 53)).

3.1 Similarity in a first-order setting

First of all, we extend our propositional language to a first-order one: we assume a set VAR of variables, a set $FUNC$ of function symbols and a set $PRED$ of predicate symbols. As usual, function and predicate symbols have an arity associated with them determining the number of arguments. (0-ary function symbols are called constants; 0-ary predicates are called atomic propositions.) We call a pair $\Sigma = (FUNC, PRED)$ a *signature*. A signature $\Sigma_1 = (FUNC_1, PRED_1)$ is a *subsignature* of $\Sigma_2 = (FUNC_2, PRED_2)$ if $FUNC_1 \subseteq FUNC_2$ and $PRED_1 \subseteq PRED_2$. We can also speak about the intersection $\Sigma_1 \sqcap \Sigma_2$ of two signatures Σ_1 and Σ_2 in the obvious way (just take the intersections of the sets of function symbols and of the sets of predicate symbols).

The set $TERM(VAR, FUNC, PRED)$, or just abbreviated $TERM$, is the minimal set containing VAR and closed under the construction $g(t_1, ..., t_n)$ for function symbols $g \in FUNC$ and $t_i \in TERM$ $(i = 1, ..., n)$ where the arity of g is n. The set $AT(VAR, FUNC, PRED)$, or AT, of atomic formulas is given as the smallest set closed under the constructions

- $P(t_1, ..., t_n)$ for function symbols $P \in PRED$ and $t_i \in TERM$ $(i = 1, ..., n)$ where the arity of P is n, and
- $t_1 = t_2$ for $t_1, t_2 \in TERM$.

The set $\mathcal{L}(VAR, FUNC, PRED)$, usually abbreviated $\mathcal{L}$, of first-order formulas is the minimal set closed under the classical connectives and the construction $\forall x\ \varphi$ and $\exists x\ \varphi$ with $x \in VAR$ and $\varphi \in \mathcal{L}$, and containing the set AT of atomic formulas. We denote the set of free variables of a formula φ by $FV(\varphi)$.

As usual, a language $\mathcal{L}(VAR, FUNC, PRED)$ over signature $\Sigma = (FUNC, PRED)$ is interpreted on Σ-structures of the form $w = (\mathcal{A}, \Sigma, \Phi, \Pi)$, where $\mathcal{A}$ is a domain of values for the interpretation of the variables and constants, Φ

is a function such that, for all $g \in FUNC$, $\Phi : g \mapsto (\mathcal{A}^n \rightarrow_{part} \mathcal{A})$ where n is the arity of g, and Π is a function such that, for all $P \in PRED$, $\Pi : P \mapsto (\mathcal{A}^n \rightarrow_{tot} \{tt, ff\})$ where n is the arity of P. (Here $X \rightarrow_{part} Y$ and $X \rightarrow_{tot} Y$ stand for the classes of partial and total functions from X to Y, respectively.) When convenient we may also denote $\Pi(P)$ as a subset of $\mathcal{A}^n$. We denote the class of Σ-structures as $STRUCT(\Sigma)$, while the class of Σ-structures with fixed domain $\mathcal{A}$ is denoted $STRUCT(\mathcal{A}, \Sigma)$. We omit the standard clauses for the interpretation of the language on these structures, since they can be found in any textbook on logic (such as e.g. (van Dalen 89)).

Given a structure $w = (\mathcal{A}, \Sigma, \Phi, \Pi)$, with $\Sigma = (FUNC, PRED)$. Let $\mathcal{A}_0$ be another domain and let $\Sigma_0 = (FUNC_0, PRED_0)$ be another signature. We define the substructure $w \downarrow \mathcal{A}_0, \Sigma_0$ of w as the structure $(\mathcal{A} \cap \mathcal{A}_0, \Sigma \sqcap \Sigma_0, \Phi \downarrow \mathcal{A}_0, \Sigma_0, \Pi \downarrow \mathcal{A}_0, \Sigma_0)$, where $\Phi \downarrow \mathcal{A}_0, \Sigma_0$ is a function interpreting only the function symbols in $FUNC \cap FUNC_0$ in the domain (and range) $\mathcal{A} \cap \mathcal{A}_0$, and similarly for $\Pi \downarrow \mathcal{A}_0, \Sigma_0$.

In this more refined set-up we can also be more precise about the similarity mapping. Instead of just stipulating a mapping from atomic formulas to atomic formulas, we now define a mapping between signatures $\Sigma_1 = (FUNC_1, PRED_1)$ and $\Sigma_2 = (FUNC_2, PRED_2)$ where we assume $FUNC_1, FUNC_2 \subseteq FUNC$ and $PRED_1, PRED_2 \subseteq PRED$. In fact, we shall define a similarity mapping based on a signature *isomorphism*:

A pair (T^1, T^2) is a signature isomorphism from $\Sigma_1 = (FUNC_1, PRED_1)$ to $\Sigma_2 = (FUNC_2, PRED_2)$ if

1. T^1 is a bijective, arity-preserving mapping from $FUNC_1$ to $FUNC_2$.
2. T^2 is a bijective, arity-preserving mapping from $PRED_1$ to $PRED_2$.

