A Critical Reexamination of Default Logic, Autoepistemic Logic, and Only Knowing*

Joseph Y. Halpern

IBM Research Division
Almaden Research Center, Dept. K53/802
San Jose, CA 95120
email: halpern@almaden.ibm.com

Abstract

Fifteen years of work on nonmonotonic logic has certainly increased our understanding of the area. However, given a problem in which nonmonotonic reasoning is called for, it is far from clear how one should go about modeling the problem using the various approaches. We explore this issue in the context on two of the best-known approaches, Reiter's default logic [1980] and Moore's autoepistemic logic [1985], as well as two related notions of "only knowing", due to Halpern and Moses [1984] and Levesque [1990]. In particular, we return to the original technical definitions given in these papers, and examine the extent to which they capture the intuitions they were designed to capture.

1 Introduction

Fifteen years of work on nonmonotonic logic has certainly increased our understanding of the area. However, given a problem in which nonmonotonic reasoning is called for, it is far from clear how one should go about modeling the problem using the various approaches. Moreover, after modeling the problem, we can rarely be sure that we will get all and only the conclusions that we "ought" to get, even in cases where there is agreement about

*Research sponsored in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080. The United States Government is authorized to reproduce and distribute reprints for governmental purposes. A preliminary version of this paper appears in *Computational Logic and Proof Theory (Proceedings of the Kurt Gödel Symposium*, Lecture Notes in Computer Science, vol. 713, Springer-Verlag, 1993, pp. 43–60. This version is very similar to one that will appear in *Computational Intelligence* in 1997.

what these are (cf. the discussion in [Sombé 1990]). We explore this issue in the context of two of the best-known approaches to nonmonotonic reasoning, Reiter's default logic [1980] and Moore's autoepistemic logic [1985], as well as two related notions of "only knowing", due to Halpern and Moses [1984] and Levesque [1990].

Our approach is to return to the original technical definitions given in these papers, and examine the extent to which they capture the intuitions they were designed to capture. For example, we show that several variations of Reiter's original definition of extension are all equally consistent with the intuitions presented in his paper. We show that one leads to autoepistemic logic and Levesque's "only knowing" approach, while another leads to the "only knowing" approach of Halpern and Moses. These connections can be viewed as extending results of Konolige [1988] and Marek and Truszczyński [1989, 1993]. The goal of this paper, however, is not so much to establish connections, but to point out gaps in our intuitions and understanding, and to show how the same intuition can lead in a number of different directions.

2 Default Logic

We start by considering Reiter's default logic. Although Reiter considers first-order default logic, we consider only propositional default logic here, to simplify the exposition. It is straightforward to extend all of our arguments to the first-order case.

A *default* is an expression of the form $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma}$, where α, $\beta_1, \ldots, \beta_m$, and γ are propositional formulas. We call α the *prerequisite* of the default, $\beta_1, \ldots, \beta_m$ its *justifications*, and γ its *conclusion*. Roughly speaking, such a default can be interpreted as "if α holds and it is consistent to believe each of $\beta_1, \ldots, \beta_m$, then one may believe γ" [Reiter 1980]. A default theory Δ is a pair (W, D), where W is a set of propositional formulas and D is a set of defaults.

Reiter starts by listing three rather noncontroversial properties that any acceptable set E of beliefs based on a default theory $\Delta = (W, D)$ ought to have:

C1. It should contain W: $W \subseteq E$.

C2. It should be deductively closed: $Th(E) = E$.[1]

C3. It should be *closed under the defaults in* D: if $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in E$, and $\neg\beta_1, \ldots, \neg\beta_m \notin E$ (so that each of $\beta_1, \ldots, \beta_n$ is consistent with E), then $\gamma \in E$.

We say that any set E satisfying C1–C3 above is *consistent with* Δ.

[1] A theory E is *deductively closed* with respect to a logic $\mathcal{S}$ if, whenever $\alpha_1, \ldots, \alpha_n \in E$ and $(\alpha_1 \wedge \ldots \wedge \alpha_n) \Rightarrow \beta$ is provable in $\mathcal{S}$, then $\beta \in E$. Unless we explicitly mention a logic, we take deductive closure to be with respect to propositional logic. $Th(E)$ is the deductive closure of E with respect to propositional logic.

Reiter is not interested in all sets consistent with Δ. Rather, he focuses only on certain consistent sets. Immediately after defining the three properties that characterize consistent sets, he goes on to define a function Γ, that we denote here Γ_1^Δ. (The superscript Δ emphasizes the dependence on Δ, and the subscript 1 is to pave the way for the definitions of Γ_2^Δ, Γ_3^Δ, and Γ_4^Δ below.) Γ_1^Δ is a function that, given a set of propositional formulas, returns another set of propositional formulas. If S is a set of propositional formulas, then $\Gamma_1^\Delta(S)$ is the least set S' satisfying the following three properties:

D1. $W \subseteq S'$

D2. $Th(S') = S'$

D3. If $\frac{\alpha:M\beta_1,\ldots,M\beta_m}{\gamma} \in D$, $\alpha \in S'$, and $\neg\beta_1,\ldots,\neg\beta_m \notin S$, then $\gamma \in S'$.

A set E of formulas is an *extension* of Δ if $\Gamma_1^\Delta(E) = E$, i.e., if E is a fixed point of the operator Γ_1^Δ. Reiter views the extensions of Δ as describing all possible "acceptable" sets of beliefs based on Δ.

In [Reiter 1980], a number of properties of the fixed points of Γ_1^Δ are described. However, the question that is never asked or answered in [Reiter 1980] is precisely why this particular operator Γ_1^Δ is considered. What is the intuition behind it? The fixed points of Γ_1^Δ consist of some, but (as we shall see) not all of the sets consistent with Δ. There are other operators quite similar to Γ_1^Δ that give other sets of fixed points (all of which are also consistent with Δ). So why should we be particularly interested in the fixed points of Γ_1^Δ? This question is particularly important given the number of known examples of the occasionally counterintuitive nature of extensions (see below).

To understand the issue better, let us consider a number of variants of Γ_1^Δ. We focus on two issues here:

1. What is the role of the set S in $\Gamma_1^\Delta(S)$? One way to think of it is as a tentative set of beliefs, or a context, against which to compute the defaults. In that case, $\Gamma_1^\Delta(S)$ is what you should believe, given context S. But if S is a tentative set of beliefs, why don't we require that $S \subseteq \Gamma_1^\Delta(S)$?

2. In D3, for a default rule of the form $\frac{\alpha:M\beta_1,\ldots,M\beta_m}{\gamma}$, why do we use S' for α, and S for $\beta_1,\ldots,\beta_m$? If S is meant to be a tentative set of beliefs, why not use S both for α and $\beta_1,\ldots,\beta_m$?[2]

By varying the two parameters discussed above, we are led to three new operators, that we denote Γ_2^Δ, Γ_3^Δ, and Γ_4^Δ. Γ_2^Δ is the same as Γ_1^Δ, except that we replace D3 by D3$'$, where the clause "$\alpha \in S'$" is replaced by "$\alpha \in S$". Γ_3^Δ is the same as Γ_1^Δ, except that we add the clause

[2]Alternatively, we could use S' for both. To cut down the number of alternatives, we have ignored this possibility here.

D0. $S \subseteq S'$.

Finally, Γ_4^Δ is the same as Γ_3^Δ, except that D3 is replaced by D3′.[3]

Let $fp(\Gamma_i^\Delta)$ be the set of fixed points of Γ_i^Δ, $i = 1, \ldots, 4$. Notice that $fp(\Gamma_1^\Delta)$ consists of precisely Reiter's extensions. It is straightforward to check that the fixed points of $\Gamma_1^\Delta, \ldots, \Gamma_4^\Delta$ are all consistent with Δ.

Proposition 2.1: *All the sets in $fp(\Gamma_i^\Delta)$, $i = 1, \ldots, 4$ are consistent with Δ.*

Proof: See Appendix A. ∎

The following examples bring out some of the differences between these operators.

Example 2.2: Suppose that $\Delta = (\emptyset, \{\frac{true:Mp}{false}\})$. Then it is easy to see that $fp(\Gamma_1^\Delta) = fp(\Gamma_2^\Delta) = \emptyset$, while $fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta) = \{Th(W) : \neg p \in W\}$. To see this in the case of Γ_1^Δ, observe that if $\neg p \notin S$, then the default guarantees that $false \in \Gamma_1^\Delta(S)$, so that by D2, we have that $\Gamma_1^\Delta(S)$ is the inconsistent set of all formulas, and $S \neq \Gamma_1^\Delta(S)$. On the other hand, if $\neg p \in S$, then the default does not apply, and $\Gamma_1^\Delta(S)$ just consists of all propositional tautologies. Again, $S \neq \Gamma_1^\Delta(S)$. An identical argument works for Γ_2^Δ. On the other hand, it is easy to show that $fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta) = \{Th(W) : \neg p \in W\}$. Intuitively, the default says that if p is consistent, we should believe *false*. Since we never believe *false*, then we can conclude p is inconsistent, and thus $\neg p$ holds. It follows that any deductively closed set containing $\neg p$ will be consistent with Δ, and this is precisely what $fp(\Gamma_3^\Delta)$ and $fp(\Gamma_4^\Delta)$ give us. ∎

