Defeasible Reasoning as a Cognitive Model

G. Aldo Antonelli[*]
Stanford University
`aldo@csli.stanford.edu`

November 1996

One of the most important developments over the last twenty years both in logic and in Artificial Intelligence is the emergence of so-called *non-monotonic logics*. These logics were initially developed by McCarthy [10], McDermott & Doyle [13], and Reiter [17]. Part of the original motivation was to provide a formal framework within which to model cognitive phenomena such as *defeasible inference* and *defeasible knowledge representation*, i.e., to provide a formal account of the fact that reasoners can reach conclusions tentatively, reserving the right to retract them in the light of further information.

This initial intuition has given rise to a wealth of formal frameworks, from circumscription (McCarthy [11]) to default logic (Reiter [17]) to autoepistemic logic (Moore [14]), to cite only a few. All these frameworks, however, single out a notion of inferential consequence which is either computationally hard, or mathematically not well-behaved (essentially because of what has come to be known as the "multiple-extensions" problem), or both. In this paper, we will claim that such computational complexity and mathematical behavior turn out to be problematic if we want to adhere to the original motivation for the development of non-monotonic logics, and that they might ultimately make non-monotonic formalisms not viable as *cognitive models*.

On the other hand, it is possible to argue, as we do here, for a different approach to defeasible inference that builds upon, and shares certain basic intuitions with, the wealth of technical results in non-monotonic reasoning. At the same time this alternative approach exhibits a different mathematical behavior which it is hoped will lead to a more tractable inferential model. In this paper, we are going to show how to recast the basic notions of at least some defeasible inferential and representational frameworks in such a way as to give rise to formalisms whose mathematical properties might make them somewhat better suited as cognitive models.

Throughout we will keep in the background the following two desiderata, which any formal account of defeasible inference should meet:

d1 It should be possible to define in a natural way a notion of *defeasible consequence*, in analogy to the classical notion of logical consequence;

d2 Such a notion of defeasible consequence should be *feasible*, i.e., as easy to compute as could be expected, relatively to the language (propositional, first-order, modal, etc.) being used.

The two desiderata are connected, in that a *natural* notion of defeasible consequence is likely also to meet the second desideratum. It also seems appropriate to think that any notion that fails to meet **d1** or **d2** might ultimately turn out not to be a viable model of human cognition. Many

---

[*] Thanks to Dagfinn Føllesdal for useful comments and criticisms.

defeasible representational frameworks fail to meet **d1**, or **d2**, or both, and this seems to cast some doubt on their viability as cognitive models. On the other hand, the notion of defeasible inference introduced in [2] and in the present paper allows for a natural definition of a defeasible inference relation, which, it is hoped, will also turn out to meet **d2**.

It should be mentioned that this paper aims to be motivational, and that a complete treatment of the formal details can be found in [1, 2]. In the following pages, we will present a particular inferential formalism, i.e., a simplified version of Reiter's default logic. We will provide the fundamental definitions and illustrate the basic facts concerning this version of default logic, in order to assess its viability both as a formal tool and a cognitive model. We will then proceed to impose a "new twist" on the basic notions in such a way as to preserve the fundamental intuitions while at the same time obviating some of the drawbacks of the standard version. Finally we will mention how this relates to the so-called "truth-gap" approaches to the theory of truth and how it can be extended to other defeasible formalisms.

1 Basic Definitions in Default Logic

We are going to work with a classical propositional language $\mathcal{L}$, of the usual sort. In particular, $\mathcal{L}$ is obtained from propositional constants $p, q, r, \ldots$ using the propositional connectives $\wedge, \vee, \neg, \rightarrow$. We will employ the classical notion of logical consequence, according to which a sentence φ is a logical consequence of a set of sentences S if and only if φ is true on any truth-value assignment on which all sentences in S are true. Similarly, φ is inconsistent with a set S of sentences if and only if $\neg\varphi$ is a consequence of S.