We can extend a signature isomorphism to a function T on $TERM$ and AT as follows:

Let $T^1 : FUNC \rightarrow FUNC'$ and $T^2 : PRED \rightarrow PRED'$ be arbitrary functions, then we define $T : AT(VAR, FUNC, PRED) \rightarrow AT(VAR, FUNC', PRED')$ on atoms as follows:

- $T(x_i) = x_i$ for $x_i \in VAR$
- $T(f) = T^1(f)$ for $f \in FUNC$
- $T(P) = T^2(P)$ for $P \in PRED$
- $T(f(t_1, ..., t_n)) = T(f)\ (T(t_1), ..., T(t_n))$ if f has arity n
- $[T(t_1 = t_2)] = [T(t_1) = T(t_2)]$
- $T(P(t_1, ..., t_i)) = T(P)\ (T(t_1), ..., T(t_n))$ if P has arity n

An isomorphism between signatures together with a bijection between the domains of the structures based on these signatures induces a similarity mapping between these structures .

Assuming a signature isomorphism (T^1, T^2) from $\Sigma = (FUNC, PRED)$ to $\Sigma' = (FUNC', PRED')$, and a bijective function T^0 from the domain $\mathcal{A}$ to the domain $\mathcal{A}'$, the *similarity mapping* $R_{\mathbf{T}}$ from $STRUCT(\mathcal{A}, \Sigma)$ to structures $STRUCT(\mathcal{A}', \Sigma')$ induced by $\mathbf{T} = (T^0, T^1, T^2)$ is given by:

for all $w = (\mathcal{A}, \Sigma, \Phi, \Pi) \in STRUCT(\mathcal{A}, \Sigma), R_{\mathbf{T}}(w) = (\mathcal{A}', \Sigma', \Phi', \Pi') \in STRUCT(\mathcal{A}', \Sigma')$ satisfying

1. $\Phi(f)(a_1, ..., a_n)$ exists $\Leftrightarrow$ $\Phi'(T^1(f))(T^0(a_1), ..., T^0(a_n))$ exists, and
   $T^0(\Phi(f)(a_1, ..., a_n)) = \Phi'(T^1(f))((T^0(a_1), ..., T^0(a_n))$
   for all $f \in FUNC$ and $(a_1, ..., a_n) \in \mathcal{A}^n$
   Note: for constants this means $T^0(\Phi(c)) = \Phi'(T^1(c))$

2. $\Pi(P)(a_1, ..., a_n) = \Pi'(T^2(P))(T^0(a_1), ..., T^0(a_n))$
   for all $P \in PRED$ and $(a_1, ..., a_n) \in \mathcal{A}^n$

Let $\overline{t}^w[a_1/x_1, ..., a_n/x_n]$ be the interpretation of the term $t(x_1, ..., x_n)$ in structure w under the valuation $x_1 \mapsto a_1, ...x_n \mapsto a_n$. (This can be defined inductively, which we omit here.)

Lemma 1 *Let $w \in STRUCT(\mathcal{A}, \Sigma)$ and $w' = R_{\mathbf{T}}(w)$ for $\mathbf{T} = (T^0, T^1, T^2)$. For any $a_1, ...a_n \in \mathcal{A}$ and any term t:*

$$T^0(\overline{t}^w[a_1/x_1, ..., a_n/x_n]) = \overline{T(t)}^{w'}[T^0(a_1)/x_1, ..., T^0(a_n)/x_n]$$

$\Diamond$

In the sequel of this paper we employ the notation $w \models \varphi\ [a_1/x_1, ..., a_n/x_n]$ (for $a \in \mathcal{A}$) meaning that φ is true in w under the valuation $x_1 \mapsto a_1, ...x_n \mapsto a_n$.

The translation lemma above for terms gives rise to a first-order version of rendition of atomis formulas as put in the following proposition.

Proposition 2 *Let $\Sigma = (FUNC, PRED)$ and $\Sigma' = (FUNC', PRED')$. Suppose $w \in STRUCT(\mathcal{A}, \Sigma)$, and $w' = R_{\mathbf{T}}(w) \in STRUCT(\mathcal{A}', \Sigma')$ for $\mathbf{T} = (T^0, T^1, T^2)$. Let T be the function $T : AT(VAR, FUNC, PRED) \rightarrow AT(VAR, FUNC', PRED')$ from the above definition, in terms of T^1 and T^2.*
Then for any $a_1, ..., a_n \in \mathcal{A}$ and $P(x_1, ..., x_n) \in AT(VAR, FUNC, PRED)$:
$w \models P(x_1, ..., x_n)\ [a_1/x_1, ..., a_n/x_n] \iff$
$w' \models T(P)(x_1, ..., x_n)\ [T^0(a_1)/x_1, ..., T^0(a_n)/x_n]$ $\Diamond$