Example 2.3: Suppose $\Delta = (\emptyset, \frac{p:M\,true}{p})$. It is easy to see that $fp(\Gamma_1^\Delta) = \{Th(true)\}$. Thus, Δ has only one extension: the set of all tautologies. However, it is also easy to see that $fp(\Gamma_2^\Delta) = \{Th(true), Th(p)\}$, and that $fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta)$ consists of all deductively closed propositional theories. Notice that every deductively closed theory is in fact consistent with Δ. ∎

Example 2.4: Suppose that $\Delta = (\emptyset, \{\frac{true:Mp}{false}, \frac{\neg p \wedge q:M\,true}{\neg p \wedge q}, \frac{\neg p \wedge \neg q:M\,true}{\neg p \wedge \neg q}\})$. Combining the arguments given in the previous examples, it is not hard to show that $fp(\Gamma_1^\Delta) = \emptyset$, $fp(\Gamma_2^\Delta) = \{Th(\neg p \wedge q), Th(\neg p \wedge \neg q)\}$, and $fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta) = \{Th(W) : \neg p \in W\}\}$. Thus $fp(\Gamma_1^\Delta)$ differs from $fp(\Gamma_2^\Delta)$, and both differ from $fp(\Gamma_3^\Delta)$. ∎

Notice that in both of these examples, we have $fp(\Gamma_1^\Delta) \subseteq fp(\Gamma_2^\Delta) \subseteq fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta)$, and $fp(\Gamma_3^\Delta)$ and $fp(\Gamma_4^\Delta)$ consist of all theories consistent with Δ. Somewhat surprisingly, this is a general phenomenon.

[3] These are certainly not the only possible variants of Γ_1^Δ. Others are considered, for example, in [Brewka 1991; Mikitiuk and Truszczyński 1995; Schaub 1992]. The ones considered here are perhaps the simplest, however, and suffice to bring out the point that Reiter's original intutions could have easily led to a number of different definitions.

Theorem 2.5: *For all default theories Δ we have*

(a) $fp(\Gamma_1^\Delta) \subseteq fp(\Gamma_2^\Delta) \subseteq fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta)$,

(b) $fp(\Gamma_3^\Delta)$ *consists of all theories consistent with* Δ.

Proof: See Appendix A. ∎

Example 2.4 shows that the containments in part (a) are, in general, strict.

This result already tells us that if the goal is truly to capture consistent sets, as characterized by C1–C3, then we should be using Γ_3^Δ or Γ_4^Δ, rather than Reiter's Γ_1^Δ. Moreover, the result also shows that the definition of Γ_3^Δ (and similarly Γ_4^Δ) is robust, in that it does not depend on whether $\alpha \in S$ or $\alpha \in \Gamma_3^\Delta(S)$. Nevertheless, as the examples suggest, we typically do not want to consider *all* sets consistent with Δ. There are too many of them. We really do want to focus on only some of them, although perhaps not necessarily the ones picked out by Γ_1^Δ.

A good argument can be made that we should focus on *minimal* fixed points, the ones that are not subsets of any other fixed points. These are the ones where the agent can be viewed as knowing as little as possible beyond what is forced by the fact the agent's beliefs should be consistent with W. The case for minimal fixed points is strengthened when we look at things proof-theoretically (which is certainly consistent with Reiter's original intentions). If $\Delta = (W, \emptyset)$, so that there are no defaults, then $fp(\Gamma_1^\Delta) = Th(W)$; that is, $fp(\Gamma_1^\Delta)$ is the deductive closure of W, and thus consists precisely of what is provable from W (if we have a complete propositional proof system). By way of contrast, $fp(\Gamma_3^\Delta) = \{Th(W') : W' \supseteq W\}$; that is, $fp(\Gamma_3^\Delta)$ consists of all deductively closed theories containing W. Note, however, that the only *minimal* fixed point of $fp(\Gamma_3^\Delta)$ is $Th(W)$.

Let $mfp(\Gamma_i^\Delta)$ denote the set of minimal fixed points of Γ_i^Δ, $i = 1, 2, 3$. Then the following result describes the relation between the minimal fixed points of each of the operators we are considering.

Theorem 2.6: *For all default theories Δ, we have*

(a) $mfp(\Gamma_1^\Delta) = fp(\Gamma_1^\Delta)$,

(b) $mfp(\Gamma_1^\Delta) \subseteq mfp(\Gamma_2^\Delta)$,

(c) $mfp(\Gamma_1^\Delta) \subseteq mfp(\Gamma_3^\Delta) = mfp(\Gamma_4^\Delta)$.

Proof: Part (a) follows from Theorem 2.4 in [Reiter 1980]. Parts (b) and (c) are proved in Appendix A. ∎

Example 2.4 shows that, in general, the containment in parts (b) and (c) is strict, and that $mfp(\Gamma_2^\Delta)$ and $mfp(\Gamma_3^\Delta)$ are in general incomparable.

While we are often interested in minimal fixed points, as the following example, due to Poole [1989], shows, we are not necessarily always interested in minimal fixed points.

Example 2.7: Suppose $\Delta = (p_1 \vee p_2, \{\frac{true:M(q_1 \wedge \neg p_1)}{q_1}, \frac{true:M(q_2 \wedge \neg p_2)}{q_2}\})$. (In Poole's example, p_1 is *Broken(Left-arm)*, p_2 is *Broken(Right-arm)*, q_1 is *Usable(Left-arm)*, and q_2 is *Usable(Right-arm)*.) It is not hard to show that $mfp(\Gamma_1^\Delta) = mfp(\Gamma_2^\Delta) = \{Th((p_1 \vee p_2) \wedge q_1 \wedge q_2)\} \subseteq mfp(\Gamma_3^\Delta)$. This seems unreasonable: We know that one of p_1 or p_2 must hold, which suggests that one of the defaults should be disabled. Yet, in the only minimal fixed point of Γ_1^Δ and Γ_2^Δ, both defaults "fire". More reasonable seem to be the theories $Th(p_1 \wedge q_2)$ and $Th(p_2 \wedge q_1)$. Both are in fact minimal fixed points of $fp(\Gamma_3^\Delta)$, although neither is in $fp(\Gamma_1^\Delta)$ or $fp(\Gamma_2^\Delta)$ (since in this case $fp(\Gamma_1^\Delta) = mfp(\Gamma_1^\Delta) = fp(\Gamma_2^\Delta)$). Even if we focus on $fp(\Gamma_3^\Delta)$, it would be nice to have a mechanism that allows us to ignore $Th((p_1 \vee p_2) \wedge q_1 \wedge q_2)$. ∎

What can we say about default reasoning in light of these results? We see that there are a number of variants of Reiter's definition, each of which specifies (typically different) sets of formulas consistent with a given default theory Δ. Other variants appear in the literature (see, for example, [Brewka 1991; Delgrande and Jackson 1991; Łukaszewicz 1988; Mikitiuk and Truszczyński 1995; Rounds and Zhang 1993; Schaub 1992]). Why should we prefer one to the other? Marek and Truszczyński [1989, 1993] provide an elegant proof-theoretic characterization of Γ_1^Δ showing that it can be viewed as, in some sense, the outcome of applying a particular proof procedure using default rules. This can be viewed as providing some justification for Reiter's original definition. However, as we show in Appendix B, minor variations of Marek and Truszczyński's characterization give natural proof-theoretic characterizations of Γ_2^Δ and Γ_3^Δ (and thus Γ_4^Δ) as well. While there is a sense in which Γ_1^Δ can be viewed as giving more "grounded" proofs (see the discussion in Appendix B), if we are to understand default reasoning proof theoretically, it seems that we still need a better understanding of what counts as a proof, and what it means to be "grounded". As Poole's example and others show, this issue is a subtle one. Perhaps semantic models corresponding to each of these proof methods might help clarify matters.

We conclude this section with a comparison of our results and those of Marek and Truszczyński [1989, 1993]. As mentioned above, Marek and Truszczyński provide a proof-theoretic characterization of default reasoning. More precisely, they define a notion of *strong proof* from a set of defaults and observe that it follows from results of Reiter [1980] that it can be used to characterize the fixed points of Γ_1^Δ. (The details are reviewed in Appendix B.) They also define a notion of *weak proof*. As shown in Appendix B (Theorem B.2), weak proofs can be used to characterize the fixed points of Γ_2^Δ. Once this connection between weak proofs and Γ_2^Δ is established, the containment $fp(\Gamma_1^\Delta) \subseteq fp(\Gamma_2^\Delta)$ in part (a) of Theorem 2.5 and the containment $mfp(\Gamma_1^\Delta) \subseteq mfp(\Gamma_2^\Delta)$ in part (b) of Theorem 2.6 also follows from results of [Marek and Truszczyński 1989; Marek and Truszczyński 1993]. Marek and Truszczyński also define a notion of a *minimal set for* Δ. It follows from part (b) of Theorem 2.5 that $mfp(\Gamma_3^\Delta) = mfp(\Gamma_4^\Delta)$ consists of precisely the minimal sets for Δ. In [Marek and Truszczyński 1993, Theorem 3.92], it is shown that every extension of Δ is a minimal set. Given the connection between minimal sets and

$mfp(\Gamma_3^{\Delta})$, this is just part (c) of Theorem 2.6.

