Computer Science Department Technical Report
University of California
Los Angeles, CA 90024-1596

CONDITIONAL ENTAILMENT: BRIDGING TWO APPROACHES TO DEFAULT REASONING

Hector Geffner
Judea Pearl

July 1991
CSD-910027

TECHNICAL REPORT
R-157

Conditional Entailment: Bridging two Approaches to Default Reasoning

Hector Geffner
hector@ibm.com
T. J. Watson Research Center
P. O. Box 704, Room H1-K10
Yorktown Heights, N.Y. 10598

Judea Pearl
judea@cs.ucla.edu
Computer Science Department
Boelter Hall 4731, UCLA
Los Angeles, CA 90024

1 Introduction

The main task of a formal characterization of default inference is to account for the set of conclusions that a given set of facts and defaults legitimizes. This requires a language for expressing facts and defaults, and a specification of how expressions in such a language are to be interpreted. Classical logic is not well suited for the task because of its failure to accommodate the non-monotonic behavior of default inference, where conclusions may be retracted in light of new information. This limitation of classical logic has led to the development of several non-monotonic logics [McCarthy, 1980, McDermott and Doyle, 1980, Reiter, 1980, Moore, 1985] in which defaults are treated as rules for *extending* a set of beliefs in the absence of conflicting evidence. When defaults are in conflict, however, multiple extensions often arise which do not always reflect equally plausible scenarios [Reiter and Criscuolo, 1983, Hanks and McDermott, 1987]. In such cases, the 'undesired' scenarios need to be pruned by such devices as cancellation axioms or priorities [McCarthy, 1986], or by a proper reformulation of the problem (e.g. [Lifschitz, 1987]).

More recently, it has been noted that some of the spurious interactions among conflicting defaults can be eliminated by treating defaults as conditional assertions [Delgrande, 1987, Kraus *et al.*, 1989, Geffner and Pearl, 1990]. Whereas extensional interpretations regard the default "if p then normally q" as a soft reason to believe q given the truth of p, conditional interpretations regard it as a hard but context dependent constraint by which q is true (or highly probable) in the context determined by p and possibly some background knowledge. As a result, conditional interpretations properly 'dissolve' certain spurious conflicts among defaults, such as those arising from "if a then c" and "if a and b then $\neg c$". In a context where a and b are known to be true, the second default is permitted to constraint the truth of $\neg c$, while the first one is rendered inapplicable, leaving the truth of c unconstrained.

In spite of these virtues, however, conditional interpretations fail to account for a number of desirable inferences which are captured by extensional formalisms. These limitations have to do with the way *irrelevant* information is handled. For instance, given a default "if p then q" both extensional and conditional interpretations conclude q given the evidence p; conditional interpretations, however, are unable to maintain that conclusion when an additional but irrelevant piece of evidence e is taken into account. The reason is that conditional interpretations treat all evidence as *relevant* unless otherwise proven, and hence refrain from maintaining q in the presence of e. Indeeed, while extensional interpretations generate conflicts out of braveness, conditional interpretations eliminate conflict out of sheer hesitancy.

The question arises whether a unifying framework can be developed which combines the virtues of both the extensional and conditional interpretations. An earlier attempt in this direction was a proposal to enhance a conditional interpretation based on probabilities with a syntactic criterion for distinguishing relevant from irrelevant evidence [Geffner, 1988, Geffner and Pearl, 1990] (see also [Delgrande, 1987]). However, while the results, for the most part, were satisfactory, the theoretical underpinnings were not. A more promising approach, based on model-theoretic considerations has recently been advanced by Lehmann [1989] and Pearl [1990]. These proposals, however, while better motivated than earlier ones, turned out less successful: some useful inferences fail to be captured, while anomalous ones are introduced.[1]

In this paper we develop an alternative model-theoretic interpretation of defaults, called ***conditional entailment,*** which finally closes the gap between extensional and conditional interpretations and which exhibits the best features of both. Conditional entailment is closely related to prioritized circumscription except that priorities among defaults are not provided by the user but are automatically extracted from the knowledge base. Conditional

[1]See section 4.

entailment thus shows that the difference between the conditional (probabilistic or model-theoretic) and extensional interpretations can be reduced to a particular *ordering* on defaults; an idea previously suggested in [Pearl, 1990].

The paper is organized as follows. First (Section 2) we briefly review two conditional interpretations of defaults, one probabilistic and one model-theoretic. Then we define the semantics and proof-theory of conditional entailment (Section 3), and discuss related proposals (Section 4). Finally (Section 5) we address the computation of conditional entailment and the limitations of conditional entailment as an account of default reasoning.

2 Conditional Interpretations of Defaults

This section provides a brief survey of recent work on conditional interpretations of defaults.[2] Conditional interpretations assume that default theories can be structured into two components: a *background context* K containing generic information about the domain of interest, and an *evidence set* E containing information specific to the particular situation at hand. Intuitively, K contains the relevant *rules*, while E contains observational *facts* (see [Geffner and Pearl, 1990]).

The background context K is assumed to comprise a sentential component L and a default component D. Defaults are denoted by expressions of the form $p \rightarrow q$, where p and q denote sentences referred as the default antecedent and consequent respectively. The expression $\mathtt{dog(fido)} \rightarrow \mathtt{can\_bark(fido)}$, for instance, represents a default stating that "normally, if Fido is a dog, Fido can bark." We use *default schemas* of the form $p(x) \rightarrow q(x)$, where p and q are wffs with free variables among those of x, to denote the collection of defaults $p(a) \rightarrow q(a)$ that results from substituting x by all tuples a of ground terms in the language.

[2]See [Geffner, 1989] for details and proofs.

In the probabilistic interpretation of defaults [Adams, 1975, Pearl, 1988, Geffner, 1989], the background K is viewed as imposing a constraint over probability distributions which are later conditioned on the evidence E to provide the degree of belief on arbitrary sentences. We call these distributions *admissible* and define them as follows:

Definition 1 *A probability distribution P_K is* ϵ-admissible *relative to a background $K = \langle L, D \rangle$ when P_K assigns unit probability to every (strict) sentence s in L, i.e. $P_K(s) = 1$, and probabilities $P_K(q \mid p) > 1-\epsilon$ and $P_K(p) > 0$ to each default $p \rightarrow q$ in D.*

In other words, a probability distribution is ϵ-admissible when it renders the sentences in L *certain*, while leaving a range ϵ of uncertainty for the defaults in D. If the conditional probability of a proposition p given a body of evidence E approaches one as ϵ approaches zero, then we say that p is ϵ-entailed by E:

Definition 2 *A proposition p is ϵ-entailed by a default theory $T = \langle K, E \rangle$ when for any $\epsilon' > 0$, there exists an $\epsilon > 0$, such that $P_K(p \mid E) > 1 - \epsilon'$ for any ϵ-admissible probability distribution P_K.*

The entailment relation so defined is non-monotonic relative to E and captures several of the essential aspects of defaults [Pearl, 1988, Geffner, 1989]. Effective procedures for testing ϵ-entailment are reported in [Goldszmidt and Pearl, 1989].