T can be lifted to formulas $\varphi \in \mathcal{L}(VAR, FUNC, PRED)$ (providing a base for prediction again) as follows:

- $T(\forall x_1 \varphi(x_1, ..., x_n)) = \forall x_1\ T(\varphi)(x_1, ..., x_n)$
- $T(\exists x_1 \varphi)(x_1, ..., x_n)) = \exists x_1\ T(\varphi)(x_1, ..., x_n)$
- $T(\varphi \vee \psi) = T(\varphi) \vee T(\psi)$
- $T(\varphi \wedge \psi) = T(\varphi) \wedge T(\psi)$
- $T(\varphi \rightarrow \psi) = T(\varphi) \rightarrow T(\psi)$
- $T(\neg\varphi) = \neg T(\varphi)$

Proposition 3 *Let $\Sigma = (FUNC, PRED)$ and $\Sigma' = (FUNC', PRED')$. Suppose $w \in STRUCT(\mathcal{A}, \Sigma)$, and $w' = R_{\mathbf{T}}(w) \in STRUCT(\mathcal{A}', \Sigma')$ for $\mathbf{T} = (T^0, T^1, T^2)$. Let T be the function $T : AT(VAR, FUNC, PRED) \rightarrow AT(VAR, FUNC', PRED')$ from the above definition, in terms of T^1 and T^2.*
then for any formula $\varphi \in \mathcal{L}(VAR, FUNC, PRED)$ with $FV(\varphi) = \{x_1, ..., x_n\}$ and any $a_1, ..., a_n \in \mathcal{A}$

$$w \models \varphi\ [a_1/x_1, ..., a_n/x_n] \iff w' \models T(\varphi)\ [T^0(a_1)/x_1, ..., T^0(a_n)/x_n]$$

In particular, if $FV(\varphi) = \emptyset$ then

$$w \models \varphi \iff w' \models T(\varphi)$$

$\Diamond$

3.2 First-order modal logic of analogy

We now enhance our Kripke models to cater for our first-order language. Essentially, worlds are changed from simple valuations on propositional atoms to first-order structures containing a signature $(FUNC, PRED)$ and a domain of interpretation for the variables and constants.

For the accessibility relations we use relations induced by similarity mappings $STRUCT(\mathcal{A}, \Sigma)$ to structures $STRUCT(\mathcal{A}', \Sigma')$: given a set $\{\mathbf{T}_i | i = 1, ..., n\}$ of the form $\mathbf{T}_i = (T_i^0, T_i^1, T_i^2)$ as above, we define $R_{\mathbf{T}_i} \subseteq STRUCT(\Sigma) \times STRUCT(\Sigma')$ as:

$$R_{\mathbf{T}_i}(w_1)(w_2) \Leftrightarrow R_{\mathbf{T}_i}(w_1) = w_2.$$

Note that since we have stipulated that $R_{\mathbf{T}_i}$ is a triple of bijective functions we can again define its inverse $R_{\mathbf{T}_i}^{-1}$ as:

$$R_{\mathbf{T}_i}^{-1}(w_1)(w_2) \Leftrightarrow R_{\mathbf{T}_i}(w_2)(w_1).$$

Our notion of Kripke model now becomes the following. We assume a universal domain $\mathcal{A}$ and signature $\Sigma = (FUNC, PRED)$, and consider worlds as first-order structures of type $w = (\mathcal{A}', \Sigma', \Phi', \Pi')$, where $\mathcal{A}' \subseteq \mathcal{A}$, Σ' is a subsignature of Σ, and Φ and Π are interpretation functions of the function and predicate symbols, respectively, of the signature Σ' in the domain $\mathcal{A}'$, as explained above. The class of all Σ'-structures for all Σ' that are subsignatures of Σ is denoted by $STRUCT$. The worlds in our first-order Kripke models will be just a subset of $STRUCT$. Note that this more liberal than in the propositional setting where we took as universe of worlds just *all* valuation functions. We think that in the first-order setting this liberality is more realistic, since we are very unlikely to consider all possible Σ'-structures. On the other hand, in the propositional setting we also considered *partial* valuations in order to cope with similarities between worlds with different relevant propositions.

In the first-order setting that we propose, we do this a little different. For convenience, we will work with similarity mappings between structures which are signature isomorphic and for which there is a bijection between the domains, but we will allow that the actual structures in the Kripke models have a richer structure (but irrelevant as to the similarity under consideration). Technically, we allow for this by looking at *relevant* subdomains and subsignatures. So, for example, we might consider a similarity mapping between two structures $w_1 = (\mathcal{A}_1, \Sigma_1, \Phi_1, \Pi_1)$ and $w_2 = (\mathcal{A}_2, \Sigma_2, \Phi_2, \Pi_2)$ with respect to the restriction to the domain/signature pairs $\sigma_0 = (\mathcal{A}_0, \Sigma_0)$ and $\sigma_0' = (\mathcal{A}_0', \Sigma_0')$ which means that actually only there is a similarity mapping from the substructure $w_1 \downarrow \mathcal{A}_0, \Sigma_0$ to the substructure $w_1 \downarrow \mathcal{A}_0', \Sigma_0'$ of w_2. We denote such a mapping as $\mathbf{T}(\sigma_0, \sigma_0')$.

