A Semantics for Default Logic

David W. Etherington[1]

Artificial Intelligence Principles Research Department
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974-2070
ether%allegra(abtl.csnet

Abstract

In spite of the importance of well-understood semantics for knowledge representation systems, proponents of default logic have tended to ignore the lack of a general model-theoretic semantics for the formalism. This shortcoming is addressed by the presentation of such a model-theory. This characterization differs in some ways from traditional semantics. These differences are explained and motivated, and some applications of the semantics are discussed.

1. Introduction

In his development of default logic, Reiter [1980] provides a fixed-point characterization of the extensions of a default theory, but no model-theoretic semantics. Default logic has attracted attention as a formal system for non-monotonic reasoning. This is due, in part, to the intuitive clarity with which default information can be expressed. This clarity has ameliorated the delayed development of a semantics for the formalism; people have tended to be satisfied with intuitive characterizations of the extensions of default theories, together with the Tarskian semantics of the individual extensions.

While this transition to model-theory may be relatively painless, there is still a need for a semantic characterization of default logic as a whole. Etherington [1982, 1983] loosely characterized the semantics of default logic, observing that defaults "can be viewed as extending the first-order knowledge about an incompletely-specified world. [They] select restricted subsets of the models of the underlying first-order theory". Lukaszewicz [1985] recently formalized this idea for a restricted class of default theories. In this paper, we generalize this work to cover the entire class of default theories. We also recast the semantics in a more intuitive and familiar form, making it more useful and accessible. Proofs of the results quoted here may be found in [Etherington 1986].

In what follows, we present a semantics for default logic. This is subsequently evaluated and shown to provide useful insight into the formalism. We begin by providing the briefest of introductions to default logic. More details can be found in [Reiter 1980].

[1] Parts of this work were done at the University of British Columbia, and supported in part by an I.W. Killam Predoctoral Scholarship and by NSERC grant A7642.

2. Default Logic

Default logic allows new inference rules to be added to a standard first-order logic. These rules sanction their conclusions provided the set of beliefs satisfies the conditions outlined in their premises. Unlike standard logic, these premises may refer both to what is known and to what is not known. The latter property allows rules to be added that specify inferences to be made only when specific information is missing.

A *default* is any expression of the form:[2]

$$\frac{\alpha(\vec{x}) : \beta_1(\vec{x}), \ldots \beta_n(\vec{x})}{\omega(\vec{x})}$$

where $\alpha(\vec{x})$, the $\beta_i(\vec{x})$, and $\omega(\vec{x})$ are all formulae whose free variables are among those in $\vec{x} = x_1,\ldots,x_m$. α, the β_i, and ω are called the *prerequisite*, *justifications*, and *consequent* of the default, respectively. Two classes of single-justification defaults are distinguished. Those with $\beta(\vec{x}) = \omega(\vec{x})$ are said to be *normal*; those with $\beta(\vec{x}) = \omega(\vec{x}) \wedge \gamma(\vec{x})$ are called *seminormal*.

Defaults serve as rules of inference or conjecture, augmenting those provided by first-order logic. Under certain conditions, they sanction inferences that could not be made within a conventional framework. If their prerequisites are known and their justifications are "consistent" (*i.e.*, their negations are not provable), then their consequents can be inferred. A *default theory* is an ordered pair consisting of a set of defaults, D, and a set of first-order formulae, W. D can be viewed as extending the definite knowledge of W to provide information not derivable from W.

Since defaults allow reference to what is not provable in the determination of what is provable, the "theorems" of a default theory are not so easy to generate as are those of a first-order theory. What is provable both determines and is determined by what is not provable. To avoid this apparent circularity, the theorems of a default theory are defined by a fixed-point construction. An *extension*, E, for a default theory is required to contain all the known facts, be closed under the $\vdash$ relation, and contain the consequent of any default whose prerequisite is satisfied by E and whose justifications are consistent with E. Furthermore, E contains only those facts required by the above conditions.