An alternative conditional interpretation of defaults appeals to models instead of probabilities. The role of probability distributions is filled by *preferential model structures* [Lewis, 1973, Shoham, 1988, Kraus *et al.*, 1989, Makinson, 1989]:

Definition 3 *A* preferential model structure *is a pair $\langle \mathcal{I}, < \rangle$, where $\mathcal{I}$ denotes a non-empty collection of interpretations, and '$<$' denotes an irreflexive and transitive order relation over $\mathcal{I}$.*

For a particular structure $\langle \mathcal{I}, < \rangle$ and two interpretations M and M' in $\mathcal{I}$, the notation $M < M'$ is read as saying that M is *preferred* to M'. Moreover, when M is a model of T (i.e. a model of L and E) and $\mathcal{I}$ contains no model of T preferred to M, then M is also said to be a *preferred model* of T in that structure.

In the same way in which the probabilistic interpretation only considers *admissible* probability distributions, so does the model-theoretic interpretation only considers *admissible* preferential model structures, viewing defaults as constraints over the preference relation '$<$':

Definition 4 *A well-founded*[3] *preferential model structure* $\langle \mathcal{I}, < \rangle$ *is* admissible *relative to a background* $K = \langle L, D \rangle$ *iff every interpretation in* $\mathcal{I}$ *satisfies* L, *and for every default* $p \rightarrow q$ *in* D, *(a)* $\mathcal{I}$ *contains a model of* p, *and (b)* q *is true in all preferred models of* p.

Preferential entailment is defined analogously to ϵ-entailment:

Definition 5 *A default theory* $T = \langle K, E \rangle$ preferentially entails (p-entails) *a proposition* p *iff* p *is true in all the preferred models of* E *of every preferential model structure admissible with* K.

Thus, while in ϵ-entailment the background K defines the admissible probability distributions which are then conditioned upon the evidence E, in p-entailment, the background K defines the admissible preferential model structures from which the preferred models of E are selected. Interestingly, as suggested early by Adams [1975, 1978], and noted recently by Lehmann and Magidor [1988], ϵ-entailment and p-entailment coincide for finite propositional languages, and they both accept an elegant and simple proof-theory:

Theorem 1 *For finite propositional languages, if* $T = \langle K, E \rangle$ *is a default theory, the following conditions are equivalent: 1)* T ϵ*-entails* q, *2)* T *p-entails* q, *3) the expression* $E \mathrel{|\!\!\approx_K} q$ *is derivable from the rules:*

[3]A structure $\langle \mathcal{I}, < \rangle$ is well-founded if for every non-preferred interpretation M in $\mathcal{I}$, $\mathcal{I}$ contains a preferred interpretation M' such that $M' < M$.

Rule 1 (Conditionals) $p \mathrel{|\!\!\approx_K} q$ *if* $p \rightarrow q \in D$

Rule 2 (Deduction) *If* $E, L \vdash p$ *then* $E \mathrel{|\!\!\approx_K} p$

Rule 3 (Augmentation) *If* $E \mathrel{|\!\!\approx_K} p$ *and* $E \mathrel{|\!\!\approx_K} q$ *then* $E, p \mathrel{|\!\!\approx_K} q$

Rule 4 (Reduction) *If* $E \mathrel{|\!\!\approx_K} p$ *and* $E, p \mathrel{|\!\!\approx_K} q$ *then* $E \mathrel{|\!\!\approx_K} q$

Rule 5 (Disjunction) *If* $E, p \mathrel{|\!\!\approx_K} r$ *and* $E, q \mathrel{|\!\!\approx_K} r$ *then* $E, p \vee q \mathrel{|\!\!\approx_K} r$

Pearl [1989] has referred to rules 1–5 as a default reasoning *core*, suggesting that they constitute a basic set of principles that any reasonable account of defaults is bound to obey. As discussed in the introduction, while the core captures certain patterns of inference that escape traditional non-monotonic logics, it also misses patterns which the latter do capture. The extension of the core that we will pursue in this paper provides the benefits of the two approaches and originates from the following simple observation.

Let us say that we encode defaults $p \rightarrow q$ as 'abnormality' sentences of the form $p \wedge \neg \mathtt{ab}_\mathtt{i} \Rightarrow q$ with unique abnormality predicates $\mathtt{ab}_\mathtt{i}$. Namely, we take a default theory $T = \langle K, E \rangle$ with a background $K = \langle L, D \rangle$ and map it into a new theory $T' = \langle K', E \rangle$ with a background $K' = \langle L', D' \rangle$, where D' is empty and L' comprises the formulas in L as well as the abnormality sentences corresponding to the defaults in D. Furthermore, let us say that we define the consequences of T in terms of the models of T' which are *minimal* in the set of abnormalities they sanction, as advocated in [McCarthy, 1986].

Notice first, that by restricting attention to models which are minimal we will be automatically capturing the 'independencies' that characterize extensional non-monotonic formalisms. Namely, given a default "birds fly" we would not only be able to conclude that a bird flies, but also that a *red* bird flies. On the other hand, as with other extensional formalisms, the minimal model semantics will be missing patterns such as specificity preferences, that are captured by both ϵ-entailment and p-entailment. Given that for finite propositional languages ϵ-entailment and p-entailment can be completely characterized by rules 1–5, it is natural to ask then which rules among 1–5 the minimal model semantics fails to capture.

It turns out that under reasonable assumptions, rule 1, the conditional rule (called defaults rule elsewhere), is the only rule that the minimal model semantics fails to capture. In other words, the minimization of 'abnormality' renders a semantics that complies with all the rules sanctioned by ϵ-entailment and p-entailment, with the exception of rule 1. This observation suggests that a promising way of enhancing a conditional interpretation of defaults with 'extensional' features would be to twist the minimal model semantics in such a way that rule 1 becomes valid. That is indeed what we are about to do in Section 3. The mechanism for validating rule 1 will be to impose certain *priorities* among abnormalities as a function of the defaults in K. By considering only minimal models which sanction abnormalities of lower priority, we will get an entailment relation that is stronger than both p-entailment and minimal models, a relation we call *conditional entailment.*

3 Conditional Entailment

3.1 Preliminary Definitions

Conditional entailment deals with default theories $T = \langle K, E \rangle$ in which each default in K is associated a unique ***assumption*** of 'normality.' Arbitrary default theories $T' = \langle K', E \rangle$ can be expressed in this assumption-based format by replacing each default schema $p(x) \rightarrow q(x)$ in K' by a sentence $p(x) \wedge \delta_i(x) \Rightarrow q(x)$ and a default schema $p(x) \rightarrow \delta_i(x)$, where δ_i denotes a new and unique assumption predicate, which summarizes the normality conditions required for concluding $q(x)$ in the context of $p(x)$. This encoding 'trick' is similar to McCarthy's [1986] 'abnormality' formulation and Poole's [1988] default naming conventions. Our choice for assumptions as primitive objects is just a matter of convenience. We call the theories in the resulting format ***assumption-based default theories.***

We will represent assumptions by literals of the form $\delta_i(a)$, where a is a tuple of ground terms, and will use the symbols δ, δ', ..., as variables ranging over the assumptions in the underlying language $\mathcal{L}$. Similarly, we use the notation $\Delta_{\mathcal{L}}$ to refer to the collection of all such assumptions.