A (first-order) Kripke model is an ordered tuple $\mathcal{M} = (STRUCT', \{R_{\mathbf{T}_i(\sigma_i, \sigma_i')} \mid i = 1, ..., n\})$, where $STRUCT' \subseteq STRUCT$ and $\sigma_i = (\mathcal{A}_i, \Sigma_i)$ with $\mathcal{A}_i \subseteq \mathcal{A}$ and Σ_i is a subsignature of Σ, and similarly for σ_i'.

Here the relations $R_{\mathbf{T}_i(\sigma_i, \sigma_i')}$ give the similarity mappings that are considered. Note that in view of the above, with each such mapping it is specified with respect to which restricting signature one should take the similarity.

On the basis of this Kripke model, we can interpret modal operators of type $\Box_{\mathbf{T}_i(\sigma_i, \sigma_i')}$ and derived modalities $\Box^{-1}_{\mathbf{T}_i(\sigma_i, \sigma_i')}$, $\Box_i$, $\Box$ and $\Box^*$, as in the propositional case: for instance,

$$\mathcal{M}, w \models \Box_{\mathbf{T}_i(\sigma_i, \sigma_i')} \varphi \Leftrightarrow \textit{for all } w' \textit{ with } R_{\mathbf{T}_i(\sigma_i, \sigma_i')}(w, w') : \mathcal{M}, w' \models \varphi.$$

$\Box^{-1}_{\mathbf{T}_i(\sigma_i, \sigma_i')}$, $\Box_i$, $\Box$ and $\Box^*$ are based on the (derived) relations $R^{-1}_{\mathbf{T}_i(\sigma_i, \sigma_i')}$, $R_i = R_{\mathbf{T}_i(\sigma_i, \sigma_i')} \cup R^{-1}_{\mathbf{T}_i(\sigma_i, \sigma_i')}$, $R = \bigcup_{i=1}^{n} R_i$, and R^*, the transitive, reflexive closure of the relation R, respectively. So, for instance,

$$\mathcal{M}, v \models \Box^* \varphi \Leftrightarrow \textit{for all } w \textit{ with } R^*(v, w) : \mathcal{M}, w \models \varphi,$$

and similarly for the other operators. Also we may use the duals $\Diamond_{\mathbf{T}_i(\sigma_i, \sigma_i')}$, $\Diamond^{-1}_{\mathbf{T}_i(\sigma_i, \sigma_i')}$, $\Diamond_i$, $\Diamond$ and $\Diamond^*$ of these operators, defined as usual.

As usual we define validity of a formula φ in a model $\mathcal{M}$, denoted $\mathcal{M} \models \varphi$, by $\mathcal{M}, w \models \varphi$ for all worlds w in $\mathcal{M}$, and validity of φ, denoted $\models \varphi$, by $\mathcal{M} \models \varphi$ for all models $\mathcal{M}$.

We can now put Proposition 3 in modal terms if we abuse our language slightly:

Theorem 2

$$\models \varphi \leftrightarrow \Box_{\mathbf{T}_i} \mathbf{T}_i(\varphi)[T^0(x_1)/x_1, ..., T^0(x_n)/x_n]$$

$\Diamond$

Note that this formula contains the substitution of a variable x by the formula $T^0(x)$, not to be confused with $T(x)$, which is just equal to x itself. The interpretation of this nonstandard formula is as follows: in a world w of a model $\mathcal{M}$, if the variable x has the value a, $T^0(x)$ denotes the element $T^0(a)$.

Apart from the validities we have seen in the propositional case, we now also get validities involving the first-order elements of the logic. A very infamous formula that comes up in the context of first-order modal logic is the so-called Barcan formula:

$$\forall x \Box_{\mathbf{T}_i(\sigma_i, \sigma_i')} \varphi \leftrightarrow \Box_{\mathbf{T}_i(\sigma_i, \sigma_i')} \forall x \varphi.$$

As explained in (Gamut 91) especially the implication of this formula for the dual operators, viz.

$$\exists x \Diamond_{\mathbf{T}_i(\sigma_i, \sigma_i')} \varphi \leftrightarrow \Diamond_{\mathbf{T}_i(\sigma_i, \sigma_i')} \exists x \varphi.$$

is in most contexts rather counter-intuitive. However, in the present context the Barcan formula is not valid in general, as one can see easily by the following counterexample:

Take $\varphi = Px$ and two similar worlds with disjoint non-empty domains. Then it is easy to see that the lefthandside of the Barcan formula is false (or not even defined, as the predicate P must be true for elements which are not in the domain), whereas the righthandside may well be true. However, (again) abusing language slightly we can express a modified version of the Barcan formula which is true:

$$\models \forall x \Box_{\mathbf{T}_i(\sigma_i, \sigma_i')} \varphi[T^0(x)/x] \leftrightarrow \Box_{\mathbf{T}_i(\sigma_i, \sigma_i')} \forall x \varphi$$

Note that again this formula contains the substitution of x by the formula $T^0(x)$, to be interpreted as earlier.