As mentioned, default logic is intended to represent defeasible inferences of the the following form: "*If Tweety is a bird, then in the absence of information to the effect thet Tweety is a penguin, infer that Tweety flies*". The inference is defeasible because when we acquire further information to the effect that Tweety is indeed a penguin, the conclusion of the inference is retracted. Such a defeasible inference rule can be represented by a "default" of the following form:

$$\frac{\texttt{Tweety is a bird} : \texttt{Tweety is not a penguin}}{\texttt{Tweety flies}}.$$

In general, a default δ is an expression of the form

$$\frac{\zeta : \eta}{\theta},$$

where ζ, η, θ are sentences from $\mathcal{L}$. The expressions ζ, η, θ are called the *pre-requisite*, the *justification*, and the *conclusion* of δ, respectively. The intuitive meaning of δ is that if ζ is known, and η is consistent with what is known, then we are entitled to infer θ ("by default"). To say that η is consistent with what is known is to say that we have no information to the effect that $\neg\eta$ is true.

In the particular case in which a default δ has no pre-requisite or, equivalently, its prerequisite is a tautology, we say that δ is *categorical.* Other interesting classes of defaults have been identified. Reiter [17] singles out the notion of a *normal* default: this is a default whose justification is the same as its conclusion. Similarly, Reiter & Criscuolo [18] define *semi-normal* defaults to be defaults of the form

$$\frac{\zeta : \eta \wedge \theta}{\theta},$$

in which the conclusion occurs as a conjunct in the pre-requisite.

Given the above construal of defaults, we can introduce the notion of a default theory. By a *default theory* we understand a pair (W, Δ), where W is a set of propositional axioms in $\mathcal{L}$ (a "world-description") and Δ is a *finite* set of defaults $\delta_1, \ldots, \delta_n$. A default theory (W, Δ) is *categorical*, *normal*, or *semi-normal*, according as all defaults in Δ are categorical, normal, or semi-normal. As we will see, categorical default theories form a natural and well-behaved class.

We are now going to lay down some of the basic definitions and terminological conventions. Let (W, Δ) be a default theory, S a set of $\mathcal{L}$-sentences, and δ a default from Δ. Then we say:

1. δ is *admissible* in S if and only if the prerequisite of δ is a consequence of $S \cup W$;
2. δ is *conflicted* in S if and only if the conclusion of δ is inconsistent with $S \cup W$;
3. δ is *pre-empted* in S if and only if the justification of δ is inconsistent with $S \cup W$.

If Γ is a set of defaults, we say that δ is admissible, conflicted, or pre-empted in Γ according as δ is admissible, pre-empted or conflicted in the set of all conclusions of defaults from Γ.

Reiter [17] defines the notion of an *extension* for a default theory. Intuitively, an extension for a default theory (W, Δ) is a deductively closed, consistent set of formulas containing W and the consequents of a maximal subset of Δ. An extension for a default theory is a set of sentences that we consider *warranted* on the basis of the theory (there is no implicit assumption here that such an extensions is unique: in principle a default theory might warrant more than one set of sentences). Here we provide the simpler definition for a categorical default theory, where the definition is simpler in two ways.

First of all, our extensions are going to be sets of defaults rather than sets of sentences. This move is by now fairly standard in the literature, and it has the advantage of making clear that what matters are the defaults "triggered" or "fired" in any extension. From this, we can recover the sets of sentences warranted by the extension simply by taking the logical consequences of the conclusions of defaults in the extension.

The second, more substantial simplification with respect to Reiter's definition is to restrict ourselves to categorical default theories. The general case of arbitrary default theories has the added complication of having to account for a kind of minimality of extensions. The point is that extensions for arbitrary default theories have to minimize, subject to other constraints, the number of admissible defaults, i.e., the number of defaults whose pre-requisite follows logically from the conclusions of other defaults in the extension. Here we sidestep the issue completely by considering only categorical default theories (the reader is referred to [2] for a treatment of the general case).