[2] We omit the "M" preceding each 0_n since they are redundant in the positional notation.

3. A Semantics for Default Theories

As mentioned earlier, default logic's semantics can be viewed in terms of restrictions of the set of models of the underlying theory. The first-order theory partially specifies a world, which is further specified by the defaults. Each default can be viewed as extending the world-description by restricting the set of possible worlds assumed to contain the "rear" world, at the same time constraining how other defaults may further extend the world-description. Lukaszewicz [1985J formalizes this intuition for normal default theories. Because of the well-behaved nature of these theories, this is straightforward. The resulting semantics amounts to considering the Tarskian semantics of each of the partial extensions constructed by proceeding monotonically from *W* toward an extension by repeatedly satisfying the next (according to some arbitrary ordering of the defaults) applicable normal default by making its consequent true. If ever no more defaults from *D* are applicable, the resulting set of models characterizes an extension. Since each step affirms a formula consistent with those affirmed previously, the set of models contracts monotonically. The intersection of the sets of models from each stage (or the last set, if there are only finitely many) is precisely the set of models of the extension.

Lukaszewicz' approach applies only to normal defaults, since the required property of semi-monotonicity, which allows each normal default to be treated independently, does not hold for non-normal defaults [Reiter 1980, Theorem 3.2]. Lukaszewicz proposes addressing this problem by translating non-normal defaults to normal defaults. He argues that, of single-justification defaults, only normal and seminormal defaults have reasonable interpretations. Other defaults are therefore translated to seminormal defaults by conjoining the consequent to the justification:

$$\frac{\alpha : \beta}{\gamma} \rightarrow \frac{\alpha : \beta \wedge \gamma}{\gamma}.$$

The translation from seminormal to normal, which is somewhat more controversial, strengthens the consequent to make it identical to the justification:

$$\frac{\alpha : \beta \wedge \gamma}{\gamma} \rightarrow \frac{\alpha : \beta \wedge \gamma}{\beta \wedge \gamma}.$$

This makes sense, Lukaszewicz argues, so long as α's that are also γ's are typically β's. That is, so long as one could reasonably augment the theory with: $\frac{\alpha \wedge \gamma : \beta}{\beta}$.

One can imagine situations where this is not appropriate. For example, a system for legal reasoning might have a rule suggesting that those with motives who *might* be guilty should be suspects:

$$\frac{has\text{-}motive(x) : guilty(x)}{suspect(x)}$$

It is clearly reasonable to translate this to:

$$\frac{has\text{-}motive(x) : suspect(x) \wedge guilty(x)}{suspect(x)},$$

allowing that there may be reasons not to suspect someone even without knowing their innocence. It is *not* reasonable to follow through by assuming the guilt of all suspects:

$$\frac{has\text{-}motive(x) : suspect(x) \wedge guilty(x)}{suspect(x) \wedge guilty(x)}$$

Furthermore — for whatever purpose they may serve — the scheme does not apply to defaults with multiple justifications. Thus, while Lukaszewicz' semantics covers many cases, there are reasons to want a semantics that covers more than normal defaults. To this we now turn.

Because of the failure of semi-monotonicity for non-normal theories, simply applying one default after another does not, in general, lead to extensions. It is necessary to ensure that the application of each default does not violate the justifications of already-applied defaults. In the semantic domain, this means that Lukaszewicz' simple approach of sequentially restricting the set of models of the first-order theory, to satisfy each default in turn, will not necessarily lead to the set of models of an extension. Etherington [1986] develops extra machinery that, essentially, introduces a backtracking (or pruning) mechanism to the semantics to deal with this problem. While this extension is not particularly difficult conceptually, the technical details are somewhat formidable and the result appears (superficially) cially ad *hoc*. This is partly because the resulting semantics looks radically different from familiar Tarskian or Kripkean (or even circumscriptive) model-theories. We present an equivalent semantics below, in a very different guise. This new presentation makes the salient features of the semantics more apparent, as well as easing comparisons with the semantic theories of other nonmonotonic systems.