This formula as well as its dual form seem to be quite intuitive in this context.

3.3 Example: solar system - atom

Let us consider a Kripke model with two worlds w_1 and w_2. World w_1 represents the solar system:

$w_1 = (\mathcal{A}_1, \Sigma_1, \Phi_1, \Pi_1)$, where
$\mathcal{A}_1 = \{sun, planet_1, planet_2, ..., planet_n\}$;
$\Sigma_1 = (FUNC_1, PRED_1)$ with $FUNC_1 = \{Sun, Successor\_planet\}$, $PRED_1 = \{Large, Small, Middle, Yellow, Planet, Reflecting, Emitting, Attracts, Orbits\}$;
$\Phi(Sun) = sun$, $\Phi(Successor\_planet)(planet_i) = planet_{i+1}$ for $i = 1, ..., n-1$, $\Phi(Successor\_planet)(sun) = planet_1$, and $\Phi(Successor\_planet)(planet_n)$ is undefined;
$\Pi(Large) = \{Sun\}$, $\Pi(Small) = \{planet_1, ..., planet_n\}$, $\Pi(Middle) = \Pi(Yellow) = \Pi(Emitting) = \{sun\}$, $\Pi(Planet) = \Pi(Reflecting) = \{planet_1, ..., planet_n\}$, $\Pi(Attracts) = \{(sun, planet_1), ..., (sun, planet_n)\}$, $\Pi(Orbits) = \{(planet_1, sun), ..., (planet_n, sun)\}$,
In the world w_1 the following formulas are true: $\forall x\ Small(x) \rightarrow Planet(x)$, $\forall x\ Planet(x) \rightarrow Orbits(x, Sun)$, $\forall x \forall y\ (Successor\_planet(x) = y) \rightarrow Planet(y))$, $\neg Planet(Sun)$.

The world w_2 represents the atom.

$w_2 = (\mathcal{A}_2, \Sigma_2, \Phi_2, \Pi_2)$, where
$\mathcal{A}_2 = \{nucleus, electron_1, electron_2, ..., electron_n\}$;
$\Sigma_2 = (FUNC_2, PRED_2)$ with $FUNC_2 = \{Nucleus, Successor\_shell\}$,
$PRED_2 = \{Large, Small, Centre, Positive, Spin, Electron, Attracts\}$;
$\Phi(Nucleus) = nucleus$, $\Phi(Successor\_shell)(electron_i) = electron_{i+1}$ for $i = 1, ..., n-1$, $\Phi(Successor\_shell)(nucleus) = electron_1$, and $\Phi(Successor\_shell)(electron_n)$ is undefined;
$\Pi(Large) = \{nucleus\}$, $\Pi(Small) = \{electron_1, ..., electron_n\}$, $\Pi(Centre) = \Pi(Positive) = \{sun\}$, $\Pi(Spin) = \Pi(Electron) = \{electron_1, ..., electron_n\}$, $\Pi(Attracts) = \{(nucleus, electron_1), ..., (nucleus, electron_n)\}$.

In the world w_2 the following formulas are true: $\forall x\ Small(x) \rightarrow Electron(x)$, $\forall x \forall y\ (Successor\_shell(x) = y) \rightarrow Electron(y))$, $\neg Electron(Nucleus)$

Now let us consider $\Sigma_0 = (FUNC_0, PRED_0)$ with $FUNC_0 = FUNC_1$ and $PRED_0 = \{Large, Small, Centre, Planet, Emitting\}$, and $\Sigma_0' = (FUNC_0', PRED_0')$ with $FUNC_0' = FUNC_2$ and $PRED_0' = \{Large, Small, Centre, Attracts, Electron\}$. Furthermore, let $\sigma_0 = (\mathcal{A}, \Sigma_0)$ and $\sigma_0' = (\mathcal{A}, \Sigma_0')$. We can now define a similarity mapping $\mathbf{T}_1(\sigma_0, \sigma_0') = (T^0, T^1, T^2)$ by taking: T^0 a bijection between $\mathcal{A}_1$ and $\mathcal{A}_2$ such that $T^0(planet_i) = electron_i$, $T^0(sun) = nucleus$, T^1 a bijection between $FUNC_1$ and $FUNC_2$ such that $T^1(Successor\_planet) = Successor\_shell$, $T^1(Sun) = Nucleus$, and T^2 a bijection between $PRED_1 \cap$

$PRED_0$ and $PRED_2 \cap PRED'_0$ such that $T^2(Large) = Large$, $T^2(Small) = Small$, $T^2(Middle) = Centre$, $T^2(Planet) = Electron$, $T^2(Attracts) = Attracts$.