We say that a set Γ of defaults is a *classical extension* for a categorical default theory (W, Δ) if and only if it satisfies the following equation:

$$\Gamma = \{\delta \in \Delta : \delta \text{ is not pre-empted in } \Gamma\}.$$

It is well known that if Γ is a classical extension for (W, Δ) then the set of conclusions of defaults in Γ is inconsistent with W if and only if W is already inconsistent. Similarly, if Γ is a classical extension for (W, Δ) and δ is neither conflicted nor pre-empted in Γ, then δ is in Γ.

Given the above definition of an extension for a (categorical) default theory, it is worth noting that it is not obvious how to determine whether a default theory has an extension and, if so, how to construct one. First of all, not all default theories have extensions, as we can see by considering a default theory (W, Δ) where W is empty and Δ contains only the default

$$\delta = \frac{: p}{\neg p},$$

where p is atomic. Suppose Γ were an extension for the theory: then there are two cases, according as δ is in Γ or not. If it is, then the justification of δ, i.e., p, is inconsistent with the set of conclusions of defaults from Γ, so that δ is pre-empted in Γ and therefore cannot be in Γ after all (given that Γ satisfies the equation defining extensions). On the other hand, if δ is not in Γ, then there are no defaults in Γ so that no atomic sentence can be inconsistent with the set of conclusions of defaults in Γ, and in particular δ is not pre-empted in Γ; so δ is in Γ after all. This shows that, under the hypothesis that Γ is an extension, δ is in Γ of and only if it is not in Γ, an obvious contradiction. We conclude that no extensions exist.

On the other hand, default theories can have multiple extensions. To see this, consider the default theory (W, Δ), where W is empty and Δ comprises the two defaults:

$$\delta_1 = \frac{:p}{\neg q} \quad \text{and} \quad \delta_2 = \frac{:q}{\neg p}$$

It is immediate to check that no extension can trigger both defaults, otherwise they would both be pre-empted in such an extension. Moreover, any extension for the theory must trigger at least one of them: for if neither belonged to the extension, neither would be pre-empted in such an extension and therefore would have to belong to the extension after all. So there are two classical extensions Γ_1 and Γ_2, each triggering one default and pre-empting the other one.

This is a general phenomenon: default theories can have multiple classical extensions, that are all incomparable with respect to the subset relation. It follows, in particular, that there cannot be a unique minimal (classical) extension of a default theory. This fact turns out to be problematic if we are interested in defining a notion of defeasible consequence for a default theory, in analogy to the classical notion of logical consequence. Indeed, it would seem that we ought to be interested in such a notion, as expressed in desideratum **d1**, if we want not only to take the name default *logic* seriously, but also to develop a viable model of cognition. This means that we would like to define a relation $\mid\!\sim$ that a default theory has to the sentences that are warranted by it, interpret such a relation as *defeasible consequence*, and write, say, $(W, \Delta) \mid\!\sim \varphi$ just in case φ is a defeasible consequence of (W, Δ).

In general, given a default theory having multiple classical extensions, there are several ways in which we can define a relation $\mid\!\sim$: (*i*) we can decide to be *creduluous* and say that $(W, \Delta) \mid\!\sim \varphi$ precisely when φ follows from (the set of conclusions of defaults in) *some* extension of (W, Δ); (*ii*) we can arbitrarily pick an extension Γ among the many possible, and decide that $(W, \Delta) \mid\!\sim \varphi$ just in case φ follows from (conclusions of defaults in) Γ; or (*iii*) we can be *skeptical* and say that $(W, \Delta) \mid\!\sim \varphi$ just in case φ follows from (conclusions of defaults in) *all* extensions of (W, Δ).

All three alternatives have drawbacks. Alternative (*ii*) is not acceptable for the arbitrariness of our choice of Γ. Alternative (*i*) can lead us sometimes to endorse contradictory statements (as in the case of Γ_1 and Γ_2 above). Alternative (*iii*) is the one that best resonates with certain intuitions about defeasible reasoning, e.g., the fact that defeasible reasoners should be cautious in drawing their inferences (see Horty *et al.* [7] for a general argument in favor of skepticism in defeasible reasoning), but goes about it the wrong way. It is certainly a funny way to be skeptical to generate *all* possible extensions of a theory and then take the intersection. If feasibility of computation is an issue at all, this is by far the least resource-oriented approach, and certainly not a satisfactory implementation of skepticism. There seems to be no natural way using classical extensions to satisfy desideratum **d1** without pre-empting **d2**.