Definition: $\Gamma_1 \geq_\delta \Gamma_2$ **(δ prefers Γ_1 to Γ_2)**
Consider a default, $\delta = \frac{\alpha : \beta_1, \ldots, \beta_n}{\omega}$, a set of models, Γ, and $\Gamma_1, \Gamma_2 \in 2^\Gamma$. The partial-order corresponding to δ, $\geq_\delta$, over 2^Γ is defined as follows:

$$\Gamma_1 \geq_\delta \Gamma_2 \equiv$$
$$\forall \gamma \in \Gamma_2 . \gamma \models \alpha, \exists \gamma_1, \ldots, \gamma_n \in \Gamma_2 . \gamma_i \models \beta_i, \text{ and}$$
$$\Gamma_1 = \Gamma_2 - \{\gamma \mid \gamma \models \neg\omega\}$$ ∎

Intuitively, $\geq_\delta$ captures δ's preference for more specialized world-descriptions, in which its consequent holds, over those in which its prerequisite holds and its justifications are each consistent, but which do not necessarily satisfy its consequent. The idea of preference extends to sets of defaults in the obvious way.

Definition: $\Gamma_1 \geq_D \Gamma_2$ **(D prefers Γ_1 to Γ_2)**
Consider a set of defaults, D, a set of models, Γ, and $\Gamma_1, \Gamma_2 \in 2^\Gamma$. The partial-order corresponding to D, $\geq_D$, over 2^Γ is defined as the union of the partial-orders given by the defaults of D. *I.e.*,

$$\Gamma_1 \geq_D \Gamma_2 \equiv \exists \delta \in D . \Gamma_1 \geq_\delta \Gamma_2.$$ ∎

For a normal default theory, $\Delta = (D, W)$, it is sufficient at this point to consider the $\geq_D$-maximal elements of $2^{MOD(W)}$, where $MOD(W)$ is the set of all models of W. Each of these corresponds to the set of all models of an extension of Δ, and *vice versa*.[3] As mentioned earlier, the failure of semimonotonicity for non-normal defaults means a more complex approach is necessary for full generality. We must semantically account for non-normal defaults' ability to require the continued consistency of their justifications without their explicitly acting to ensure this. We have

[3] Lukaszewicz proves this for an equivalent (though superficially quite different) formulation.

identified a property of sets of models, which we call *stability,* that provides such an account.

Definition: Stability

Let $\Delta = (D, W)$ be a default theory, and let Γ be a $\geq_D$-maximal element of $2^{MOD(W)}$. Γ is *stable* for Δ if and only if there is a $D' \subseteq D$ such that $\Gamma \geq_D MOD(W)$, and for each $\frac{\alpha : \beta_1, \ldots, \beta_n}{\omega} \in D'$, $\exists \gamma_1, \ldots, \gamma_n \in \Gamma$. $\gamma_i \models \beta_i$. ∎

In other words, a set of models is stable for a default theory, (D.VV). if it is a maximal specialization of the set of models of *W.* and does not refute the justifications of any of the defaults used in the specialization. This notion of stability is related, but not directly analogous, to stability in autoepistemic theories [Moore 1985]. In particular, it incorporates elements of what Moore calls "groundedness". For a detailed discussion of the relationship between default logic and autoepistemic logic, see [Konolige 1987].

The stable sets of models for a default theory provide a semantic interpretation for the theory. The soundness and completeness results for this semantics are given by Theorems 1 and 2. respectively.