This similarity mapping $\mathbf{T}_1(\sigma_0, \sigma'_0)$ induces an accessibility relation $R_{\mathbf{T}_1(\sigma_0,\sigma'_0)}$, and a corresponding modality $\Box_{\mathbf{T}_1(\sigma_0,\sigma'_0)}$. If we take, for simplicity, our Kripke model only consisting of the worlds w_1 and w_2, we can now show that $(w_1, w_2) \in R_{\mathbf{T}_1(\sigma_0,\sigma'_0)}$ (This is left to the reader to verify).

We can now check Proposition 3 for the formulas that hold in w_1. We know for example (abbreviating $\mathbf{T}_1(\sigma_0, \sigma'_0)$ by $\mathbf{T}$): $w_1 \models \forall x\ (Small(x) \rightarrow Planet(x))$, so by Proposition 3 we obtain that $w_2 \models \mathbf{T}_1(\forall x\ Small(x) \rightarrow Planet(x)) = \forall x\ (\mathbf{T}_1(Small)(\mathbf{T}_1(x)) \rightarrow \mathbf{T}_1(Electron)(\mathbf{T}_1(x))) = \forall x\ (Small(x) \rightarrow Electron(x))$, which is indeed true, as we have seen before.

Note that the worlds w_1 and w_2 share some predicate symbols such as 'Large' and 'Small'. This enables us to express something more abstract but nontrivial (of course, always valid 1st-order formulas hold) with respect to the similarity class induced by the similarity mapping $\mathbf{T}$, i.e. something that holds for both w_1 and w_2. We can do this by means of the modal operator $\Box_1^*$. For instance, we have that:

$$w_1 \models \Box_1^* \exists x \exists y (Large(x) \wedge Small(y) \wedge Attracts(x, y))$$

thus abstracting from which large thing exactly is attracting which small thing.

4 Defeasible Analogical Reasoning

In the previous sections we have treated reasoning by analogy in such a way that the conclusions of such reasoning are certain and inescapable. This is obviously an over-idealization and not true in general. If one knows exactly to what degree two similar structures are similar, then drawing certain conclusions seems to be a correct procedure. However, in practice this is almost never the case. We have uncertainties about the target domain which is the very reason why we try to compare it with a familiar source domain and draw some conclusions from it which are meant to be tentative and might be defeated by the discovery of new information about the target domain. In this case it seems even to be very undesirable to be able only 'hard' conclusions, since this implies that either we cannot infer interesting but uncertain things about the target domain (and this is what we are aiming for), or when we derive something which is contradictory with later observations we are faced with a hard inconsistency.

For instance, consider the solar system - Atom example. Here there is a similarity between the Sun and the nucleus of the atom, and we might transfer what we know about the Sun to the nucleus to a certain degree. This we might want to specify completely in the sense that we specify exactly which predicates are considered in the similarity mapping. For instance, the predicate 'is the

middle of the (sun) system' can be mapped onto the predicate 'is the centre of the (atom) system'. On the other hand, the predicate 'is yellow' for the sun cannot be mapped so easily to something similar in the domain of the atom. In general, however, we do not know exactly which predicates 'carry over' and which do not.

For this reason we introduce the notion of *defeasibility* for analogical reasoning. The upshot of this is that conclusions concerning certain predicates are 'likely', but not quite certain, and can be defeated by new information about the target domain (such as direct observations, although we will not specify the source of information here). Some argue that this likelihood has to do with (high) probabilities. This might well be the case in particular contexts but as in studies of default reasoning we are more inclined to look at this from a more qualitative perspective and view 'rendition rules' in analogical reasoning as *default rules*, i.e. rules that under normal circumstances can be applied, but allow exceptions if one has the disposal of additional information.

So, in a sense, what we will do, is combining our modal framework of analogical reasoning with defeasible reasoning, and, in particular, some form of default reasoning. The way we will do this is inspired by earlier work on default and counterfactual reasoning we have done (Meyer & Van der Hoek 93, 95).

Before engaging into the formal details we will give the general idea. Suppose we have two domains that we think are similar in certain respects. And suppose we have information φ about the source domain that is likely to translate to the target domain (so, technically we are then speaking about a predicate that is in the domain of our similarity mapping, but which must be considered defeasible). Then we can use our framework to derive its translation $T(\varphi)$ for the target domain. Now, we first check whether this assertion $T(\varphi)$ is *consistent* with what we know already about the target domain. If this is not the case, we forget about the conclusion, If it is, we draw the tentative conclusion that $T(\varphi)$ holds in the target domain. In the formalism we will use, it will be indicated or marked as tentative by means of a modal operator, which basically states that the formula is true in a selected or *preferred* subset of the worlds that describe the (knowledge about the) target domain. Naturally, this allows for the possibility that when new information about the target domain is obtained, this will lead us to a (set of) world(s) that are not within this preferred subset of worlds, which means that the tentatively believed conclusion about the target domain will not be true after all.