It is also important to observe that classical extensions for default theories cannot be constructed (when they exist) by means by any *cumulative process*, of the sort in which defaults are successively assessed for some kind of property which can guarantee their belonging to the extension being constructed. On the contrary, we first have to "guess" a set Γ of defaults and then check that it does indeed satisfy the equation defining extensions.

This makes the problem of determining whether a given default theory has an extension computationally hard, as is borne out by the investigations of Kautz & Selman [8]. Indeed, Papadimitriou & Sideri take this to be the *basic* computational problem in default logic ([15, p. 348]). It is certainly this kind of considerations that has motivated researchers in the area to identify ever broader classes of default theories for which the problem is tractable, and in particular classes of default theories that can be guaranteed always to have an extension.

One obvious source of intractability in default logic is given by the ubiquitous consistency checks (required to verify that a default δ is not pre-empted in the putative extension Γ). But they are by no means the only source. It is mainly through the work of Kautz & Selman [8] that it has become clear that a source of complexity just as important is the fact that there seems to be no other way to check for the existence of an extension than to check all possible sequences of applications of defaults. Kautz & Selman define *unary* default theories, i.e., default theories in a restricted language comprising essentially only conjunctions of atomic sentences and negations of atomic sentences. For such a language, consistency checks are computationally easy, since a sentence φ is consistent with sentences $\psi_1, \ldots, \psi_n$ if and only if no conjunct of φ occurs negated as a conjunct in any of the sentences $\psi_1, \ldots, \psi_n$ (and conversely no negative conjunct of φ occurs unnegated as a conjunct in any of the sentences $\psi_1, \ldots, \psi_n$).

Nonetheless, Kautz & Selman establish the following facts: (*i*) the problem of determining whether a unary default theory has an extension is NP-complete (i.e., by all accounts, intrinsically intractable); and (*ii*) the problem of determining whether a given literal belongs to some extension of a given unary default theory is also NP-complete. These results seem to point to the fact that classical extensions for default logic fail to meet desideratum **d2** as well.

This intrinsic intractability has led people to try to sidestep the problem by identifying classes of default theories for which extensions can always be guaranteed to exist. For instance, one such class is the class of *ordered* default theories, identified by Etehrington [3]. Etherington's approach has been generalized by Papadimitriou & Sideri [15], who identify the broader class of the "even" default theories. The reader is referred to [2] or the original papers for more details, but it is in any case difficult to imagine how an even broader class of default theories that always have extensions could be identified. This is to say that the result of Papadimitriou & Sideri seems optimal, insofar as it goes.

It should be recalled that the notion of extension employed in default logic is intrinsically "two-valued," in the sense that it contains the consequences of a maximal set of defaults whose justifications are consistent with the extension itself. In other words, the triggering of a default can only be prevented if its justification is explicitly refuted. The approach to default logic that will be presented below seeks to circumvent this restriction by identifying a "three-valued" notion of extension for default logic, analogous to the one put forward in [1] for defeasible inheritance networks, and equally inspired by Kripke's [9] approach to the theory of truth. In particular, we will provide a notion of extension that (*i*) subsumes the classical one and (*ii*) according to which *any* default theory has an extension.