Given a default theory $T = \langle K, E\rangle$ with background $K = \langle L, D\rangle$, we further identify the models of T as the interpretations that satisfy the sentences in both L and E. We will also refer to the set of assumptions violated by an interpretation I as the *gap* of I, and denote it as $\Delta[I]$. Similarly, we use the notation $E \vdash_K p$ to express that a proposition p is a *deductive* consequence of T, i.e. $E, L \vdash p$. Finally, we will refer to a set of assumptions Δ logically consistent with T as an *argument,* and say that Δ is an argument *for* a proposition p if $E, \Delta \vdash_K p$, and an argument *against* p if $E, \Delta \vdash_K \neg p$.

3.2 Model Theory

Conditional entailment is a specialization of the notion of preferential entailment discussed in section 2 for the case in which the *preference order on interpretations* is determined by a given *priority ordering on assumptions.* We call the resulting structures *prioritized preferential structures* and define them as follows:

Definition 6 *A* prioritized preferential structure *is a quadruple $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec\rangle$, where $\mathcal{I}_{\mathcal{L}}$ stands for the set of interpretations over the underlying language $\mathcal{L}$, $\Delta_{\mathcal{L}}$ stands for the set of assumptions in $\mathcal{L}$, '$\prec$' stands for an irreflexive and transitive priority relation over $\Delta_{\mathcal{L}}$, and '$<$' is a binary relation over $\mathcal{I}_{\mathcal{L}}$, such that for two interpretations M and M', $M < M'$ holds iff $\Delta[M] \neq \Delta[M']$ and for every assumption δ in $\Delta[M] - \Delta[M']$ there exists an assumption δ' in $\Delta[M'] - \Delta[M]$ such that $\delta \prec \delta'$.*

Figure 1 illustrates the preference on two interpretations M and M' determined by an arbitrary priority ordering on assumptions depicted by means

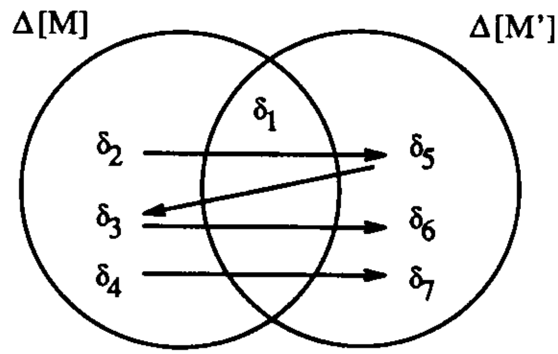

Figure 1: Preference on interpretations in prioritized structures: $M < M'$

of arrows. An arrow connecting an assumption δ_i to an assumption δ_j expresses that δ_i has lower priority than δ_j, i.e. $\delta_i \prec \delta_j$. To check then whether M is preferred to M', it is sufficient to check that each assumption δ in $\Delta[M] - \Delta[M']$ is linked by an arrow to an assumption δ' in $\Delta[M'] - \Delta[M]$ (and that $\Delta[M] \neq \Delta[M']$). Note that the assumptions violated by both M and M' (e.g., δ_1) play no role in determining the preferences between M and M'.

We assume that priority orderings do not contain infinite ascending chains $\delta_1 \prec \delta_2 \prec \delta_3 \prec \cdots$. Prioritized structures are called *preferential* because the priority ordering '$\prec$' on assumptions determines a strict partial order '$<$' on interpretations.

Lemma 1 *If the quadruple $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$ is a prioritized preferential structure, then the pair $\langle \mathcal{I}_{\mathcal{L}}, < \rangle$ is a preferential model structure.*

Such partial order on interpretations regards the relation $\delta \prec \delta'$ as a preference to sustain the assumption δ' over the assumption δ in cases of conflict. A similar mapping from *predicate* priorities to preferences occurs in Przymusinski's characterization of the perfect model semantics of general logic programs [Przymusinski, 1987] and in McCarthy's prioritized circumscription [McCarthy, 1986, Lifschitz, 1985]. Moreover, like in these frameworks, the induced order on interpretations establishes a preference which favors models violating a minimal set of assumptions:

Lemma 2 *In any prioritized preferential structure, if M is a preferred model of a theory T, then M is* minimal *in $\Delta[M]$, i.e. there is no model M' of T such that $\Delta[M'] \subset \Delta[M]$.*

While the minimality of preferred models will endow conditional entailment with the features common to traditional non-monotonic logics, the focus on a particular class of priority orderings which reflect the structure of K will account for the desirable features of conditional interpretations.

Let us say that a set of assumptions Δ is *in conflict* with a default $p \rightarrow \delta$ in K, when Δ constitutes an argument against δ in the context $\langle K, \{p\}\rangle$, i.e. when $p, \Delta \vdash_K \neg\delta$ and $p \not\vdash_K \neg\Delta$. Our basic premise is that the user who provided the default $p \rightarrow \delta$ truly means to assert δ when p represents all the evidence. Hence, it is natural to assume that s/he implicitly regards the violation of one of the assumptions in Δ as less important than the violation of δ. *Admissible priority orderings* capture this intuition. More precisely, we shall say that an assumption δ *dominates* a set of assumptions Δ whenever Δ or a subset of it is in conflict with a default $p \rightarrow \delta$ in K, and accordingly, we define *admissible* priority orderings as follows:

Definition 7 *A priority order '$\prec$' over $\Delta_{\mathcal{L}}$ is* admissible *with a background context K iff every set Δ of assumptions dominated by an assumption δ contains an assumption δ' such that $\delta' \prec \delta$.*

Admissible prioritized preferential structures are the structures induced from admissible priority orderings:

Definition 8 *A prioritized preferential structure $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec\rangle$ is* admissible *with a background $K = \langle L, D\rangle$ iff the priority ordering '$\prec$' is admissible with K.*

Finally, *conditional entailment* is defined in terms of the preferred models of the *admissible* prioritized structures:

Definition 9 *A proposition q is* conditionally entailed *by a default theory $T = \langle K, E \rangle$, iff q holds in all the preferred models of T of every prioritized preferential structure admissible with K.*

Conditional entailment combines the two target notions: minimality and conditionality. Indeed, for *finite propositional languages* the following result can be proven:

Theorem 2 *If an assumption-based default theory T preferentially entails (ϵ-entails) a proposition p, then T also conditionally entails p.*

Furthermore, if we say that a theory $T = \langle K, E \rangle$ is *p-consistent* when there is a preferential structure admissible with its background K, *conditionally consistent* when there is a prioritized structure admissible with K, and *ϵ-consistent* when for any positive ϵ there is a probability distribution ϵ-admissible with K, we can also show that conditional entailment remains well-behaved as long as p-entailment and ϵ-entailment are well-behaved:[4]

Theorem 3 *If an assumption-based default theory $T = \langle K, E \rangle$ is p-consistent (ϵ-consistent), then T is also conditionally consistent.*

Next, we will illustrate the behavior of conditional entailment by means of a simple example. We will find it convenient to write $\Delta \prec \delta$ as an abbreviation of "there exists a δ' in Δ such that $\delta' \prec \delta$." Thus, the admissibility of the priority order '$\prec$' with respect to K amounts to testing whether the relation $\Delta \prec \delta$ holds for all pairs Δ, δ such that δ dominates Δ. Clearly this test needs be applied only to *minimal* Δ's.

[4]For the proof of this theorem, see [Geffner, 1989].