Formally we go about as follows:

We extend our modal language with some new operators. Firstly, we introduce operators $\Box'_{\mathbf{T}_i}$. These operators will be used to point to a (preferred) subset of the set of worlds that pertain to the target domain with respect to the similarity mapping $\mathbf{T}_i$. For further convenience, we also introduce modal operators $[S]$, where S is a subset of the set VAL of all worlds. This will facilitate speaking about the source and target domains, since these operators give us so-to-speak a direct pointer to the worlds pertaining to these, whereas the

modalities associated with the similarity mappings do this in a rather indirect manner. Finally, we also consider the duals $\Diamond'_{\mathbf{T}_i}$ and $< S > \varphi$ defined as usual.

We enrich our Kripke structures with some extra elements:

An (enriched) Kripke model is an ordered tuple $\mathcal{M} = (VAL, \{R_{\mathbf{T}_i} | i = 1, ..., n\}, \{R'_{\mathbf{T}_i} | i = 1, ..., n\})$, where, for all i and j, $R'_{\mathbf{T}_i} \subseteq R_{\mathbf{T}_i}$.

The additional operators are interpreted on these enriched Kripke models as follows:

$$\mathcal{M}, v \models \Box'_{\mathbf{T}_i}\varphi \Leftrightarrow \textit{for all } w \textit{ with } R'_{\mathbf{T}_i}(v, w) : \mathcal{M}, w \models \varphi$$

and

$$\mathcal{M}, v \models [S]\varphi \Leftrightarrow \textit{for all } w \in S : \mathcal{M}, w \models \varphi$$

One now immediately obtains the following validities:

1. $\models \Box_{\mathbf{T}_i}\varphi \rightarrow \Box'_{\mathbf{T}_i}\varphi$
2. $\models [S_1]\varphi \rightarrow [S_2]\varphi$ for $S_1 \supseteq S_2$
3. $\models < S_1 > \top \rightarrow ([S_1][S_2]\varphi \leftrightarrow [S_2]\varphi)$
4. $\models < S_1 > \top \rightarrow ([S_1] < S_2 > \varphi \leftrightarrow < S_2 > \varphi)$

Using the expressibility of our modal language we can now give defeasible versions of analogical derivation rules:

Instead of the (strict) rule

$$\varphi \rightarrow \Box_{\mathbf{T}_i}(\mathbf{T}_i(\varphi))$$

which was a validity in the previous sections, where we based the modality $\Box_{\mathbf{T}_i}$ on a strict rendition function $\mathbf{T}_i$, we now relax this and use the defeasible rule:

$$\varphi \wedge \Diamond_{\mathbf{T}_i}(\mathbf{T}_i(\varphi)) \rightarrow \Box'_{\mathbf{T}_i}(\mathbf{T}_i(\varphi)) \qquad (1)$$

which expresses formally what we described informally above: if the $\mathbf{T}_i$-translation of the source domain information φ is compatible with the information about the target domain determined by the similarity mapping $\mathbf{T}_i$, then we prefer a set of worlds within the target domain satisfying this translated information $\mathbf{T}_i(\varphi)$.

In the example of the solar system - atom we might, for instance, when we are not completely sure about this, try to render the predicate (light) emitting in the solar system to a predicate (radiation) emitting (added to the signature) in the atom system to be used as a kind of hypothesis. For this rendition one now can use the above *defeasible* rule

$$\mathit{Emitting}(\mathit{Sun}) \wedge \Diamond_{\mathbf{T}_i}(\mathbf{T}_i(\mathit{Emitting}(\mathit{Sun}))) \rightarrow \Box'_{\mathbf{T}_i}(\mathbf{T}_i(\mathit{Emitting}(\mathit{Sun})))$$

yielding the rule:

$$Emitting(Sun) \wedge \Diamond_{\mathbf{T}_i}(Emitting(Nucleus)) \rightarrow \Box'_{\mathbf{T}_i}(Emitting(Nucleus))$$

This means that *unless there is information in the target domain that it is not the case that* $Emitting(Nucleus)$ *is true*, it is assumed that it is so.

Note that the target domain is generally dependent on the similarity mapping $\mathbf{T}_i$ and the world where you are (in or rather pertaining to the source domain), since in general we allow the same similarity mapping $\mathbf{T}_i$ to point to different target domains from different source domains.