Theorem 1 — Soundness

If E is an extension for Δ. then $\Gamma = \{M \mid M \models E\}$ is stable for Δ. ∎

Theorem 2 — Completeness

If Γ is stable for Δ, then Γ is the set of models for some extension of Δ. (*I.e.*, $\{ \omega \mid \forall \gamma \in \Gamma. \gamma \models \omega \}$ is an extension for Δ.) ∎

Example 1

Consider the default theory:

$$\Delta = \left(\left\{ \frac{A : B \wedge \neg C}{B}, \frac{A : C \wedge \neg B}{C} \right\}, \{A\} \right).$$

This theory has the following maximal world-descriptions: $\{ M \mid M \models A.B \}$, $\{ M \mid M \models A.C \}$. both of which are stable, so the theory has two extensions, $Th(\{A, B\})$ and $Th(\{A, C\})$. ∎

Example 2

The incoherent theory:

$$\Delta = \left(\left\{ \frac{: \neg A}{A} \right\}, \{ \} \right)$$

has only one maximal world-description: $\{ M \mid M \models A \}$, which is not stable. Hence this theory has no extension. ∎

4. Discussion

We are now faced with the question of how well motivated this semantics is. It is clearly different from traditional semantic theories. For example, in Tarskian semantics the truth values of the atomic formulae determine those of every other formula. Similarly, in Kripke semantics, a set of worlds, an accessibility relation, and an assignment of truth values to the atomic formulae determine what is true in the structure. It is common to have some notion of a structure, and of what it means for any primitive construct of the language to be satisfied by a structure. Logical operators and connectives are also given interpretations, allowing the truth of any construct of the language to be determined *vis-a-vis* any structure. Finally, validity and satisfiability are defined in terms of these other notions.

Because of the indexical nature of defaults, it does not seem possible to provide them with such static interpretations. The proof-theory of default logic places upper-, as well as lower-, bounds on the states of knowledge that can be taken as satisfying the theory. These bounds, however, are determined by *the nay knowledge is extended.* This requires information that is not inherent in typical semantic structures.

For example, imagine the defaults:

$$\left\{ \frac{\alpha : \beta \wedge \gamma}{\beta}, \frac{\beta : \neg\gamma}{\neg\gamma} \right\}.$$

Assuming a semantics that interprets defaults in terms of partial world-descriptions, the world-descriptions themselves clearly provide insufficient information. Consider a world-description, Γ, such that $\Gamma \models \alpha$. β and $\Gamma \models \neg(\beta \wedge \gamma)$. Although it seems that the second default should be applicable, the only justification for a model-theory such as this comes from viewing defaults as operations that transform theories. On this view, the set of models $[M \mid \Gamma \models \gamma]$ is the theory which corresponds to the partial description Γ (in the usual Tarskian sense). The default is understood as associated with the defaults encode the transformations that the defaults induce on these theories. This view helps explain both why individual world-descriptions alone are insufficient to determine the satisfaction of defaults and why the semantics deals with satisfaction of default theories as a whole, rather than of individual defaults in isolation.

There are related semantic characterizations in the literature. For example, there are the minimal-model semantics of circumscription [McCarthy 1980] and of various other forms of closed-world reasoning [Etherington 1986]. These characterize the semantics of certain non-monotonic formalisms in terms of the minimal elements of an ordered set of structures. The presentation above goes beyond this in the complexity of the structures concerned and in the "post-filtering" of the set of minimal models, but is closely related in spirit. This topic is taken up in detail in [Etherington 1986, 1987], but some of the flavour can be given in the space available here.

The semantics of a theory under the various forms of "closed-world" assumption can be defined in terms of a restriction of the set of models of the underlying theory. according to some principle of minimization (typically according to the subset inclusion partial-order over the extensions of some predicates). The above model-theory for default logic similarly provides a principle for determining which models of a first-order theory characterize acceptable belief-sets, in this case based on maximal satisfaction of the set of defaults. There are several significant differences, however. First, rather than an ordering on individual

models, an ordering is imposed on sets of models. Second, the ordering is defined in terms of accessibility via a default, rather than strictly in terms of general criteria and intrinsic features of the models themselves. Finally, each extension is determined by a single extremum of the ordering, rather than by the set of all extrema.

The first of these differences, as mentioned previously, is necessary to encode the constraints of the theory. It allows a distinction between what is believed and what is merely consistent with what is believed. It also allows a natural representation of the incompleteness of default theories, which — unlike the models of a first-order theory — do not decide every formula.