Example 1 Consider a background context $K = \langle L, D \rangle$ with sentences:[5]

$$\mathtt{b}(x) \wedge \delta_1(x) \Rightarrow \mathtt{f}(x)$$
$$\mathtt{p}(x) \wedge \delta_2(x) \Rightarrow \neg\mathtt{f}(x)$$
$$\mathtt{p}(x) \Rightarrow \mathtt{b}(x)$$
$$\mathtt{r}(x) \Rightarrow \mathtt{b}(x)$$

and defaults $\mathtt{b}(x) \rightarrow \delta_1(x)$ and $\mathtt{p}(x) \rightarrow \delta_2(x)$. We can read the symbols b, f, p, and r, as standing for the predicates 'bird,' 'fly,' 'penguin,' and 'red-bird,' respectively.

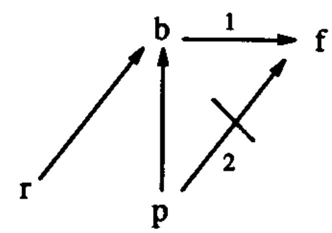

Figure 2: Strict Specificity

A priority ordering '$\prec$' will be admissible with K if the relation $\Delta \prec \delta_i(a)$ holds for any minimal assumption set Δ dominated by an assumption $\delta_i(a)$, for $i = 1, 2$ and any term a in the language. First note that there is no assumption set Δ in conflict with instances of the default schema $\mathtt{b}(x) \rightarrow \delta_1(x)$, since for any ground term a there are interpretations which satisfy K and $\mathtt{b}(a)$ and which violate no assumption. As a result, assumptions of the form $\delta_1(a)$ do not dominate any assumption set and thus impose no constraint on the admissible priority orderings. Assumptions of the form $\delta_2(a)$, on the other hand, dominate a single minimal set of assumptions $\Delta = \{\delta_1(a)\}$, since for any ground term a, $\mathtt{p}(a), \Delta \models_K \neg\delta_2(a)$ holds, while $\mathtt{p}(a) \models_K \neg\Delta$ does not. As a result, a priority order '$\prec$' will be admissible with K iff the relation

[5] Formulas are implicitly universally quantified, so we usually write $p(x) \Rightarrow b(x)$ instead of $\forall x.\, p(x) \Rightarrow b(x)$.

$\delta_1(a) \prec \delta_2(a)$ is satisfied for every ground term a in the language. We also write in these cases $\delta_1(x) \prec \delta_2(x)$.

Provided with this characterization of the prioritized structures admissible with K, we can now turn to analyze the propositions which are conditionally entailed in the different contexts of interest. For example, for an individual Tim (t), the preferred models of b(t) in K are the models which violate no assumption. As a result, both assumptions $\delta_1(\mathtt{t})$ and $\delta_2(\mathtt{t})$ are conditionally entailed by b(t), as are the propositions f(t) and ¬p(t) (i.e., Tim is presumed to be a normal flying bird, and therefore, not a penguin).

A different scenario arises when we consider the evidence p(t) instead of b(t). In this case, every interpretation satisfying the evidence and the background context is forced to render one of assumptions $\delta_1(\mathtt{t})$ or $\delta_2(\mathtt{t})$ false. Thus, two classes of minimal models arise: a class $\mathcal{C}_{\{1\}}$ comprised of the models M_1 which violate the assumption $\delta_1(\mathtt{t})$, and a class $\mathcal{C}_{\{2\}}$ comprised of the models M_2 which violate the assumption $\delta_2(\mathtt{t})$. However, since $\delta_1(\mathtt{t}) \prec \delta_2(\mathtt{t})$, models in the former class are preferred to models in the latter class, because (see definition 6) $\Delta[M_2] - \Delta[M_1] = \{\delta_2(\mathtt{t})\}$, $\Delta[M_1] - \Delta[M_2] = \{\delta_1(\mathtt{t})\}$, and so $M_1 < M_2$. It follows then, that $\mathcal{C}_{\{1\}}$ represents the class of preferred models of p(t) in K, and therefore, that the propositions $\delta_2(\mathtt{t})$ and ¬f(t) are conditionally entailed. Similar conclusions are indeed legitimized by preferential entailment and ϵ-entailment.

Finally, consider the scenario in which the target context is enhanced with the information that Tim is also a red bird, i.e. $T' = \langle K, E' \rangle$, with $E' = \{\mathtt{p(t)}, \mathtt{r(t)}\}$. In this case, neither ϵ-entailment nor p-entailment constrain the preferred models of T'. Conditional entailment, on the other hand, guarantees that the preferred models of T' are minimal, and therefore, that they belong to one of the two classes $\mathcal{C}_{\{1\}}$ and $\mathcal{C}_{\{2\}}$ of minimal models, where $\mathcal{C}_I$ stands for the collection of models M of T' with a gap $\Delta[M] = \{\delta_i(\mathtt{t}) \,|\, i \in I\}$. However, as we showed above, models in $\mathcal{C}_{\{1\}}$ are preferred to models in $\mathcal{C}_{\{2\}}$.[6] As a result, the assumption $\delta_2(\texttt{t})$ and the proposition $\neg\texttt{f}(\texttt{t})$ are conditionally entailed by T'. Note that, on the other hand, neither proposition is legitimized by either ϵ-entailment or p-entailment, nor by minimality considerations alone.

The example above illustrates different contexts built on top of a background which forces every admissible priority ordering '$\prec$' to satisfy the relation $\delta_1(a) \prec \delta_2(a)$, for all ground terms a in the language. This means that every admissible priority *relation* '$\prec$' must include all tuples of the form $\langle\delta_1(a), \delta_2(a)\rangle$. Such relations may include additional tuples as well, e.g. $\langle\delta_1(\mathrm{a}), \delta_2(\mathrm{b})\rangle$, but those tuples are not necessary for the relations to be admissible. We will say that an admissible priority relation is *minimal* when no set of tuples can be deleted without violating the admissibility constraints. For instance, in the example above, there is a *single* minimal admissible ordering which includes all and only the tuples of the form $\langle\delta_1(a), \delta_2(a)\rangle$ for ground atoms a. It is natural to ask then whether conditional entailment can be computed by restricting attention to *minimal* admissible priority orderings only. The answer is yes. Indeed, if we can obtain an admissible priority ordering '$\prec$' by deleting certain tuples from an admissible priority ordering '$\prec'$,' the preferred models in the structure $\langle\mathcal{I}_\mathcal{L}, <', \Delta_\mathcal{L}, \prec'\rangle$ will be a subset of the preferred models of the structure $\langle\mathcal{I}_\mathcal{L}, <, \Delta_\mathcal{L}, \prec\rangle$. Thus, if we say that an admissible prioritized preferential structure $\langle\mathcal{I}_\mathcal{L}, <, \Delta_\mathcal{L}, \prec\rangle$ is *minimal* if the relation '$\prec$' is a minimal admissible priority ordering, the following alternative characterization of conditional entailment results:

Lemma 3 *A proposition q is* conditionally entailed *by a default theory $T = \langle K, E\rangle$ iff q holds in all preferred models of T of every* minimal *prioritized preferential structure admissible with K.*

[6]Note that these classes do not contain the same interpretations as in the context above. Still, since they possess the same gaps, and the preference relation on classes depends exclusively on such gaps, they are ranked in the same way.