The present proposal embodies a "cautious" or "skeptical" approach to default logic, which in turn is substantiated in a three-valued notion of *general extension* introduced below. A default δ can be prevented from being triggered if and only if it is either explicitly conflicted or *potentially pre-empted*, whereas with Reiter's notion of extension, δ can be prevented from being triggered if and only if it is *explicitly pre-empted*, and within this difference lies the more cautious nature of general extensions. Moreover, contrary to Reiter's extensions, extensions of the sort proposed here can be obtained as the limit of a genuine inductive or cumulative construction. This feature makes general extensions somewhat better behaved mathematically, e.g., by imposing on general extensions a non-trivial algebraic structure, witness the fact that there is a *least* general extension for many default theories. In contrast, any two of Reiter's extensions for a given default theory are incomparable with respect to the subset relation. This leaves hope that at least in some cases the computational intractability characterizing default logic might be to some extent obviated.

2 General Extensions for Categorical Default Logic

We are now ready to introduce our notion of "general extension" for (categorical) default theories. A classical extension plays a two-fold role: on the one hand it explicitly comprises the defaults that are triggered (and these are the defaults that belong to the extension), and on the other hand it implicitly specifies the set of defaults that are ruled out (because they do not belong to the extension). The basic intuition at the basis of the notion of a general extension is that these two roles are going to be decoupled, by explicitly identifying a set of defaults as triggered as well as explicitly identifying a set of defaults as ruled out. This leads to the following definition.

A *general extension* for a categorical default theory (W, Δ) is a pair (Γ^+, Γ^-) of sets of defaults from Δ, simultaneously satisfying the following two equations:

$$\Gamma^+ = \{\delta : \delta \text{ is neither conflicted in } \Gamma^+ \text{ nor pre-empted in } \Delta - \Gamma^-\};$$
$$\Gamma^- = \{\delta : \delta \text{ is conflicted or pre-empted in } \Gamma^+\}.$$

In other words, Γ^+ is the set of all defaults that are neither conflicted in Γ^+ nor pre-empted in $\Delta - \Gamma^-$, while Γ^- is the set of all defaults that are either conflicted or pre-empted in Γ^+. Notice that before allowing a default in Γ^+ we make sure that it is not potentially pre-empted, i.e., that it is not pre-empted by any other defaults *that have not already been explicitly ruled out.*

It is possible to show that general extensions indeed generalize the notion of classical extension, and that general extensions always exist. Given a categorical default theory (W, Δ) having a a classical extension Γ^+, we can obtain a general extension for (W, Δ) by putting

$$\Gamma^- = \{\delta : \ \delta \text{ conflicted or pre-empted in } \Gamma^+\}.$$

Then (Γ^+, Γ^-) is a general extension for (W, Δ). (The reader is referred to [2] for a proof.)

The second important observation that we need to make is that every categorical default theory has a general extension. Here we skecth some of the ideas of the proof. Let (W, Δ) be a categorical default theory; we can assume that W is consistent, because otherwise $(\emptyset, \Delta)$ is a general extension. It is possible to obtain a general extension (Γ^+, Γ^-) as the limit of a process that occurs in stages. In particular, Γ^+ is constructed as the union of a sequence $\Gamma^+_0, \Gamma^+_1, \ldots$; similarly, Γ^- is also constructed as the union of the sequence $\Gamma^-_0, \Gamma^-_1, \ldots$; in turn, the sets Γ^+_n and Γ^-_n are inductively defined using an iterative process similar to the one in [6] or [1].

At the beginning of the process we assume that no defaults are triggered and no defaults are ruled out, by putting $\Gamma^+_0 = \Gamma^-_0 = \emptyset$. For the inductive step, let Γ^+_{n+1} be a *maximal* set of defaults such that:

1. The conclusions of defaults in $\Gamma^+_n \cup \Gamma^+_{n+1}$ form a consistent set;
2. no default $\delta \in \Gamma^+_{n+1}$ is pre-empted in $\Delta - \Gamma^-_n$.