3 Autoepistemic Logic

Although notions of knowledge and belief do not appear explicitly in the technical definitions of default logic, they certainly appear in all the intuitions behind it. The justifications of a default are typically taken to be facts consistent with what we know. Indeed, the M in Reiter's notation $M\beta$ for the justification of a default is a standard notation for the dual of the modal operator L for belief, so that $M\beta$ means $\neg L\neg\beta$, which can be read "it is not the case that the agent believes $\neg\beta$" or "β is consistent with the agent's beliefs". McDermott and Doyle [1980] were perhaps the first to consider a nonmonotonic modal logic, where beliefs are made explicit using a modal operator. Moore's Autoepistemic (AE) logic is an attempt to provide a rational reconstruction of McDermott and Doyle's logic. We briefly review Moore's semantics of autoepistemic logic here.

An AE theory T is viewed as the total set of beliefs held by an agent. The question that AE logic is interested in answering is "What are the possible (or, perhaps better, justified) belief sets for an agent with initial premises A?"

A *propositional interpretation* is just an assignment of truth values to formulas in the language that is consistent with the usual semantics for $\wedge$ and $\neg$ in propositional logic, and an arbitrary assignment of truth values to primitive propositions and to formulas of the form $L\varphi$, where φ is an arbitrary formula. By treating formulas of the form $L\varphi$ just as it does primitive propositions, an AE interpretation makes no *a priori* assumptions about the properties of belief. It is quite possible for an AE interpretation to assign Lp the truth value **true** and to assign $L(p \wedge p)$ the truth value **false**.

We view a propositional interpretation as a description of what is true in the world. Recall that we view T as the set of beliefs of the agent. Thus, we want to have $\varphi \in T$ iff $L\varphi$ is true in the world. To capture this intuition, Moore defines an *AE interpretation of* T to be a propositional interpretation in which $L\varphi$ is true iff φ is in T. An AE interpretation for T describes a world where T is the full set of beliefs of an agent. An *AE model of* T is an AE interpretation of T where the formulas in T are true. Not every AE theory has an AE model; for example, $\{Lp\}$ does not: any AE interpretation of $\{Lp\}$ satisfies $\neg Lp$ (since $p \notin \{Lp\}$), and so cannot satisfy Lp and be an AE model of this formula. With these basic definitions, we can now consider what the reasonable belief sets are for an agent with initial premises A.

Although Moore's notions allow for reasoners with rather weak reasoning powers, he focuses on what he calls "ideally rational agents". These are agents with perfect introspective powers, who believe all the logical consequences of their beliefs. The fact that the agent believes all the logical consequences of her beliefs means that if she believes p and believes q, then she also believes $p \wedge q$. The fact that she has perfect introspective powers means that if she believes p, then she believes that she believes p, so that Lp is in her belief set if p is. Moreover, if she doesn't believe r, then she believes that she does not believe r, so that $\neg Lr$ is in her belief set if r is not.

To capture these intutions, Moore defines the notion of *semantic completeness*. T is *semantically complete* if T contains every formula that is true in every AE model of T. As Moore shows, an agent is ideally rational in his sense exactly if his belief set is semantically complete. To make this precise, following Stalnaker [1993], Moore defines a set T to be *stable* if the following three properties hold.

St1. T is deductively closed under propositional reasoning.

St2. If $\varphi \in T$, then $L\varphi \in T$.

St3. If $\varphi \notin T$, then $\neg L\varphi \in T$.

Note that St1 captures the fact that an agent's beliefs are closed under logical consequence, while St2 and St3 capture the agents' introspective powers. Moore then proves:

Lemma 3.1: [Moore 1985] *T is stable if and only if T is semantically complete.*

A stable set containing the formulas in A describes the beliefs of an ideally rational agent (in Moore's sense) who believes A. Moore does not consider all stable sets as being justified. A theory T is *sound with respect to A* if every AE interpretation of T in which all the formulas of A are true is an AE model of T. Thus, T is sound with respect to A if, in all worlds where agent believes exactly the formulas in T and the formulas in A are true, the formulas in T are true as well. The terminology "sound" is perhaps unfortunate here, since it suggests some notion of provability. Soundness is perhaps better thought of as a type of semantic notion of entailment: the truth of the formulas in A as well as the beliefs in T entails the truth of the formulas in T.

Moore defines T to be a *stable expansion of A* if T is a stable set containing the formulas in A which is sound with respect to A. Soundness is intended to capture Moore's notion of justification. How reasonable is it?

Konolige [1988] makes the case that the notion is too weak, in that it allows some stable sets which are not really justified. For example, as he observed, the formula $Lp \Rightarrow p$ has two stable expansions: T_1, whose only nonmodal formulas are the valid formulas, and T_2, whose nonmodal formulas consist of p and all its logical consequences. T_1 seems quite reasonable; it corresponds to an agent that has no nontrivial beliefs about the world. In T_2, the agent believes that p is true. According to Moore's approach, given that $Lp \Rightarrow p$ is true, then this is justified, since the belief in p causes p to be true. But is this really reasonable? Should the fact that the agent believes p be viewed as providing sufficient justification for p to be true, even in the presence of $Lp \Rightarrow p$?

Konolige [1988] introduced several stronger variants of the notion of expansion in order to provide stronger notions of justification, with the ultimate hope of providing a notion that corresponded to default logic.[4] But, as some of the examples provided by Konolige himself suggest, the notion of justification is no better understood than the notion of default proof. The attempt to connect default logic to AE logic helps clarify some of the issues we have raised here, so we briefly review the results and Konolige and the later work of Marek and Truszczyński.

Define the *kernel* of an AE theory T to be all the nonmodal formulas in T. It is well known [Moore 1985] that a stable set is determined by its kernel. Given a default $d = \frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma}$, we associate with it the modal formula $tr_L(d) = L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m \Rightarrow \gamma$, where M is an abbreviation for $\neg L \neg$. A default theory $\Delta = (W, D)$ is translated to $tr_L(\Delta) = W \cup \{tr_L(d) : d \in D\}$. It is easy to see that under this translation, the sets consistent with Δ—by Theorem 2.5, these are just the fixed points of Γ_3^Δ—are precisely the kernels of the stable sets containing $tr_L(\Delta)$. In fact, there is an even tighter connection with our earlier results, as the following proposition shows.

Theorem 3.2: *Let Δ be a default theory.*

(a) A set of formulas is a fixed point of Γ_3^Δ iff it is the kernel of a stable set containing $tr_L(\Delta)$.

(b) A set of formulas is a fixed point of Γ_2^Δ iff it is the kernel of a stable expansion of $tr_L(\Delta)$.

Proof: See Appendix C. We remark that in light of the connection between the fixpoints of Γ_2^Δ and notion of weak proof established in Theorem B.2 of Appendix B, part (b) of this result also follows from results of Marek and Truszczyński [Marek and Truszczyński 1993; Marek and Truszczyński 1993]. A direct proof is provided here as well, for the sake of completeness. ∎

This theorem shows that stable expansions in AE logic essentially correspond to the fixed points of the operator Γ_2. The difference between default logic and AE logic is thus essentially captured by the differences between the operators Γ_1 and Γ_2.

While Konolige's examples show that Moore's notion of justification might be too weak, other examples show that it might be too strong. As has often been observed (for example, in [Halpern and Moses 1984]), the formula Lp has no stable expansion. Intuitively, this is because, according to Moore's approach, the agent's belief in p has no justification. Some authors (e.g., Lifschitz, as quoted in [Shvarts 1990]) have argued that Lp should have a unique stable expansion, the same one as p.

The question of whether Lp "should" have a stable expansion is perhaps best thought of in the context of what seems to be a deeper question, namely, the role of the set A of premises. What does it mean for a formula to be in A? How does a formula get into A in

[4] As pointed out in [Marek and Truszczyński 1989], the variants considered in [Konolige 1988] did not quite realize this hope. However, another variant proposed later by Konolige [1989] and one proposed by Marek and Truszczyński [1989] do realize it.

the first place? Moore does not consider this question in [Moore 1985], but certainly one plausible notion is that A represents information that the agent has received from some trusted source.[5] Certainly we can imagine the trusted source telling an agent something about the external world, such as p. We can also make sense of the trusted source telling an agent a formula such as $p \Rightarrow Lp$ that connects the real world to the agents subjective world. Indeed, Moore's example "if I had a brother then I would know about it" has this form. But what would it mean for the trusted source to tell the agent "Lp"? How can the trusted source have access to the agent's internal state? If the trusted source says "Lp" and the agent does not in fact believe p, how should we interpret this? One interpretation is that it is nonsensical for the trusted source to say "Lp"; under this interpretation, it is not surprising that Lp has no stable expansion. But if it is nonsensical for the trusted source to say "Lp", should it not be equally nonsensical for the trusted source to say "$p \wedge Lp$"? Unfortunately, $p \wedge Lp$ has a unique stable expansion, which is the same as that of p.

Summarizing, the situation for AE logic looks very similar to that for default logic. We have the standard notion of stable expansion, and a number of variants (besides those discussed in [Konolige 1988], others can be found in [Konolige 1992; Marek and Truszczyński 1989; Marek and Truszczyński 1993; Shvarts 1990; Schwarz and Truszczyński 1992]; we discuss the approach of [Shvarts 1990; Schwarz and Truszczyński 1992] in more detail in Appendix D). All of these can be viewed as different ways of describing what it means for a stable expansion to be "justified", given A. Alternatively, under Konolige's translation, they can be viewed as different ways of choosing a theory consistent with a default theory. A better understanding of the role of A might lead to a better understanding of whether default logic or AE logic (or one of their many variants) is the appropriate approach to apply in a given situation.