In the example above, we can thus compute conditional entailment by considering a *single* structure $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$, where the priority ordering is such that $\delta \prec \delta'$ holds iff $\delta = \delta_1(a)$ and $\delta' = \delta_2(a)$ for some ground term a in the language. Often, however, multiple minimal structures will need to be considered (see [Geffner, 1989]).[7]

Example 2 We now consider a slightly different background context K comprising the sentences:

$$\mathtt{a} \wedge \delta_1 \Rightarrow \mathtt{w}$$

$$\mathtt{u} \wedge \delta_2 \Rightarrow \neg \mathtt{w}$$

$$\mathtt{u} \wedge \delta_3 \Rightarrow \mathtt{a}$$

$$\mathtt{f} \wedge \delta_4 \Rightarrow \mathtt{a}$$

and defaults of the form $p_i \rightarrow \delta_i$ for each sentence $p_i \wedge \delta_i \rightarrow q_i$. The background K can be understood as expressing defaults such as "adults work," "university students do not work," "university students are adults," and "Frank Sinatra fans are adults." It has the same structure as example 1 (fig. 2), except that all rules are now defeasible.

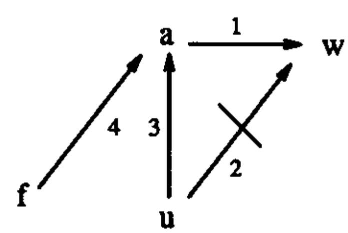

Figure 3: Default Specificity

[7]A background K' requiring multiple minimal structures can be obtained from the above K by replacing the 'strict' link $\mathtt{p}(x) \Rightarrow \mathtt{b}(x)$ by the strict link $p(x) \Rightarrow p'(x)$ and the default $p'(x) \rightarrow b(x)$ (encoded as the sentence $p(x) \wedge \delta_3(x) \Rightarrow p'(x)$ and default $p(x) \rightarrow \delta_3(x)$). A priority ordering '$\prec$' will then be admissible relative to K', if for any ground term a *either* $\delta_2(a) \prec \delta_1(a)$ or $\delta_2(a) \prec \delta_3(a)$ holds.

There are two relevant dominance relations in this background context. First, the assumption δ_2 dominates the set $\Delta = \{\delta_1, \delta_3\}$ as Δ constitutes an argument against the default $u \rightarrow \delta_2$, i.e. $u, \Delta \models_K \neg\delta_2$ and $u \not\models_K \neg\Delta$. Likewise, the assumption δ_3 dominates the set $\{\delta_1, \delta_2\}$. Thus, any priority ordering '$\prec$' admissible with K must be such that both relations $\{\delta_1, \delta_3\} \prec \delta_2$ and $\{\delta_1, \delta_2\} \prec \delta_3$ must hold. Moreover, due to the asymmetric and transitive character of priority orderings, such constraints imply $\delta_1 \prec \delta_2$ and $\delta_1 \prec \delta_3$. To show that this is the case, let us first assume $\delta_2 \prec \delta_3$. Then, by the asymmetry of the priority order we must have $\delta_3 \not\prec \delta_2$, and therefore, from the constraints above, $\delta_1 \prec \delta_2$. Now assume $\delta_2 \not\prec \delta_3$. If $\delta_1 \prec \delta_2$ does not hold, the constraints above imply $\delta_1 \prec \delta_3$ and $\delta_3 \prec \delta_2$ in contradiction with the transitivity of '$\prec$'. Thus, in either case the relation $\delta_1 \prec \delta_2$ must hold. By symmetry, we conclude that $\delta_1 \prec \delta_3$ must hold as well.[8]

With these space of admissible priority orderings, let us first consider a context $T = \langle K, E\rangle$, with $E = \{\mathtt{f}\}$. Since there is an interpretation that satisfies T and every assumption in the language, the single preferred class in every admissible prioritized structure is the class of models which violate no assumption. In particular, the assumptions δ_1 and δ_4 are conditionally entailed by T, as are the propositions a and w. Note that these inferences involve default chaining, a pattern which is not sanctioned by either ϵ-entailment or by p-entailment.

A different situation arises when the proposition u is observed. The context $T' = \langle K, E'\rangle$, with $E' = \{\mathtt{f}, \mathrm{u}\}$, gives rise to three classes of minimal models: $\mathcal{C}_{\{1\}}$, $\mathcal{C}_{\{2\}}$ and $\mathcal{C}_{\{3,4\}}$, where $\mathcal{C}_I$, as usual, stands for the class of models M of T' such that $\Delta[M] = \{\delta_i : i \in I\}$. However, since $\delta_1 \prec \delta_2$ and $\delta_1 \prec \delta_3$, any model M in $\mathcal{C}_{\{1\}}$ is preferred to any model M' in $\mathcal{C}_{\{2\}}$ and any model M'' in $\mathcal{C}_{\{3,4\}}$. Hence $\mathcal{C}_{\{1\}}$ represents the class of preferred models of T', and therefore, all assumptions other than δ_1 are conditionally entailed by T', as

[8]Note that on 'specificity' grounds the priority of δ_3 over δ_1 does not appear justified. However, without it we would not be able to conclude a from u.

are the propositions a and $\neg$w.

3.3 Proof Theory

Conditional entailment provides a characterization of the propositions which are entailed by a given default theory. It does not provide, however, effective methods for computing such propositions. In this section we will focus on such methods. We develop a number of *syntactic* criteria for testing conditional entailment some of which are amenable to implementation in ATMS-type of systems (section 5.1).

Like the proof-theory of classical deduction is structured around the notion of *proofs*, the proof-theory of conditional entailment is structured around the notion of *arguments* [Loui, 1987, Pollock, 1987]. Recalling from section 2, an argument Δ in a theory $T = \langle K, E \rangle$ refers to a set of assumptions logically consistent with T. Moreover, Δ is an argument *for* a proposition p if $E, \Delta \models_K p$, and an argument *against* p if $E, \Delta \models_K \neg p$. In the former case we also say that Δ *supports* p. If Δ is not logically consistent with T, then Δ is called a *conflict set*. *Two arguments are in conflict* when their union is a conflict set. Likewise, an assumption is *free* in T when it does not belong to any *minimal* conflict set in T and is *bound* otherwise. Default theories T are *bound* when they give rise to a *finite* number of bound assumptions. All theories considered so far are bound as are the theories to which the syntactic account below applies.

The first condition for assertability is a simple consequence of the minimality of preferred models within the class of prioritized preferential structures. Hereafter, we will assume a context $T = \langle K, E \rangle$.

Lemma 4 *An assumption is conditionally entailed if there are no arguments against it.*

A similar condition is both sound and complete for circumscriptive theories, provided that assumptions are identified with negative literals and that

T includes the *unique names* and *domain closure* axioms [Gelfond and Przymusinska, 1986]. In the context of conditional entailment, however, such a condition is too weak, since an assumption may face counterarguments and still be conditionally entailed. The "birds fly–penguins don't" example above provides one such case. If Tim is a penguin, the assumption $\delta_2(\mathtt{t})$: 'Tim does not fly, because it is a normal penguin' faces the counterargument $\delta_1(\mathtt{t})$: 'Tim flies, because it is a normal bird.' Nonetheless, $\delta_2(\mathtt{t})$ is conditionally entailed while $\delta_1(\mathtt{t})$ is not.