(Observe that such a maximal set of defaults always exists, although it might not be unique.) For Γ^-_{n+1}, put

$$\Gamma^-_{n+1} = \{\delta : \delta \text{ conflicted or pre-empted in } \Gamma^+_{n+1}\}.$$

Then, it is possible to show that $\Gamma^+_n \subseteq \Gamma^+_{n+1}$ and $\Gamma^-_n \subseteq \Gamma^-_{n+1}$, by induction on n. This establishes that the sequences Γ^+_n and Γ^-_n (for $n \geq 0$) are increasing. So it is natural to take the limit of such sequences, by putting:

$$\Gamma^+ = \bigcup_{n\geq 0} \Gamma^+_n;$$
$$\Gamma^- = \bigcup_{n\geq 0} \Gamma^-_n.$$

At this point it is a matter of calculation to verify that (Γ^+, Γ^-) is an extension for (W, Δ). It is perhaps worth mentioning that this part requires the hypothesis that Δ contains finitely many defaults, and that this is the only point where the hypothesis is invoked.

It is worth noting that the construction given above, which provides a general extension for any categorical default theory is *non-deterministic*, in that at different stages different choices are possible. These choices occur in the definition of Γ^+_{n+1}, whenever we have to supply a maximal sets of defaults satisfying certain conditions. Since in principle more than one such maximal set might exist, we obtain different general extensions for the same default theory.

However, any non-determinism is eliminated if we assume that no default can be conflicted (with respect to any set of sentences) without being already pre-empted, and this can be achieved by restricting ourselves to semi-normal default theories. In fact, in such theories, the conclusion of any default occurs as a conjunct in the justification, so that if the default is conflicted then it's already pre-empted.

This fact allows us to modify the above construction by setting Γ^+_{n+1} equal to the set

$$\{\delta : \delta \text{ not pre-empted in } \Delta - \Gamma^-_n\};$$

since no default can be conflicted without being already pre-empted, this assures that no default is conflicted, and in particular that the conclusions of defaults in Γ^+_{n+1} form a consistent set.

The upshot of this is that as long we restrict ourselves to semi-normal (categorical) default theories, then these default theories have a *unique* minimal general extension, i.e., the extension produced by the above mentioned construction. As we will see, this is crucial for the satisfaction of desideratum **d1**.

3 Examples

Consider first the default theory (W, Δ), where W is empty and Δ comprises the two defaults:

$$\frac{: p}{\neg q} \quad \text{and} \quad \frac{: q}{\neg p}.$$

As we have seen, this theory has two classical extensions, according to which default is triggered. In addition to these, the theory has one general extension, in which no default is triggered and none is ruled out.

Similarly, the default theory in which W is empty, and Δ comprises only the default

$$\frac{: p}{\neg p},$$

has one general extension, namely, $(\emptyset, \emptyset)$, in which the default is neither triggered nor ruled out. We already know that such a theory has no classical extensions.

It might seem that minimal extensions are quite uninteresting, given that no defaults are triggered and none are ruled out. It turns out that this is not the case. Consider for instance the theory where W is empty, but Δ comprises the defaults

$$\frac{:p}{\neg p} \quad \text{and} \quad \frac{:q}{r}$$

The second default has nothing to do with the first, and so it should be triggered in any extension. However, the presence of the first default preents the theory from having any classical extension. On the other hand, the theory does have one general extension, namely

$$(\frac{:q}{r}, \emptyset).$$

Indeed, the first default cannot be triggered in any extension, but there is no obstacle that prevents triggering the second, which is in fact contained in all extensions of the theory, including any minimal one.

Now, consider a different sort of theory, in which W is empty and $\Delta = \{\delta_1, \ldots, \delta_n\}$, for some $n > 1$. Suppose also that for all k such that $1 \leq k < n$,

$$\delta_k = \frac{:p_k}{\neg p_{k+1}},$$

whereas

$$\delta_n = \frac{:p_n}{\neg p_1}.$$

So, Δ is a sequence of defaults, each one of which pre-empts the next one, and the last one of which pre-empts the first. It is easy to check that this theory has no classical extensions if n is odd. (It has, of course one general extension even when n is odd, namely $(\emptyset, \emptyset)$.) On the other hand, if n is even, say $n = 2m$, beside the general extension $(\emptyset, \emptyset)$, there are two classical extensions, namely

$$(\{\delta_{2k+1} : 0 \leq k < m\}, \{\delta_{2k+2} : 0 \leq k < m\}),$$

and

$$(\{\delta_{2k+2} : 0 \leq k < m\}, \{\delta_{2k+1} : 0 \leq k < m\}).$$

For the details the reader is referred to [1, 15].