4 Only Knowing

Intuitively, Moore's notion can be thought of as characterizing what an agents who knows (or believes) only A is justified in believing. Levesque [1990] tries to formalize this intuition with a logic for "only knowing". We discuss Levesque's formalism here, and an alternate approach to due to Halpern and Moses [1984], which we hereafter call the HM approach.

Levesque considers a K45 notion of belief, and introduces a modal operator O, where $O\alpha$ is read "only believes α".[6] The O operator is best understood in terms of another operator introduced by Levesque denoted N. While $L\alpha$ says "α is true at all the worlds that the agent considers possible", $N\alpha$ is viewed as saying "α is true at all the worlds that the agent does *not* consider possible". Then $O\alpha$ is defined as an abbreviation for

[5] We remark that Moore himself is comfortable with this interpretation of A [private communication, 1992]. But there are certainly other interpretations, as we shall see in the next section.

[6] For the reader not familiar with modal logic, we review the relevant logics in Appendix D.

$L\alpha \wedge N\neg\alpha$. Thus, $O\alpha$ holds if α is true at all the worlds that the agent considers possible, and only at these worlds. We can read $L\alpha$ as saying "the agent knows at least α", while $N\neg\alpha$ says "the agent knows at most α" (for if he knew more, than he would not consider possible all the worlds where α is true). Thus, $O\alpha$ can be viewed as saying "the agent knows (or believes) exactly α", or "the agent knows only α".

In the case of a single agent, we can identify worlds with truth assignments. A structure consists of a pair (W, w), where W is a set of worlds (i.e., truth assignments), intuitively, the set of worlds that the agent considers possible, and w is a world, intuitively representing the "real" world. As usual, $L\varphi$ is true if φ is true at all worlds that the agent considers possible, i.e., all the worlds in W. To capture the intuition that $N\alpha$ is true if α is true at all the worlds that the agent does not consider possible, Levesque defines $N\alpha$ as follows:

$$(W, w) \models N\alpha \text{ if } (W, w') \models \alpha \text{ for all } w' \notin W.$$

Two important features of this definition are worth mentioning here. First, the set of all worlds is absolute, and does not depend on the structure: it is the set of all truth assignments. Thus, the set of impossible worlds given that W is the set of worlds that the agent considers possible is just the complement of W (relative to the set of all truth assignments). Second, when evaluating the truth of α at an "impossible world" w', we do not change W, the set of worlds that the agent considers possible. Intuitively, this means that the agent's beliefs are held constant, although we may vary what is true in the real world.

It is also perhaps worth observing that, at least in some circumstances, $O\alpha$ holds even if it might not seem appropriate to say "the agent believes only α", at least, according to the standard English usage of "believes only". For example, it is not hard to show that $O(Lp \Rightarrow p) \wedge Lp$ is satisfiable. But in a situation where this is satisfied, the agent believes only $Lp \Rightarrow p$, yet also believes p. This is not a problem with the definition *per se*; it is a hint, though, that interpreting O as "believes only" might not always be appropriate.

As we mentioned above, Levesque shows how we can understand Moore's AE logic in terms of only knowing. Notice that we can associate with every stable set T a unique set of worlds W_T such that $\alpha \in T$ iff $W_T \models \alpha$ (i.e., $W_T, w \models \alpha$ for all $w \in W_T$). In fact, it is easy to see that W_T consists of all worlds w such that, for every propositional formula α, if $\alpha \in T$ then $w \models \alpha$. Levesque shows that T is a stable expansion of α iff $W_T \models O\alpha$. We can actually strengthen Levesque's result slightly:

Theorem 4.1: *Suppose T is a stable set. Then*

(a) T is sound with respect to α iff $W_T \models N\neg\alpha$.

(b) α is in T iff $W_T \models L\alpha$.

Theorem 4.1 formalizes the relationship between Levesque's notion of "only knowing" and Moore AE logic.

Levesque's approach suggests a way of understanding the role of the set A in AE logic that does not involve A being the formulas told to the agent by some trusted source. Rather, the set A is thought of as characterizing an agent's beliefs. With this viewpoint, we can understand the fact that Lp has no stable expansions as saying that an agent can never believe only Lp. The fact that p and $p \wedge Lp$ have the same stable expansion simply means that an agent that believes only p also can be viewed as believing only $p \wedge Lp$.

While this is a reasonable interpretation, it leaves open the question of how an agent comes to believe "only A". If it is not the result of being told something by a trusted source, then how does it happen? Levesque is, in fact, interested in knowledge bases that are told things via a TELL operation. Unfortunately, it is not the case in general that the stable expansion(s) of A corresponds to what an agent believes if it is told precisely A, although this is the case if A consists of nonmodal formulas.

We now turn to the approach considered in [Halpern and Moses 1984]. Halpern and Moses were trying to capture the notion of what it means for an agent to say "all I know is α". A number of alternative definitions of "all I know" are given, which are proved equivalent. We discuss two of them here.

One definition is in terms of stable sets: If all an agent knows is α, then his knowledge is characterized by that (consistent) stable set that contains α, and whose kernel is minimum among those stable sets that contain α. If there is no minimum consistent stable set containing α, then α is said to be *dishonest*: it does not make sense to say that all an agent knows is A. An example of a dishonest formula is $Lp \vee Lq$. Intuitively, an introspective agent that claims to know only $Lp \vee Lq$ is being dishonest, for the agent must know which of Lp or Lq he knows.

An alternative definition is in terms of Kripke structures: What an agent knows if all he knows is α is the set of formulas β such that $W \models \beta$, where W is the unique set of worlds such that (a) $W \models \alpha$ and (b) if $W' \models \alpha$, then $W' \subseteq W$. If no such set W exists, then α is said to be dishonest. Notice that, unlike Levesque's definition, in the HM definition the agent's beliefs are not kept constant. We consider situations where the set of possible worlds varies, in an effort to find the maximum set of possible worlds, which corresponds to minimizing the agent's beliefs.

Although the phrase "all I know" is used both by Levesque and by Halpern and Moses, there are clearly some significant differences between the two notions. In [Halpern and Moses 1984], the difference was attributed to the fact that Moore's notion (and hence Levesque's notion) is based on a K45 notion of belief, whereas the HM notion is based on S5. This is not the true cause of the difference: As we observe in Appendix D, in the case of one agent, we get the same notion of stable set whether we consider S5 or K45 (modulo the inconsistent set).[7] The HM definitions apply *with no change* whether we consider K45 or S5. (Indeed, notice that the definition above mentioned neither K45 nor S5.) The major difference is that in Levesque's definition, the set of worlds that the agent

[7]This is no longer true once we have many agents in the picture; see [Halpern 1994].

considers possible is kept fixed; in the HM definition, it is not. As a consequence, the HM definition comes perhaps a little closer to capturing the natural-language semantics of "all I know". For example, a formula such as $Lp \Rightarrow p$ is honest under the HM approach: An agent's knowledge if all she knows in $Lp \Rightarrow p$ is characterized by the stable set T_1; we can ignore the "anomalous" set T_2. The formula Lp is honest as well; it makes sense to say "all I know is that I know p" (this is the same as saying "all I know is p"). By way of contrast, $O(Lp)$ is inconsistent.

Interestingly, the distinction between the two approaches can be captured in terms of the operators Γ_2^Δ and Γ_3^Δ. Indeed, the following result is almost immediate from Theorems 3.2 and 4.1.

Theorem 4.2: *Suppose that Δ is a default theory and α is the conjunction of the formulas in $tr_L(\Delta)$. Then*

(a) if $W \models O\alpha$, then there is a fixed point A of Γ_2^Δ such that for each non-modal formula β, we have $W \models \beta$ iff $\beta \in A$,

(b) α is honest iff Γ_3^Δ has a unique minimal fixed point. If Γ_3^Δ has a unique minimal fixed point, then β is in the minimal fixed point if an agent who knows only α also knows β.

The HM approach is not without its problems. For one thing, as pointed out in [Schwarz and Truszczyński 1992], this approach is not conservative with respect to introducing explicit definitions. Suppose all an agent knows is α, and then the agent defines a new proposition q (one not appearing in α) to be equivalent to some formula φ. We might hope that adding the formula $q \Leftrightarrow \varphi$, which takes φ to be the definition of q, does not affect what the agent knows about formulas not involving q. Such a property certainly holds for standard monotonic modal logics. However, suppose α is the formula *true*, and let α' be the definitional formula $q \Leftrightarrow Lp$. Clearly, α is honest, but α' is not (since α' has two minimal models, one where $Lq \wedge Lp$ holds and another where $L\neg q$ holds).

Another problem with this approach is that it does not capture standard monotonic reasoning very well. For example, we might hope that if all we know is $p \Rightarrow Lp$, then we could conclude $\neg p$: if all we know is $p \Rightarrow Lp$, then we don't know p (i.e., Lp does not hold), so we can conclude $\neg p$. On the other hand, the formula $p \Rightarrow Lp$ is a dishonest formula in the HM approach. It has two minimal models: one where $p \wedge Lp$ holds, and one where $\neg p \wedge L\neg p$ holds.