In order to capture these conclusions by syntactic means, we need to consider the priority orderings induced by K. Indeed, the reason $\delta_2(\mathtt{t})$ survives its conflict with $\delta_1(\mathtt{t})$ is because its priority is higher; that is, $\delta_1(\mathtt{t}) \prec \delta_2(\mathtt{t})$ holds in every admissible priority ordering. The assertability conditions below take these constraints into account. Recall that we write $\Delta \prec \delta$ as an abbreviation of the expression "$\exists \delta' \in \Delta$ such that $\delta' \prec \delta$."

Lemma 5 *An assumption δ is conditionally entailed if for every argument Δ against δ and every admissible priority ordering '$\prec$', the relation $\Delta \prec \delta$ holds.*

Note that it is sufficient to consider only the *minimal* arguments Δ against δ; if $\Delta \prec \delta$ holds, so will $\Delta' \prec \delta$ for any superset Δ' of Δ.

The condition introduced by lemma 5 leads to the correct handling of the example above. Given the evidence $E = \{\mathtt{p}(\mathtt{t})\}$ ('Tim is a penguin'), $\Delta = \{\delta_1(\mathtt{t})\}$ is the only (minimal) argument against $\delta_2(\mathtt{t})$ and since $\delta_2(\mathtt{t})$ has a priority higher than $\delta_1(\mathtt{t})$, lemma 5 permits us to derive $\delta_2(\mathtt{t})$, from which 'Tim does not fly' follows.

However, while lemma 5 refines lemma 4, it is not yet complete. This can be illustrated by converting the *strict* 'links' in Example 1 into *default* 'links', resulting in a structure analogous to that of Example 2. By semantic arguments, we showed then that the assumption δ_4 is conditionally entailed by $T = \langle K, E \rangle$, with $E = \{\mathtt{u}, \mathtt{f}\}$. Yet, the condition in lemma 5 does not authorize this conclusion: $\Delta = \{\delta_1, \delta_2\}$ is an argument against δ_4 for which the relation $\Delta \prec \delta_4$ fails to hold.

Intuitively, the reason δ_4 is conditionally entailed by T in spite of the counterargument Δ, is that Δ contains an assumption δ_1 which is *defeated*. In other words, δ_1 is in conflict with two 'better' assumptions δ_2 and δ_3 which knock the argument Δ out, leaving the assumption δ_4 unchallenged.

This suggests that lemma 5 should be extended by considering multiple conflicts at the same time. For that, some definitions will be handy. We write below $\Delta' \prec \Delta$ as an abbreviation of the expression "for every δ in Δ, $\Delta' \prec \delta$."

Definition 10 *Given a priority ordering '$\prec$', an argument Δ* defeats *an argument Δ' if the two arguments are in conflict and the relation $\Delta' \prec \Delta$ holds. In this case we say that Δ is a* defeater *of Δ'.*

Definition 11 *An argument Δ is* protected *from a conflicting argument Δ' iff for every priority ordering admissible with K, Δ contains a defeater of Δ'.*

Intuitively, when an argument Δ is protected from a conflicting argument Δ' it means that Δ is stronger than Δ', and that from the point of view of Δ, the conflicting argument Δ' can be ignored. When all such conflicting arguments can be so ignored, we say that Δ is *stable:*

Definition 12 *An argument is* stable *iff it is protected from every conflicting argument.*

As suggested above, a stable argument is better than any of its competitors, and propositions supported by stable arguments are conditionally entailed:

Lemma 6 *A proposition is conditionally entailed if it is supported by a stable argument.*

The notion of stable arguments is a very powerful one indeed, and it accounts for most of the natural inferences authorized by conditional entailment. Nonetheless, lemma 6 does not yet provide a complete syntactic account of conditional entailment. For example in a theory T comprised of the sentences $\delta_1 \Rightarrow \neg\delta_3$, $\delta_2 \Rightarrow \neg\delta_3$ and $\neg\delta_1 \vee \neg\delta_2$, where δ_3 has a lower priority than δ_1 and δ_2, the disjunction $\delta_1 \vee \delta_2$ is conditionally entailed even though it is not supported by any stable argument (neither $\Delta_1 = \{\delta_1\}$ or $\Delta_2 = \{\delta_2\}$ are stable as they conflict with each other). To account for such conclusions we will need to consider *disjunctive* arguments as well. We will accommodate such arguments by considering the assertability conditions of disjunctive collection of arguments which we call *covers*. For instance, $C = \{\Delta_1, \Delta_2\}$, with $\Delta_1 = \{\delta_1\}$ and $\Delta_2 = \{\delta_2\}$ will turn out to be a *stable cover*, thus legitimizing the disjunction $\delta_1 \vee \delta_2$ and any proposition supported by it.

We make the notion of *stable covers* precise by refining first the conditions under which an argument is protected:

Definition 13 *An argument* Δ *is* strongly protected *from a conflicting argument* Δ' *if for every subargument* Δ'_i *of* Δ' *in conflict with* Δ *there exists a subargument* Δ_i *of* Δ *in conflict with* Δ' *such that* $\Delta'_i \prec \Delta_i$.

Note that if an argument Δ is protected from a conflicting argument Δ'_1 but is not protected from a conflicting argument Δ'_2, Δ will *not* be *strongly* protected from the union $\Delta'_1 + \Delta'_2$ even though Δ may be protected from it. The distinction between the two notions is irrelevant for stable arguments which are *both* protected and strongly protected from every conflicting argument but is needed for dealing with *disjunctive* arguments.

Let us refer to a collection of arguments as a *cover* — where a cover is to be understood as the disjunction of the arguments it contains — and let us generalize the notions of conflicts and protection as follows:

Definition 14 *An argument* Δ *is* in conflict with a cover *if* Δ *is in conflict with every argument in the cover.*

Definition 15 A cover is protected *from a conflicting argument* Δ *if the cover contains an argument* Δ' *which is* strongly protected *from* Δ.

The conditions under which a *cover* is stable can be then obtained as a generalization of the conditions under which an *argument* is stable. The only difference is that, for the purpose of completeness, we only consider arguments Δ in conflict with C that have as many assumptions from C as possible. We call such conflicting arguments *definite* as they either include or rebut each of the assumptions which occur in C.

Definition 16 *A* cover *is* stable *iff it is protected from every definite conflicting argument* Δ.[9]

As expected, the conditions of lemma 6 can be strengthen by replacing stable arguments by stable covers. Furthermore, if we say that a proposition p is supported by a *cover* when it is supported by every argument in the cover, the following *complete* characterization of conditional entailment results:

Theorem 4 (Main) *A proposition p is conditionally entailed if and only if p is supported by a stable cover.*

We have thus arrived to a complete syntactic characterization of conditional entailment in terms of admissible priority orderings. We can now compute conditional entailment by looking either at models or arguments. Still, an undesirable feature of both approaches is that they presume that we have identified the set of admissible priority orderings, and therefore, that we can check whether relations of the form $\Delta' \prec \Delta$ are necessarily satisfied. This, however, is often a non-trivial task. Fortunately, it is possible to replace such a test by a corresponding *syntactic* test on K.