The second deviation arises because defaults are general, non-homogeneous, inference rules. Consequently, the partial-order relation is potentially more complex for default logic. Lifschitz" [1984. 1986] recent work allowing more general orderings may void this difference, but the question remains open.

Individual extrema determine extensions as a result of the "brave" (in McDermott's [1982] terminology) character of default logic Default logic treats each extension as an acceptable set of beliefs, with the intention that a reasoner will somehow "choose" a single extension within which to reason about the world. Other nonmonotonic formalisms (such as circumscription [McCarthy 1980. 1986]) are based on "cautious** approaches that accept a default conclusion only if it occurs in *all* acceptable sets of beliefs. One can easily construct a variant of default logic that pursues a "cautious" course. (The converse is not obviously true for all "cautious" systems: see [Etherington 1987].) Such a system would define the theorems of a default theory to be those formulae true in all extensions, with the obvious change to the semantics: the theorems would then be defined as those formulae true in all models of all stable world-descriptions.

5. Conclusions

Default logic has occasionally been criticized for its lack of a general model-theoretic semantics. To answer this criticism we have presented such a semantics. We have also tried to show the appropriateness and utility of this semantics. We showed that it has merits that justify its use, and suggested it is a useful tool for comparing default logic with other nonmonotonic formalisms.

Our semantics differs from traditional, Tarskian, model-theoretic semantics in two respects: it is global, and it is not structure-oriented. By "global" we mean that there is no notion of satisfaction of individual defaults independently of the theory in which they occur (and hence no notion of satisfaction of a theory as incremental satisfaction of its components). The "non-structure-oriented" nature corresponds to the observation that defaults serve less as components of theories than as operators taking theories into other, more complete, theories. We argue that these facets of the semantics are not only justifiable, but serve to highlight important features of the operation of the syntactic mechanisms of default logic.

References

Etherington, D.W. [1982], *Finite Default Theories,* M.Sc. thesis. Department of Computer Science. University of British Columbia.

Etherington, D.W. [1983]. *Formalizing Non-Monotonic Reasoning Systems,* Technical Report 83-1. Department of Computer Science. University of British Columbia. (Also in *Artificial Intelligence 31.* 1987. 41-85).

Etherington, D.W. [1986]. *Reasoning from Incomplete Information,* Pitman Research Notes in Artificial Intelligence, Pitman Publishing Limited, London. 1987.

Etherington. D.W. [1987]. "Relating default logic and circumscription*. *Proc. IJCAI-JO.* Milan. Italy.

Konolige. K. [1987]. "On the relation between default theories and autoepistemic logic". *Proc. IJCAI-JO,* Milan. Italy.

Lifschitz, V. [1984]. *Some Results on Circumscription.* Technical Report STAN-CS-84-1019, Stanford Universitv. Stanford. CA.

Lifschitz. V [1986], "Pointwise circumscription". *Proc. AAAI-86.* Philadelphia. PA. 406-410.

Lukaszewicz. W. [1985]. "Two results on default logic*. *Proc. IJCAI-9.* Los Angeles. CA. 459-461.

McCarthy, J. [1980]. "Circumscription — a form of non-monotonic reasoning". *Artificial Intelligence 13.* 27-39.

McCarthy. J. [1986]. "Applications of circumscription to formalizing commonsense knowledge". *Artificial Intelligence 28.* 89-116.

Moore, R.C. [1985]. "Semantical considerations on nonmonotonic logic"'. *Artificial Intelligence 25,* 75-94.

McDermott. D [1982]. "Non-monotonic logic II". *J ACM* 29.35-57.

Reiter. R. [1980]. "A logic for default reasoning". *Artificial Intelligence 13.* 81-132.

Touretzky. D.S. [1984]. *The Mathematics of Inheritance Systems.* Ph.D. dissertation. Department of Computer Science, Carnegie-Mellon University.