Of course, the HM approach was never meant to capture nonmonotonic reasoning of this sort. Rather, like Levesque's logic, it was intended to characterize a certain state of knowledge, one where it was appropriate to say "all I know is α". The arguments in [Halpern and Moses 1984] are intended to support the conclusion that the definition of "all I know" does indeed capture our intuitions regarding this notion. Does it capture our intuitions better than Levesque's definition? Notice that the two definitions agree on nonmodal formulas, so we must look at modal formulas to distinguish them. We would argue that the HM definition is more appropriate on formulas like $Lp \Rightarrow p$; on the other hand, it has the unpleasant property of not being conservative with respect to definitions, a failing from which Levesque's approach does not suffer. Again, it seems that we need to fine-tune our intuitions regarding "all I know", perhaps by studying how we come to know all we know.

5 Conclusions

We have critically reanalyzed some of the basic definitions in default logic and AE logic. Each of the approaches has some rather counterintuitive properties. An obvious question to ask is which approach is best? It is doubtful that there is a unique answer. The situation feels much like that in philosophy in the mid 60's, when there was a great deal of discussion of what was the "right" notion of knowledge. As the work on distributed systems [Halpern 1987] and situated automata [Rosenschein and Kaelbling 1986] shows, the S5 definition, long rejected by philosophers, is an appropriate one for certain applications (and inappropriate for others).

This suggests that in order to sharpen our intuitions, it might be best to focus on a few concrete application areas, where we have a clear understanding of when it is appropriate to adopt a default rule, or what it means for a formula to be in a set A of initial premises, and what it means for a default conclusion to be appropriate. As Neufeld [1989] has pointed out, surprisingly little work has been done on this. It would be interesting to consider a number of different interpretations, and consider applications where each interpretation is appropriate. We should not be surprised to discover that different approaches to nonmonotonic reasoning, and different logics, are appropriate for different interpretations of defaults and different applications.

A Proofs of Results in Section 2

We repeat the results here for convenience.

Proposition 2.1: *All the sets in* $fp(\Gamma_i^\Delta)$, $i = 1, \ldots, 4$ *are consistent with* Δ.

Proof: We prove the result for Γ_1^Δ; the proofs in the other cases are essentially the same and left to the reader. Suppose $A \in fp(\Gamma_1^\Delta)$. Thus, $A = \Gamma_1^\Delta(A)$. It is immediate from D1 and D2 that A contains W and is deductively closed, so it satisfies C1 and C2. Moreover, from D3 it follows that if $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in \Gamma_1^\Delta(A)$, and $\neg\beta_1, \ldots, \neg\beta_m \notin A$, then $\gamma \in \Gamma_1^\Delta(A)$. But since $A = \Gamma_1^\Delta(A)$, it is immediate that A satisfies C3. Thus, A is consistent. ∎

Theorem 2.5: *For all default theories* Δ *we have* *(a)* $fp(\Gamma_1^\Delta) \subseteq fp(\Gamma_2^\Delta) \subseteq fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta)$,

(b) $fp(\Gamma_3^\Delta)$ *consists of all theories consistent with* Δ.

Proof: For part (a), first suppose $A \in fp(\Gamma_1^\Delta)$. We want to show that $A \in fp(\Gamma_2^\Delta)$. Thus, we must show that $\Gamma_2^\Delta(A) = A$. Since $\Gamma_1^\Delta(A) = A$ by assumption, D1, D2, and D3 must hold if we take $S = S' = A$. But D3$'$ is identical to D3 if we take $S = S' = A$. Thus, it suffices to show that no strict subset $A' \subset A$ satisfies D1, D2, and D3$'$ with $S = A$ and $S' = A'$. But if one does, it easy to see that we can satisfy D1, D2, and D3 with $S = A$ and $S' = A'$, contradicting the assumption that $\Gamma_1^\Delta(A) = A$. It follows that $\Gamma_2^\Delta(A) = A$, so $A \in fp(\Gamma_2^\Delta)$. Hence $fp(\Gamma_1^\Delta) \subseteq fp(\Gamma_2^\Delta)$.

Next, suppose that $A \in fp(\Gamma_2^\Delta)$. We want to show that $A \in fp(\Gamma_3^\Delta)$. Since $\Gamma_2^\Delta(A) = A$ by assumption, it must be the case that D1, D2, and D3$'$ hold with $S = S' = A$. Clearly D0 also holds with $S = S' = A$. Since D3 and D3$'$ are identical when $S = S' = A$, D3 also holds under these assumptions. Thus, it remains to show that no strict subset $A' \subset A$ satisfies D0, D1, D2, and D3 with $S = A$ and $S' = A'$. But since D0 implies $A \subseteq A'$, we get a contradiction in this case. Thus, $\Gamma_3^\Delta(A) = A$, so $A \in fp(\Gamma_3^\Delta)$. It follows that $fp(\Gamma_2^\Delta) \subseteq fp(\Gamma_3^\Delta)$.

Next, suppose that $A \in fp(\Gamma_3^\Delta)$. We want to show that $A \in fp(\Gamma_4^\Delta)$. Since $\Gamma_3^\Delta(A) = A$ by assumption, it must be the case that D0, D1, D2, and D3 hold with $S = S' = A$. Since D3 and D3$'$ are identical when $S = S' = A$, D3$'$ also holds under these assumptions. Thus, it remains to show that no strict subset $A' \subset A$ satisfies D0, D1, D2, and D3 with $S = A$ and $S' = A'$. But again D0 gives us a contradiction if we assume $A' \subset A$. Thus, $\Gamma_4^\Delta(A) = A$, so $A \in fp(\Gamma_4^\Delta)$. It follows that $fp(\Gamma_3^\Delta) \subseteq fp(\Gamma_4^\Delta)$. The argument that $fp(\Gamma_4^\Delta) \subseteq fp(\Gamma_3^\Delta)$ is identical, except that we reverse the roles of D3 and D3$'$. Thus, $fp(\Gamma_3^\Delta) = fp(\Gamma_4^\Delta)$.

For part (b), we know by Proposition 2.1 that every theory in $fp(\Gamma_3^\Delta)$ is consistent with Δ. Thus, it remains to show that every theory consistent with Δ is in $fp(\Gamma_3^\Delta)$. So suppose that E is a theory consistent with Δ. Since E satisfies C1–C3, it follows that D0–D3 hold with $S = S' = E$. It easily follows that $\Gamma_3^\Delta(E) = E$, so E is in $fp(\Gamma_3^\Delta)$, as desired. ∎

Theorem 2.6: *For all default theories* Δ*, we have*

(a) $mfp(\Gamma_1^\Delta) = fp(\Gamma_1^\Delta)$,

(b) $mfp(\Gamma_1^\Delta) \subseteq mfp(\Gamma_2^\Delta)$,

(c) $mfp(\Gamma_1^\Delta) \subseteq mfp(\Gamma_3^\Delta) = mfp(\Gamma_4^\Delta)$.

Proof: As we observed in the main text, part (a) of Theorem 2.6 already follows from Theorem 2.4 in [Reiter 1980].

For part (b), suppose $A \in \mathit{mfp}(\Gamma_1^\Delta)$; we want to show that $A \in \mathit{mfp}(\Gamma_2^\Delta)$. From Theorem 2.5, we know that $A \in \mathit{fp}(\Gamma_2^\Delta)$. Suppose there is some set $A' \subseteq A$ such that $A' \in \mathit{fp}(\Gamma_2^\Delta)$. Since $\Gamma_2^\Delta(A') = A'$, it is immediate that D1 and D2 hold if we take $S = A$ and $S' = A'$. D3 also holds, for if $\frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in A'$, and $\neg\beta_1, \ldots, \neg\beta_m \notin A$, then surely $\neg\beta_1, \ldots, \neg\beta_m \notin A'$. Since D3$'$ holds with $S = S' = A'$, it follows that $\gamma \in A'$. Thus, D3 holds with $S = A$ and $S' = A'$. It follows that $A = \Gamma_1^\Delta(A) \subseteq A'$. It follows that $A' = A$, and so $A \in \mathit{mfp}(\Gamma_2^\Delta)$.

For part (c), suppose $A \in \mathit{mfp}(\Gamma_1^\Delta)$; we want to show that $A \in \mathit{mfp}(\Gamma_3^\Delta)$. From Theorem 2.5, we know that $A \in \mathit{fp}(\Gamma_3^\Delta)$. Suppose there is some set $A' \subseteq A$ such that $A' \in \mathit{fp}(\Gamma_3^\Delta)$. Identical arguments to those used in the proof of part (b) again show that we have $\Gamma_1^\Delta(A) \subseteq A'$, so that we must have $A = A'$. Hence $A \in \mathit{mfp}(\Gamma_3^\Delta)$. The fact that $\mathit{mfp}(\Gamma_3^\Delta) = \mathit{mfp}(\Gamma_4^\Delta)$ is immediate from the fact that $\mathit{fp}(\Gamma_3^\Delta) = \mathit{fp}(\Gamma_4^\Delta)$, which was shown in Theorem 2.5. ∎

B Proof-Theoretic Characterizations of Default Logic

In this appendix, we discuss the proof-theoretic characterization of default logic given in [Marek and Truszczyński 1989; Marek and Truszczyński 1993], and its relation to the operators Γ_i^Δ. Following [Marek and Truszczyński 1989; Marek and Truszczyński 1993] (with some minor changes in notation), if D is a set of defaults and S is a set of formulas, we say that there is an *n-step strong proof of φ from D and E given S* if there is a sequence $\varphi_1, \ldots, \varphi_n$ such that $\varphi_n = \varphi$, and for all i with $1 \leq i \leq n$, one of the following holds:

1. φ_i is in S,
2. φ_i is a propositional tautology,
3. there exist $j, k < i$ such that φ_k is of the form $\varphi_j \Rightarrow \varphi_i$ (so that φ_i follows from φ_j and φ_k using *modus ponens*),
4. for some $j < i$, there is a default $\frac{\varphi_j : M\beta_1, \ldots, M\beta_m}{\varphi_i} \in D$ such that $\neg\beta_k \notin E$, $k = 1, \ldots m$.