[9]A consequence of this definition is that the stability of a cover C cannot be computed by considering only the minimal arguments in conflict with C. Rather such arguments have to be extended with as many assumptions from C as possible, what can lead to a proliferation of arguments to evaluate if C is large.

Let us say that a *set* Δ *of assumptions dominates a set* Δ' if every assumption δ in Δ dominates the union $\Delta + \Delta'$. Then, due to the asymmetry and transitivity of priority orderings, the following result can be obtained:

Theorem 5 (Dominance) *For two sets of assumptions* Δ *and* Δ', *the relation* $\Delta' \prec \Delta$ *holds in every priority ordering '*$\prec$*' admissible with a* consistent background $K = \langle L, D \rangle$ *if and only if* Δ *is part of a set* Δ'' *that* dominates Δ' *in* K.[10]

Theorems 4 and 5 together thus permit us to determine whether a given proposition is conditionally entailed by purely syntactic means. For that, we only need to look for stable covers and corresponding dominance relations.

4 Related Work

Conditional entailment is a refinement of an extension of preferential entailment developed independently by Pearl [1990] and Lehmann [1989], targeted at finite propositional default theories. Like in conditional entailment, Pearl and Lehmann rank defaults according to a dominance-like criterion, and use those rankings to infer a preference relation on models. Nonetheless three important differences can be pointed out between Pearl's and Lehmann's proposals on the one hand, and conditional entailment one the other. First, both Pearl and Lehmann deal with integer rankings as opposed to strict partial orders; second, they define the rank of a model as a sole function of the highest ranked default violated by the model; and third, they only consider *one*, in a sense minimal, prioritized preferential structure, as opposed to multiple ones. The consequences of these choices are 1) preferred models are not always minimal; i.e. they do not always violate a minimal set of defaults,

[10]A background K is consistent when there exists a priority ordering admissible with K. For example, a background K with two defaults $p \rightarrow \delta$ and $p \rightarrow \delta'$ such that $p \vdash_K \neg(\delta \wedge \delta')$ is not consistent. See [Geffner, 1989] for details.

and 2) conflicts among defaults that should remain unsolved, sometimes get solved. Two examples illustrate these problems.

Given two defaults $p \rightarrow q$ and $p \rightarrow \neg r$ both accounts fail to authorize the conclusion q given both p and r. The reason is that, in the resulting world ranking, the violation of one default "costs" as much as the violation of many defaults of equal rank. In particular, this implies that exceptional subclasses (e.g., penguins) are unable to inherit properties from their parent superclasses (e.g., birds), which is a major drawback for most practical applications. The second class of problems arises from their commitment to a unique integer ranking on worlds. Consider for example two defaults $p \wedge s \rightarrow q$ and $r \rightarrow \neg q$, which render the status of q ambiguous in the presence of p, s and r. In Pearl's and Lehmann's approaches, this ambiguity is anomalously resolved *in favor* of *q* when a new default $p \rightarrow \neg q$ is added. The reason is that the introduction of $p \rightarrow \neg q$ automatically raises the ranking of the more specific default $p \wedge s \rightarrow q$ which thus becomes preferred to $r \rightarrow \neg q$. The extension of ϵ-semantics based on the principle of maximum-entropy [Goldzmidt *et al.*, 1990], remains committed to a unique integer ranking on worlds and thus inherits similar problems.

Outside the conditional camp, conditional entailment is closest to prioritized circumscription. Prioritized circumscription is a refinement of parallel circumscription, originally proposed by McCarthy [1986], and later developed by Lifschitz [1985, 1988]. Roughly, the effect of prioritized circumscription is to induce a preference for models that assign smaller extensions to predicates of higher priority. The only difference between conditional entailment and prioritized circumscription in the propositional case, is the source of these priorities: while prioritized circumscription relies on the user, conditional entailment extracts the priorities from the knowledge base.

Two other technical differences arise, however, in the first-order case. First, the priorities in prioritized circumscription are on *predicates* as opposed to *literals*.[11] Such a difference often translates into a different behavior. For instance, in the "birds fly, penguins don't" example, the conclusion $\neg$flies(tim) is conditionally entailed by penguin(tim) by virtue of the priority of the assumption δ_2(tim) ('if Tim is a penguin, Tim does not fly') over the assumption δ_1(tim) ('if Tim is a bird, Tim flies'). This behavior can be accommodated in prioritized circumscription by letting the *predicate* $\mathrm{ab}_2 = \neg\delta_2$ have a higher priority than the *predicate* $\mathrm{ab}_1 = \neg\delta_1$. However, such an encoding produces an unintended side effect which does not arise in conditional entailment: given that either Tim is a flying penguin or that Tweety is a non-flying bird, for instance, prioritized circumscription is forced to conclude the latter.

The second technical difference between conditional entailment and prioritized circumscription lies in the notion of *minimality*. In conditional entailment a model M of T is minimal iff it is has a minimal gap $\Delta[M]$; namely, if there is no model M' of T with violates a set of assumptions $\Delta[M']$ properly included in $\Delta[M]$. In particular, since assumptions are *ground literals*, M will be a minimal model of a theory $T = \{\exists x. \neg\delta_1(x)\}$ iff M satisfies every assumption in the language. Every such model will thus presume the existence of one or many *unnamed* individuals which belong to the extension of the predicate δ_1. So while the formula $\delta_1(\mathrm{a})$ will hold in all minimal models of T, the formula $\exists x. \forall y. \neg\delta_1(y) \Rightarrow y = x$ will not. The opposite is true in circumscription, where the minimality of a model is understood *semantically*, rather than *syntactically* (see for example, [Lifschitz, 1985]). In such a case no attention is paid to literals, but to individuals in the domains of the interpretations.

Which notion of minimality is to be preferred? The consensus is overwhelming in favor of the semantic notion, as model-theory enjoys a special status as a tool for specifying the meaning of formal notations. However, if we just ask which minimality criterion is the most convenient for formalizing

[11] Except for the 'pointwise' formulation in [Lifschitz, 1988].

default inference, the syntactic criterion prevails. Such a criterion permits us to reason about equality, enabling us for example to infer that an assumption $\delta_1(\mathtt{b})$ is likely to hold in spite of the violation of an assumption $\delta_1(\mathtt{a})$. For such a pattern to be captured by circumscription, the inequality between a and b need be stated explicitly, precluding the possibility of a and b denoting the same thing. Similarly, given a default "birds fly," conditional entailment, unlike circumscription, is not bound to conclude that *all* birds fly. Indeed, the treatment of equality and universals in conditional entailment is closer to Reiter's [1980] default logic than to circumscription.