This is the reason why we do *not* have 'S5-like' validities such as

- $\Diamond_{\mathbf{T}_i}(\top) \rightarrow (\Box_{\mathbf{T}_i}\Box'_{\mathbf{T}_i}(\varphi) \leftrightarrow \Box'_{\mathbf{T}_i}(\varphi))$
- $\Diamond'_{\mathbf{T}_i}(\top) \rightarrow (\Box'_{\mathbf{T}_i}\Box_{\mathbf{T}_i}(\varphi) \leftrightarrow \Box_{\mathbf{T}_i}(\varphi))$

To ease working with these defeasible rules and reasoning about a particular source and target domain, we may use the modalities $[S]\varphi$, which are 'S5-like'. The way to use these in practice is the following: identify your (fixed) source domain and the set S of worlds associated with this. Determine the target domain(s) S_i with respect to the similarity mapping(s) $\mathbf{T}_i$ under consideration. If this can be done, formula (??) can be written in a more 'rigid' or direct way:

$$[S]\varphi \wedge < S_i > (\mathbf{T}_i(\varphi)) \rightarrow [S'_i](\mathbf{T}_i(\varphi)) \text{ for some } S'_i \subseteq S_i \qquad (2)$$

When reasoning with this representation we have the disposal of the validities 1.-4. mentioned above.

5 Conclusion

In this paper we have given a semantical approach to analogical reasoning based on a possible worlds framework. Starting out from a very simple and rather naive notion of similarity and associated models and modal operators expressing such similarity in a purely propositional setting we have extended our approach to a first-order language and logic, and ended with the incorporation of the notion of defeasibility in analogical inferences. Of course, the semantics proposed does not yet capture the full complexity of all forms of reasoning by analogy. To begin with, only a qualitative form of analogical reasoning is treated, where it is abstracted away from all possible kinds of similarity measures. This issue should bear a relation with the defeasible nature of this form of reasoning, which can also be viewed both qualitatively and quantitatively. This will be subject of further study.

Acknowledgements. The authors wish to thank Wiebe van der Hoek, Dirk van Dalen, Peter van Emde Boas, Johan van Benthem, Rineke Verbrugge and

Dick de Jongh for discussions on the topic of this work. Also the partial support of ESPRIT III BRA No 6156 DRUMS (Defeasible Reasoning and Uncertainty Management Systems) is gratefully acknowledged.

References

[1] (Burstein 86)
Burstein M.H. Incremental Analogical Reasoning. *Machine Learning, an Artificial Approach* Vol. II. Michalski R. S., Carbonel J.G., Mitchel T.M. (eds.), Morgan Kaufman, 1986.

[2] (Burstein 88)
Burstein M.H. Combining Analogies in Mental Models. *Analogical Reasoning.* Helman D.H. (ed.), Kluwer Academic Publishers, Boston, 1988.

[3] (Chellas 80)
Chellas B.F. *Modal Logic: An Introduction.* Cambridge University Press, Cambridge/London 1980.

[4] (van Dalen 89)
van Dalen D. *Logic and Structure.* Springer-Verlag, second edition 1989.

[5] (Evans 68)
Evans T.G. A Program for the Solution of Geometric Analogy Intelligence Test Questions. *Semantic Information Processing.* Minsky M.L. (ed.), MIT Press, Cambridge 1968.

[6] (Gamut 91)
Gamut L.T.F. *Logic, Language and Meaning Vol. II, Intensional Logic and Logical Grammar.* University of Chicago Press, Chicago, 1991.

[7] (Hall 89)
Hall R.P. Computational Approaches to Analogical Reasoning. A Comparative Analysis. *AI Journal* 39(1) 1989, pp. 39-120.

[8] (Indurkhya 89)
Indurkhya B. Modes of Analogy. *International Workshop on AII.* Jantke K.P. (ed.), Lecture Notes in Artificial Intelligence, Springer Verlag 1989, pp. 217-229.

[9] (van Leeuwen 95)
J.C. van Leeuwen. Analogical Reasoning. A Semantical Approach, Master's Thesis, Utrecht University, 1995.

[10] (Meyer & Van der Hoek 93)
Meyer J.-J. Ch. & van der Hoek W. Counterfactual Reasoning by (Means

of) Defaults. *Annals of Mathematics and Artificial Intelligence* 9 (III-IV), 1993, pp. 345-360.

[11] (Meyer & Van der Hoek 95)
Meyer J.-J. Ch. & van der Hoek W. A Default Logic Based on Epistemic States. *Fundamenta Informaticae* 23(1), 1995, pp. 33-65.

[12] (Morgan 79)
Morgan C.G. Modality, Analogy and Ideal Experiments according to C.S. Peirce. *Synthese* 41, 1979 (1), pp. 65-83.

[13] (Peirce)
Peirce C.S. *Collected Papers of Charles Saunders Peirce Vols. I - VIII.* ed. by Ch. Hartsborne, P. Wriss & A.W. Burks, Harvard University Press, Cambridge (1933- 1958).

[14] (Tarski, Mostowski & Robinson 53)
Tarski A., Mostowski A. & Robinson R. *Undecidable Theories.* North-Holland 1953.

[15] (Thiele 86)
Thiele H. A Model Theoretic Approach to Analogy. *International Workshop on AII.* Jantke K.P. (ed.), Lecture Notes in Computer Science, Springer Verlag 1986, pp 196-208.

[16] (Winston 68)
Winston P. Learning and Reasoning by Analogy. *Communications of the ACM* 23(12), 1968, pp 683-703.