Finally, consider the theory (W, Δ) where again W is empty, and Δ contains the two defaults

$$\frac{:q}{p} \quad \text{and} \quad \frac{:q}{\neg p}.$$

This theory has two classical extensions (according to which default is fired), and these are also the only two general extensions. This example is instructive because it shows that even *minimal* general extensions need not be unique. It is however easy to change the above theory in such a way that no default is conflicted in any extension unless it is already pre-empted in it. This amounts to turning the theory into a *semi-normal* default theory by a maneuver first identified by Reiter & Criscuolo [18]. A (categorical) default is *semi-normal* if it has the form

$$\frac{:\eta \wedge \theta}{\theta}.$$

We can turn the above theory into a semi-normal one by replacing Δ by the set of defaults

$$\frac{:q \wedge p}{p} \quad \text{and} \quad \frac{:q \wedge \neg p}{\neg p}.$$

This theory gains a unique minimal general extension in which no default is triggered.

4 Assessment and Related Work

How does the present proposal fare with respect to the two desiderata **d1** and **d2**? We have already remarked that classical extensions do not seem to allow us to define a notion of defeasible consequence in a natural way. The case is not dissimilar for general extensions. As the last example of the previous section shows, there are default theories for which we cannot single out any privileged general extension, which makes it difficukt to define a natural notion of defeasible consequence. It is still true that every theory has a general extension although it might fail to have a classical extension. This is an important feature, given the intuition that every default theory should license at least some inferences, but it does not seem to help toward the satisfaction of desideratum **d1**.

Things are different when we consider semi-normal default theories. In fact, given any theory, we can always switch to a semi-normal theory in the manner advocated by Reiter & Criscuolo [18]. This switch in inconsequential we only consider classical extensions, since semi-normal default theories need not have any classical extension (see [18, §3]). But the switch becomes critical when general extensions are allowed. We have already mentioned that semi-normal default theories always have a *unique* minimal general extension. This allows us, for any given semi-normal (categorical) default theory (W, Δ), to define $(W, \Delta) \mid\!\sim \varphi$ if and only φ is a logical consequence of conclusions of defaults in Γ, where Γ is the unique minimal extension of (W, Δ).

This takes us at least some of the toward the satisfaction of desideratum **d1**, at least in the case of semi-normal default theories. This also leaves hope that this notion of defeasible consequence for semi-normal (categorical — but this restriction can be lifted) default theories might turn out to be feasible. Certainly, general extensions can be constructed from the ground up, which makes one think that the stage process by which this is achieved might be implementable in a resource-conscious fashion.

What we have being saying so far applies to categorical default theories. However, this restriction is unessential, and can be lifted at the cost of complicating our definitions and proofs somewhat. Such complication occurs orthogonally to the issues involved in the satisfaction of the two desiderata **d1** and **d2**, and is therefore better left out in a preliminary presentation. The reader is referred to [2] for a full treatment of arbitrary default theories. (The existence of unique minimal extensions for semi-normal categorical default theories extends smoothly to semi-normal default theories.)

As mentioned at the beginning inspiration for this approach derives from Kripke's work [9] in the theory of truth. In turn, Kripke's approach exploits a construction originally due to Kleene (see Feferman [4] for a strengthening of the Kripke construction as well as a historical appraisal of its predecessors). Kripke considers a language containing its own truth predicate, avoiding paradox by switching to a three-valued setting. In such a setting, a sentence is true, false or indeterminate according as its negation is false, true or indeterminate. This intuition (extended to the other connectives in the language) allows one to carry out an inductive cosntrauction not too dissimilar from the one that gives minimal general extensions for default theories. By weakening our semantics Kripke achieved a better-behaved model of how natural language can contain its own truth predicate, and the maneuver carried out here is similar in inspiration.