We say that there is a strong proof of φ from D and E given S if there is an n-step proof for some n.

We define an operator $Cn^{D,E}$ on sets of formulas by taking $Cn^{D,E}(S)$ to be the set of formulas with a strong proof from D and E, given S. It is easy to see that $Cn^{D,E}$ is monotone in S (if $S \subseteq S'$, then $Cn^{D,E}(S) \subseteq Cn^{D,E}(S')$). Define $Cn_0^{D,E}(S) = S$ and $Cn_{n+1}^{D,E}(S) = Cn^{D,E}(Cn_n^{D,E}(S))$. Finally, define $Cn_\infty^{D,E}(S) = \cup_{n=1}^\infty Cn_n^{D,E}(S)$. Marek and Truszczyński observe that the following theorem follows from results of [Reiter 1980].

Theorem B.1: *If $\Delta = (W, D)$, then $\Gamma_1^\Delta(E) = \mathrm{Cn}_\infty^{D,E}(W)$.*

This result provides a proof-theoretic characterization of Γ_1^Δ. Marek and Truszczyński also define a notion of *weak proof of φ from D and E given S*. The definition is identical to that of strong proof, except that clause 4 is replaced by:

4′. there is a default $\frac{\alpha:M\beta_1,\ldots,M\beta_m}{\varphi_i} \in D$ such that $\alpha \in E$ and $\neg\beta_k \notin E$, $k = 1, \ldots m$.

Thus, the prerequisite of the default must now be in E, rather than something that was proved earlier. This change is somewhat reminiscent of the change in going from Γ_1^Δ to Γ_2^Δ. As the following result shows, this is not just coincidental. In analogy to $Cn^{D,E}$, we define an operator $WCn^{D,E}$ on sets of formulas by taking $WCn^{D,E}(S)$ to be the set of formulas with a weak proof from D and E, given S. We can also provide an analogous definition of $WCn_\infty^{D,E}(S)$.

Theorem B.2: *If $\Delta = (W, D)$, then $\Gamma_2^\Delta(E) = \mathrm{WCn}_\infty^{D,E}(W)$.*

Proof: We first show that $\Gamma_2^\Delta(E) \subseteq WCn_\infty^{D,E}(W)$. Since, by definition, $\Gamma_2^\Delta(E)$ is the least set satisfying D1, D2, and D3′, it suffices to show that $WCn_\infty^{D,E}(W)$ satisfies these three properties. Clearly $W \subseteq WCn_1^{D,E}(W)$, so $W \subseteq WCn_\infty^{D,E}(W)$; thus, D1 is satisfied. Clauses 2 and 3 of the definition of strong (and weak) proof guarantee that $WCn_\infty^{D,E}(W)$ is deductively closed, so D2 is satisfied. Finally, we must show that $WCn_\infty^{D,E}(W)$ is closed under defaults in the sense defined by D3′. But clearly if $\frac{\alpha:M\beta_1,\ldots,M\beta_m}{\gamma} \in D$, $\alpha \in E$, and $\neg\beta_1, \ldots, \neg\beta_m \notin E$, then clause 4′ of the definition of weak proof guarantees that $\gamma \in WCn_\infty^{D,E}(W)$, as desired. It now follows that $\Gamma_2^\Delta(E) \subseteq WCn_\infty^{D,E}(W)$.

To show that $WCn_\infty^{D,E}(W) \subseteq \Gamma_2^\Delta(E)$, we show by induction on n that $WCn_n^{D,E}(W) \subseteq \Gamma_2^\Delta(E)$. The case that $n = 0$ is easy, since by definition $W \subseteq \Gamma_2^\Delta(E)$. For the inductive case, assuming that $WCn_n^{D,E}(W) \subseteq \Gamma_2^\Delta(E)$, we must show that $WCn_{n+1}^{D,E}(W) \subseteq \Gamma_2^\Delta(E)$. To do this, it suffices to show by a subinduction on k that if there is a k-step weak proof of φ from D and E given $WCn_n^{D,E}(W)$, then $\varphi \in \Gamma_2^\Delta(E)$. This follows easily from the properties D1, D2, and D3′ of $\Gamma_2^\Delta(E)$; we leave details to the reader. ∎

We can also provide a similar proof-theoretic treatment of Γ_3^Δ. Define an *ultraweak proof of φ from D and E given S* to be the same as a strong proof, except that we add one more possibility:

0. φ_i is in E.

We define an operator $UCn^{D,E}$ on sets of formulas by taking $UCn^{D,E}(S)$ to be the set of formulas with an ultraweak proof from D and E, given S. We can also provide an analogous definition of $UCn_\infty^{D,E}(S)$.

Theorem B.3: *If $\Delta = (W, D)$, then $\Gamma_3^\Delta(E) = \mathrm{UCn}_\infty^{D,E}(W)$.*

Proof: The proof is similar to that of Theorem B.2; we leave the straightforward details to the reader. ∎

Notice that the difference between the three notions of proof is the extent to which they use the formulas in E. In a strong proof, it is only used to check if the justifications in a default rules are "justified". In a weak proof, it is used to check both the prerequisite and the justifications in a default rule. Finally, in an ultraweak proof, it is used to check the prerequisite and the justifications in a default rule, and it is used to justify steps in a proof. Thus, we can view strong proofs as more "grounded" than weak proofs, which are in turn more "grounded" than ultraweak proofs. Such a notion of "groundedness" or "justification" deserves further study; perhaps an appropriate semantics for default rules might help to clarify its role.

C Proof of Theorem 3.2

Theorem 3.2: *Let Δ be a default theory.*

(a) A set of formulas is a fixed point of Γ_3^Δ iff it is the kernel of a stable set containing $tr_L(\Delta)$.

(b) A set of formulas is a fixed point of Γ_2^Δ iff it is the kernel of a stable expansion of $tr_L(\Delta)$.

Proof: Suppose $\Delta = (W, D)$. For part (a), suppose that A is a fixed point of Γ_3^Δ. By Theorem 2.1, A is consistent with Δ. Consider the unique stable set S_A with kernel A. It easily follows from Moore's construction that S_A has two properties of relevance to us: if α is a propositional formula, then (1) $L\alpha \in S_A$ iff $\alpha \in A$ and (2) $\neg L\alpha \in S_A$ iff $\alpha \notin A$. Suppose that $d = \frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$. We claim that $tr_L(d) \in S_A$. There are two cases to consider: First, suppose that $\gamma \in A$. In this case, since S_A is deductively closed, it contains all formulas of the form $\varphi \Rightarrow \gamma$, and, in particular, $tr_L(d)$. On the other hand, if $\gamma \notin A$, since A is consistent with Δ, it follows from property C3 that either $\alpha \notin A$ or $\neg\beta_j \in A$ for some $j = 1, \ldots, m$. Thus, either $\neg L\alpha \in A$ or $L\neg\beta_j \in A$. In either case, it follows from the fact that S_A is deductively closed that $tr_L(d) = (L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m \Rightarrow \gamma) \in S_A$. Clearly S_A contains W, since A does. Thus, it follows that A is the kernel of a stable set containing $tr_L(\Delta)$, namely S_A.

Conversely, suppose that A is the kernel of a stable set S containing $tr_L(\Delta)$. To show that A is the fixed point of Γ_3^Δ, it suffices to show that A is consistent with Δ. Since S contains W, so does A, and since S is deductively closed, so is A. Thus, A satisfies C1 and C2. To see that A satisfies C3, suppose that $d = \frac{\alpha : M\beta_1, \ldots, M\beta_m}{\gamma} \in D$, $\alpha \in A$, and $\neg\beta_1, \ldots, \neg\beta_m \notin A$. Since S is the unique stable set with kernel A, it follows by Moore's construction again that $L\alpha, M\beta_1, \ldots, M\beta_m \in S$. Since $tr_L(d) \in S$, it follows that $\gamma \in S$, and hence $\gamma \in A$. Thus, A satisfies C3. This means that A is consistent with Δ, and hence is a fixed point of Γ_3^Δ.