Indeed, similar approaches have been put forward in the case of logic programs. Fitting [5] offers a general framework within which to develop a three- or multi-valued semantics for logic programs, and connections between logic programs and defeasible formalisms were noticed by Przymusinski [16]. The reader is referred to [2] for a full assessment of the connections between the present proposal and those accounts. For now suffice it to say that the present account is three-valued only in spirit, and that the underlying semantics is thoroughly classical. This marks a noticeable difference with the work of Fitting and Przymusinski. Moreover, the possibility of defining a natural notion of defeasible consequence for semi-normal default theories and its relation to the desiderata **d1** and **d2** appears to be a novel and welcome feature of the proposal given here.

References

[1] G.A. Antonelli, *Defeasible Inheritance on Cyclic Networks*, manuscript, Yale Unversity, 1995.

[2] G.A. Antonelli, *General Extensions for Default Logic*, manuscript, Stanford University, 1996.

[3] D. Etherington, *Formalizing Nonmonotonic Reasoning Systems*, **Artificial Intelligence** 31 (1987), pp. 41–85.

[4] S. Feferman, *Toward Useful Type-Free Theories, I*, in R. Martin (ed.), **Recent Essays on Truth and the Liar Paradox**, Oxford Univeristy Press, Oxford 1984, pp. 237–287.

[5] M. Fitting, *The Family of Stable Models*, **J. of Logic Programming** 17 (1993), pp. 197-225.

[6] John F. Horty, *Some Direct Theories of Nonmonotonic Inheritance*, in D.M. Gabbay, C.J. Hogger & J.A. Robinson(eds), **Handbook of Logic in Artificial Intelligence and Logic Programming**, vol. 3: **Nonmonotonic Reasoning and Uncertain Reasoning**, Oxford University Press, Oxford 1994.

[7] J.F. Horty, R.H. Thomason, & D.S. Touretzky, *A Skeptical Theory of Inheritance in Nonmonotonic Semantic Networks*, **Artificial Intelligence** 42 (1990), pp. 311–48.

[8] H. Kautz & B. Selman, *Hard problems for simple default logic*, **Artificial Intelligence** 49 (1991), pp. 243–79.

[9] S. Kripke, *Outline of a Theory of Truth*, **The Journal of Philosophy** 72 (1975), pp. 690–716.

[10] J. McCarthy, *Epistemological Problems of Artificial Intelligence*, **Proceedings of the Fifth International Joint Conference on Artificial Intelligence** MIT Press, Cambridge, Mass. 1977.

[11] J. McCarthy, *Circumscription — a Form of Non-Monotonic Reasoning*, **Artificial Intelligence** 13 (1980), pp. 27–39, 171–172.

[12] J. McCarthy, *Artificial Intelligence and Logic*, in R. Thomason (ed.), **Philosophical Logic and Artificial Intelligence**, Kluwer Academic Press, Dordrecht 1989.

[13] D. McDermott & J. Doyle, *Non-Monotonic Logic I*, **Artificial Intelligence** 13 (1980), pp. 41–72.

[14] R. C. Moore, *Semantical Considerations on Nonmonotonic Logic*, **Artificial Intelligence** 25 (1985), pp. 75–94.

[15] C.H. Papadimitriou & M. Sideri, *Default Theories that Always Have Extensions*, **Artificial Intelligence** 69 (1994), pp. 347–57.

[16] T. Przymusinski, *Three-valued nonmonotonic formalisms and semantics of logic programs*, **Artificial Intelligence** 49 (1991), pp. 309–43.

[17] R. Reiter, *A Logic for Default Reasoning*, **Artificial Intelligence** 13 (1980), pp. 81–132.

[18] R. Reiter & G. Criscuolo, *On Interacting Defaults*, **Proceedings of the Seventh International Joint Conference on Artificial Intelligence**, Vancouver, B.C. (1981), pp. 270–276.

[19] B. Selman, *Tractable default reasoning*, PhD Thesis, University of Toronto (1990).