For part (b), suppose A is a fixed point of Γ_2^Δ. Consider the unique stable set S_A with kernel A. We want to show that S_A is a stable expansion of $tr_L(\Delta)$. The argument that $tr_L(\Delta)$ is in S_A is identical to that in part (a). Thus, it remains to show that S_A is sound with respect to $tr_L(\Delta)$. So suppose that I is an AE interpretation of S_A in which all the formulas in $tr_L(\Delta)$ are true. We want to show that I is an AE model of S_A. Thus, we must show that all the formulas in S_A are true under interpretation I. It clearly suffices to show that all the formulas in A are true. Since A is a fixed point of Γ_2^Δ, we have that $A = \Gamma_2^\Delta(A)$. By Theorem B.2, it follows that $A = WCn_\infty^{D,A}(W)$. Since $WCn_\infty^{D,A}(W) = \cup_{n=1}^\infty WCn_n^{D,A}(W)$, it suffices to prove by induction on n that all the formulas in $WCn_n^{D,A}(W)$, $n = 0, 1, 2, \ldots$, are true under interpretation I. Clearly the formulas in $W = WCn_0^{D,A}(w)$ are true in A, since these formulas form a subset of $tr_L(\Delta)$. Suppose that every formula in $WCn_n^{D,A}(W)$ is true under I. We want to show that every formula in $WCn_{n+1}^{D,A}(W)$ is true under I. Thus, we must show that if there is a strong proof of φ from E and A given $WCn_n^{D,A}(W)$, then φ is true under I. Suppose that there is such a strong proof of φ, that is, a sequence $\varphi_1, \ldots, \varphi_m$ such that $\varphi_m = \varphi$ and for all $i \leq m$, one of clauses 1, 2, 3, or $4'$ in the definition of strong proof holds. A straightforward induction argument shows that each φ_i is true under I. In particular, notice that if clause $4'$ holds for φ_i, then there is some default $\frac{\alpha:M\beta_1,\ldots,M\beta_m}{\varphi_i} \in D$ such that $\alpha \in A$ and $\neg\beta_1, \ldots, \neg\beta_m \notin A$. By hypothesis, the formula $tr_L(d) = L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m \Rightarrow \varphi_i$ is true. Moreover, since I is an AE interpretation of S_A, it follows that $L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m$ is true under I. It follows that φ_i is true under I. Thus, every formula in $WCn_{n+1}^{D,A}(W)$ is true under I, and hence every formula in A is true under I.

For the converse, suppose that A is the kernel of a stable expansion S_A of $tr_L(\Delta)$. We want to show that $\Gamma_2^\Delta(A) = A$. Clearly A contains W and is deductively closed, so it satisfies D1 and D2. To see that it satisfies D3$'$, suppose $d = \frac{\alpha:M\beta_1,\ldots,M\beta_m}{\gamma} \in D$, $\alpha \in A$, and $\neg\beta_1, \ldots, \neg\beta_m \notin A$. Since $\alpha \in A$ and $\neg\beta_1, \ldots, \neg\beta_m \notin A$, it follows that $L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m \in S_A$. By assumption, $tr_L(d) \in S_A$. It follows that $\gamma \in S_A$, and hence that $\gamma \in A$. Thus, A satisfies D3$'$. This means that $\Gamma_2^\Delta(A) \subseteq A$.

To see that $\Gamma_2^\Delta(A) = A$, consider the AE interpretation I of S_A in which the only propositional formulas that are true are those in $\Gamma_2^\Delta(A)$. We claim that I makes all the formulas in $tr_L(\Delta)$ true. It clearly makes all the formulas in W true, since $W \subseteq \Gamma_2^\Delta(A)$. Suppose $d = \frac{\alpha:M\beta_1,\ldots,M\beta_m}{\gamma} \in D$. Clearly if $L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m$ is not true under I, then $tr_L(d)$ is true under I. On the other hand, if $L\alpha \wedge M\beta_1 \wedge \ldots \wedge M\beta_m$ is true under I, then α must be in A, and $\neg\beta_1, \ldots, \neg\beta_m$ cannot be in A, since I is an AE interpretation of S_A, and A is the kernel of S_A. Since $\Gamma_2^\Delta(A)$ satisfies D3$'$, it follows that $\gamma \in \Gamma_2^\Delta(A)$, and hence that γ is true under I. Again, it follows that $tr_L(d)$ is true. Since S_A is a stable expansion of $tr_L(D)$, every AE interpretation of S_A which makes the formulas in $tr_L(D)$ true must make all the formulas in S_A and, in particular, all the formulas in A, true. It follows that $\Gamma_2^\Delta(A) = A$, as desired. Thus, A is a fixed point of Γ_2^Δ. ∎

D A Closer Look at Stable Expansions

We start with a review of a few of the basic notions of modal logic that we shall need. We refer the reader to one of the standard texts on modal logic, such as [Chellas 1980] or [Hughes and Cresswell 1968], for further details.

The modal logic K45 is characterized by the following axiom system, consisting of four axioms:

P. All instances of axioms of propositional logic

K. $(L\varphi \wedge L(\varphi \Rightarrow \psi)) \Rightarrow L\psi$

4. $L\varphi \Rightarrow LL\varphi$

5. $\neg L\varphi \Rightarrow L\neg L\varphi$

and two rules of inference:

R1. From φ and $\varphi \Rightarrow \psi$ infer ψ

R2. From φ infer $L\varphi$.

We get the modal logic KD45 by adding the axiom

D. $\neg L\mathit{false}$,

and the modal logic S5 by adding, instead, the stronger axiom

T. $L\varphi \Rightarrow \varphi$.

Numerous other modal logics can be constructed by considering other combinations of axioms.

The axioms **4** and **5** are called the *positive introspection axiom* and *negative introspection axiom*, respectively. They are appropriate for agents that are sufficiently introspective so that they know what they know and do not know.

We give semantics to K45, KD45, and S5 by considering Kripke structures. For these logics, we can view a Kripke structure as a pair (W, w), where W is a set of truth assignments that, intuitively, characterizes the worlds the agent considers possible, and w is a truth assignment that, intuitively, characterizes the "real world". We call such a pair a *K45 situation*. A *KD45 situation* is a K45 situation (W, w) such that $W \neq \emptyset$. An *S5 situation* (W, w) is a K45 situation such that $w \in W$. We give semantics to formulas with respect to situations. If p is a primitive proposition, then $(W, w) \models p$ if p is true under truth assignment w. Conjunctions and negations are dealt with in the standard way. For formulas of the form $L\alpha$, we define

$$(W, w) \models L\alpha \text{ iff } (W, w') \models \alpha \text{ for all } w' \in W.$$

It is well known [Hughes and Cresswell 1968] that a formula is provable in K45 (resp. KD45, S5) if and only if it is true in all K45 (resp. KD45, S5) situations.

The following result, which is implicit in the results of [Konolige 1988], establishes the connection between K45 and the notion of stable sets.

Lemma D.1: *T is stable if and only if $T = \{\alpha : (W, w) \models L\alpha\}$ for some K45 situation (W, w).*

We remark that we can replace the K45 above by S5 or KD45, if we assume that T is not the inconsistent set consisting of all formulas. Essentially, this says that (modulo the inconsistent set) the set of beliefs of an S5 agent are indistinguishable from those of a K45 (or a KD45) agent.

This lemma also suggests how we can define a notion of stable set for agents that do not have complete introspective powers. Suppose $\mathcal{S}$ is a (modal) logic, and we have a notion of $\mathcal{S}$ situation, as we do for K45, KD45, or S5. (This is well known to be the case for all the standard modal logics.) Then we can define a set T to be *$\mathcal{S}$-stable* if and only if $T = \{\varphi : (W, w) \models L\alpha\}$ for some $\mathcal{S}$-situation (W, w). As shown in [Halpern 1993; Halpern 1994], the notion of $\mathcal{S}$-stable is quite useful when dealing with modal logics with many agents. We say that T is *$\mathcal{S}$-sound with respect to A* if, for every $\mathcal{S}$-situation (W, w) such that $(W, w) \models L\alpha$ iff $\alpha \in T$ and $(W, w) \models A$ (where we take a set of formulas to be satisfied at (W, w) exactly if each formula in the set is), we have that $(W, w) \models T$. It is easy to see that a stable set T is sound with respect to A in Moore's sense exactly if it is K45-sound. Finally, we define an *$\mathcal{S}$-stable expansion of A* to be an $\mathcal{S}$-stable set containing A that is $\mathcal{S}$-sound with respect to A. This generalizes Moore's definitions to arbitrary modal logics (at least, to ones that can be captured by a class of Kripke structures).

Motivated by the original work of McDermott [1982], Marek, Schwarz and Truszczyński (for example, in [Marek, Shvarts, and Truszczyński 1991; Shvarts 1990; Schwarz and Truszczyński 1992]) consider an alternate notion of stable expansion. Given a modal logic $\mathcal{S}$, they define T to be an *$\mathcal{S}$-expansion* of A if

$$T = \{\varphi : A \cup \{\neg L\psi : \psi \notin T\} \vdash_{\mathcal{S}} \varphi\}.$$

They assume that $\mathcal{S}$ includes the necessitation rule, so that $\psi \vdash_{\mathcal{S}} L\psi$. Using this assumption, it is not hard to show that if T is an $\mathcal{S}$-expansion of A, then T is a (K45-)stable set. By varying the modal logic $\mathcal{S}$, we obtain different notions of expansion, with different properties. By taking $\mathcal{S}$ to be K45, we get back Moore's original notion of stable expansion. On the other hand, as shown in [Shvarts 1990], if we take $\mathcal{S}$ to be S4, then $\{Lp\}$ has a unique S4-expansion. Moreover, $\{Lp \Rightarrow p\}$ only has one S4-expansion, the set T_1 defined earlier. The "anomalous" expansion T_2 in which Lp holds is not an S4-expansion.

The machinery developed in Marek, Schwarz, and Truszczyński has some very nice technical properties. But what is not clear is the underlying philosophical motivation of this approach. What does it mean that an agent uses, say, S4 to reason about his knowledge, and thus, does not necessarily know what he does not know, yet his final belief set is stable, and thus has the negative introspection property? This question needs to be answered before we can comfortably apply the technical machinery.

E Proofs of Theorems 4.1 and 4.2

Theorem 4.1: *Suppose T is a stable set. Then*

(a) T is sound with respect to α iff $W_T \models N\neg\alpha$.

(b) α is in T iff $W_T \models L\alpha$.

Proof: For part (a), suppose T is sound with respect to α. To show that $W_T \models N\neg\alpha$, we must show that if $w \notin W_T$, then $W_T, w \models \neg\alpha$ or, equivalently, if $W_T, w \models \alpha$ then $w \in W_T$. Suppose that $W_T, w \models \alpha$. Let I be an AE interpretation of T such that the only propositional formulas that are true under I are the propositional formulas satisfied by w. It is easy to show, by induction on the structure of formulas, that for every formula β we have $W_T, w \models \beta$ iff β is true under interpretation I. For propositional formulas this is true by construction of I. The only nontrivial step in this argument comes if β is of the form $L\beta'$. In this case, $W_T, w \models L\beta'$ iff $\beta' \in T$ (by construction of W_T) iff $L\beta'$ is true under I (since I is an AE interpretation of T). In particular, it follows that I makes α true. Since T is sound with respect to α, it must be the case that I makes all the formulas T true. Hence, $W_T, w \models \beta$ for every formula, and in particular, every propositional formula, in T. It follows that $w \in W_T$, as desired.

Conversely, suppose $W_T \models N\neg\alpha$. We want to show that T is sound with respect to α. So suppose that I is an AE interpretation of T which makes α true. We must show that I makes every formula in T true. Let w be a truth assignment which agrees with I. As above, we can show that for all formulas β, we have $W_T, w \models \beta$ iff I makes β true. Since I makes α true, we must have that $W_T, w \models \alpha$. Since $W_T \models N\neg\alpha$, we must have that $w \in W_T$. We also know that $W_T, w \models L\beta$ for every $\beta \in T$ (since I makes all these formulas true). Since $w \in W_T$, it follows that $W_T, w \models \beta$ for every $\beta \in T$. Thus, every formula in T is true in situation (W_T, w). Hence, I makes every formula in T true, as desired.

For part (b), note that $W_T \models L\alpha$ iff $L\alpha \in T$, which is true iff $\alpha \in T$ (since T is stable). ∎

Theorem 4.2: *Suppose that Δ is a default theory and α is the conjunction of the formulas in $tr_L(\Delta)$. Then*

(a) if $W \models O\alpha$, then there is a fixed point A of Γ_2^Δ such that for each non-modal formula β, we have $W \models \beta$ iff $\beta \in A$,

(b) *α is honest iff Γ_3^Δ has a unique minimal fixed point. If Γ_3^Δ has a unique minimal fixed point, then β is in the minimal fixed point if an agent who knows only α also knows β.*

Proof: For part (a), suppose that $W \models O\alpha$. Let T be the stable set consisting of all formulas β such that $W \models \beta$. From Theorem 4.1, it follows that T is a stable expansion of $\alpha = tr_L(\Delta)$. By Theorem 3.2, it follows that the kernel of T is a fixed point of Γ_2^Δ. Part (a) now follows immediately.

For part (b), α is honest iff there is a stable set containing α whose kernel is minimal among those stable sets containing α. By Theorem 3.2, this is true iff Γ_3^Δ has a unique minimal fixed point. Moreover, if Γ_3^Δ has a unique minimal fixed point, it follows immediately from Theorem 3.2 and the HM definition of "all I know" that this minimal fixed point is the kernel of the stable set whose formulas are the formulas that are known if the agent knows only α. Thus, if β is in the miminal fixed point, then an agent who knows only α also knows β. ∎

Acknowledgments

I would like to thank Adam Grove, Daphne Koller, Kurt Konolige, Hector Levesque, Wiktor Marek, Bob Moore, Ray Reiter, Grisha Schwarz, Bob Stalnaker, and Mirek Truszczyński for discussions that inspired me to write this paper, and to ARPA for funding the work on the ARPAnet that made these discussions possible. Grisha and Moshe Vardi made a number of useful comments on an earlier draft of this paper.

References

Brewka, G. (1991). Cumulative default logic: in defense of nonmonotonic inference rules. *Artificial Intelligence 50*, 183–205.

Chellas, B. F. (1980). *Modal Logic.* Cambridge, U.K.: Cambridge University Press.

Delgrande, J. P. and W. K. Jackson (1991). Default logic revisited. In J. A. Allen, R. Fikes, and E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Second International Conference (KR '91)*, pp. 118–127. San Francisco, Calif.: Morgan Kaufmann.

Halpern, J. Y. (1987). Using reasoning about knowledge to analyze distributed systems. In J. F. Traub, B. J. Grosz, B. W. Lampson, and N. J. Nilsson (Eds.), *Annual Review of Computer Science, Vol. 2*, pp. 37–68. Palo Alto, Calif.: Annual Reviews Inc.

Halpern, J. Y. (1993). Reasoning about only knowing with many agents. In *Proc. National Conference on Artificial Intelligence (AAAI '93)*, pp. 655–661.

Halpern, J. Y. (1994). A theory of knowledge and ignorance for many agents. Research Report RJ 9894, IBM. To appear, *Journal of Logic and Computation*.

Halpern, J. Y. and Y. Moses (1984). Towards a theory of knowledge and ignorance. In *Proc. AAAI Workshop on Non-monotonic Logic*, pp. 125–143. Reprinted in K. R. Apt (Ed.), *Logics and Models of Concurrent Systems*, Springer-Verlag, Berlin/New York, pp. 459–476, 1985.

Hughes, G. E. and M. J. Cresswell (1968). *An Introduction to Modal Logic*. London: Methuen.

Konolige, K. (1988). On the relation between default and autoepistemic logic. *Artificial Intelligence 35*, 343–382.

Konolige, K. (1989). Errata to "On the relation between default and autoepistemic logic". *Artificial Intelligence 41*, 115.

Konolige, K. (1992). Ideal introspective belief. In *Proc. National Conference on Artificial Intelligence (AAAI '92)*, pp. 635–641.

Levesque, H. J. (1990). All I know: a study in autoepistemic logic. *Artificial Intelligence 42*(3), 263–309.

Łukaszewicz, W. (1988). Considerations on default logic: an alternative approach. *Computational Intelligence 4*, 1–16.

Marek, W., G. F. Shvarts, and M. Truszczyński (1991). Modal nonmonotonic logics: ranges, characterization, computation. In J. A. Allen, R. Fikes, and E. Sandewall (Eds.), *Principles of Knowledge Representation and Reasoning: Proc. Second International Conference (KR '91)*, pp. 395–404. San Francisco, Calif.: Morgan Kaufmann.

Marek, W. and M. Truszczyński (1989). Relating autoepistemic and default logics. In R. J. Brachman, H. J. Levesque, and R. Reiter (Eds.), *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pp. 276–288. San Francisco, Calif.: Morgan Kaufmann.

Marek, W. and M. Truszczyński (1993). *Nonmonotonic Logic*. Berlin/New York: Springer-Verlag.

McDermott, D. and J. Doyle (1980). Non-monotonic logic I. *Artificial Intelligence 13*(1,2), 41–72.

McDermott, D. V. (1982). Nonmonotonic logic II: nonmonotonic model theories. *Journal of the ACM 29*(1), 33–57.

Mikitiuk, A. and M. Truszczyński (1995). Constrained and rational default logics. In *Proc. Fourteenth International Joint Conference on Artificial Intelligence (IJCAI '95)*, pp. 1509–1515.

Moore, R. C. (1985). Semantical considerations on nonmonotonic logic. *Artificial Intelligence 25*, 75–94.

Neufeld, E. (1989). Defaults and probabilities; extensions and coherence. In R. J. Brachman, H. J. Levesque, and R. Reiter (Eds.), *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pp. 312–323. San Francisco: Morgan Kaufmann.

Poole, D. (1989). What the lottery paradox tells us about default reasoning. In R. J. Brachman, H. J. Levesque, and R. Reiter (Eds.), *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pp. 333–340. San Francisco: Morgan Kaufmann.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence 13*, 81–132.

Rosenschein, S. J. and L. P. Kaelbling (1986). The synthesis of digital machines with provable epistemic properties. In J. Y. Halpern (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. 1986 Conference*, pp. 83–97. San Francisco, Calif.: Morgan Kaufmann.

Rounds, W. and G. Q. Zhang (1995). Domain theory meets default logic. *Journal of Logic and Computation 5*(1), 1–26.

Schaub, T. (1992). *Considerations on Default Logics*. Ph. D. thesis, Technische Hochschule Darmstadt.

Schwarz, G. F. and M. Truszczyński (1992). Modal logic S4F and the minimal knowledge paradigm. In Y. O. Moses (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Fourth Conference*, pp. 395–404. San Francisco, Calif.: Morgan Kaufmann.

Shvarts, G. F. (1990). Autoepistemic modal logics. In R. Parikh (Ed.), *Theoretical Aspects of Reasoning about Knowledge: Proc. Third Conference*, pp. 97–109. San Francisco, Calif.: Morgan Kaufmann.

Sombé, L. (1990). *Reasoning Under Incomplete Information in Artificial Intelligence.* New York: John Wiley & Sons.

Stalnaker, R. (1993). A note on nonmonotonic modal logic. *Artificial Intelligence 64*, 183–196. This is a slightly revised version of a note that was originally circulated in 1980.