There are, however, important limitations that arise from the focus on *ground literals* as opposed to *individuals*. Sometimes, we do want to assume that a property holds about all *individuals*. For instance, when reasoning about time we may want to assume that a clear block will remain clear unless a relevant change occurs. However, a default *schema* such as $\mathtt{on}(x, y, t) \rightarrow \mathtt{on}(x, y, t+1)$ will not do; in particular it will *not* authorize us to infer $\exists x, y\, \mathtt{on}(x, y, t+1)$ from $\exists x, y\, \mathtt{on}(x, y, t)$.[12] In that case, it is clear we do want to minimize the *extension* of predicate $\mathtt{ab_i} = \neg\delta_\mathtt{i}$ associated with the persistence of on. Does this mean that we are to adopt the semantic notion of minimality? Not necessarily. It is possible to retain the benefits of the syntactic criterion and still accommodate forms of *closed world reasoning.*

Let us say that we want a predicate δ_i to be *closed* when we want the *extension* of δ_i to be *maximal* (i.e. the *complement of* δ_i to be *minimal*). Furthermore, let $\delta_i[M]$ stand for the set of tuples of ground terms t in the language such that $\delta_i(t) \in \Delta[M]$. Then, in order to *close a predicate* δ_i, it is sufficient to prune all those models M of the theory T of interest which fail to satisfy the condition $\forall x. \neg\delta_i(x) \Leftrightarrow x \in \delta_i[M]$. If we say that a model of T is a model of the *closure* of T when these closure conditions are satisfied, no empty gap model of a theory $T = \{\exists x. \neg\delta_i(x)\}$, for instance, would remain a model of the closure of T. Similarly, if the preferred models of T are

[12]The example is from [McCarthy, 1980].

selected among the models of the closure of T, a theory $T = \{\neg\delta_i(\mathrm{a})\}$ will certainly conditionally entail the sentence $\forall x . x \neq \mathrm{a} \Rightarrow \delta_i(x)$ very much like the circumscription of the complement of δ_i will. So, it is possible to retain the appealing treatment of equality and universals of Reiter's default logic, and yet accommodate on demand, the form of closed world reasoning characteristic of circumscription.

In light of the relation between the model theory of prioritized circumscription and conditional entailment, it is not surprising to find that their respective proof-theories are related as well. An elegant proof-theory for prioritized circumscription has been recently developed by Baker and Ginsberg [1989]. Baker and Ginsberg address the case in which predicates are linearly ordered; namely, circumscribed predicates are drawn from sets P_1, P_2, ..., P_n such that the priority of a predicate in a set P_i is higher than the priority of a predicate in a set P_j, if $j < i$. While differing in technical detail, the proof-theory they present has the same *dialectical* flavor as the proof-theory developed in section 4.3, which as they note, is closely related to approaches to defeasible inference based on the evaluation of arguments (e.g. [Loui, 1987, Pollock, 1987]). The differences with Baker and Ginsberg are mainly in the treatment of disjunctions, which in our case, are pushed completely into what we called *covers*. Likewise, due to the nature of the constraints on admissible priority orderings, we are forced to consider *collections of non necessarily linear priority orderings*. In this regard, the results in section 4.3 are relevant to prioritized circumscription, as they relax some of the assumptions on which the proof-theory of Baker and Ginsberg is based.

5 Discussion

In the remaining of the paper we briefly address two issues: first, how to compute conditional entailment and, more generally, certain forms of prioritized circumscription, and second, some limitations of conditional entailment as an account of default reasoning.

5.1 Computing Conditional Entailment

One way to compute conditional entailment is by developing a theorem prover along the lines of the proof-theory developed in section 3.3. Needless to say, this promises to be a formidable task. More reasonable, would be to construct a sound but incomplete account which, by capturing most patterns of interest, will be both useful and understandable.

An obvious candidate for approximating conditional entailment is given by the propositions which are supported by stable arguments. From theorem 4 we are guaranteed that such propositions are conditionally entailed, and that incompleteness results only from the exclusion of disjunctive arguments (covers). If we further commit ourselves a single minimal admissible priority ordering, we find that testing whether a given set of assumptions Δ constitutes a stable argument, can be easily accomplished in terms of the minimal conflict sets (nogoods) computed by ATMS-like systems [de Kleer, 1986]. Indeed, if T is a theory and $C_1, \ldots, C_n$ its minimal conflict sets, a set of assumptions Δ will be stable iff for every conflict C_i, $C_i \cap \Delta \neq \emptyset$, there is a conflict C_j such that 1) $(C_j - \Delta) \prec (C_j \cap \Delta)$, and 2) $C_j - \Delta \subseteq C_i - \Delta$. Each of these tests is relatively simple to perform.

Still, a significant gap remains between testing whether a particular set is stable and testing whether a proposition p is supported by a stable set. We can get closer to the latter goal by constructing stable arguments incrementally. Namely, in attempting to prove p we adopt a set of assumptions Δ_0 and try to prove it stable. If it is not stable, we incrementally extend Δ_0 with a set of assumptions Δ_1 that defeats the counterarguments of Δ_0. If we succeed, we still need to show that $\Delta_0 + \Delta_1$ is stable; namely, that Δ_1 does not introduce new undefeated counterarguments of its own. For that we need to repeat the process. This non-deterministic process will either end in success (i.e., finding a stable argument which comprises Δ_0 and thus which supports p) or by finding counterarguments which cannot be defeated. This incremental approach is closer in form to the proof-theory developed by Baker and Ginsberg [1989] and to dialectical systems such as Loui's [1987].

5.2 Default Reasoning and Conditional Entailment

Conditional entailment combines in a single framework the benefits of extensional and conditional interpretations of defaults. It is natural to ask then, whether conditional entailment provides a complete account of default reasoning. The answer, not surprisingly, is no; default reasoning appears to involve aspects like *causality*, which are not captured by either conditionals or minimality considerations, and thus, which escape conditional entailment.

Consider for example a simple version of the Yale shooting problem [Hanks and McDermott, 1987]. The problem states that a gun loaded at time t_0 is shot at a later time t_1 at a person alive at t_1. The question is to predict the fate of the person after the shooting. The relevant relations can be encoded in a background context K with sentences:

$$\texttt{loaded}_0 \wedge \delta_1 \Rightarrow \texttt{loaded}_1$$

$$\texttt{alive}_1 \wedge \delta_2 \Rightarrow \texttt{alive}_2$$

$$\texttt{shoot}_1 \wedge \texttt{loaded}_1 \Rightarrow \neg\texttt{alive}_2$$

$$\texttt{shoot}_1 \wedge \texttt{loaded}_1 \Rightarrow \neg\delta_2$$

and defaults $\texttt{loaded}_0 \rightarrow \delta_1$ and $\texttt{alive}_1 \rightarrow \delta_2$. It is simple to check that the context $T = \langle K, E \rangle$ with $E = \{\texttt{loaded}_0, \texttt{alive}_1, \texttt{shoot}_1\}$ gives rise to two classes of minimal models: a class $\mathcal{C}_1$ comprising models in which the person dies and thus the assumption δ_2 is violated; and a class $\mathcal{C}_2$ comprising models in which the gun is unloaded and thus the assumption δ_1 is violated. Since the assumption δ_1 is not constrained to have a priority higher than the assumption δ_2, nor vice versa, both classes of models are equally preferred, and the 'expected' conclusion $\neg\texttt{alive}_2$ is not sanctioned.

This and related problems require the explicit handling of *causal explanations*. Indeed, we prefer to adopt δ_1 and reject δ_2 because the commitment to δ_1 *explains* the violation of δ_2 but not vice versa. Several proposals have been reported in the literature for accommodating causal explanations in a default framework. We have recently explored some of these issues in [Geffner, 1989] and [Geffner, 1990].

References

[Adams, 1975] E. Adams. *The Logic of Conditionals.* D. Reiter, Dordrecht, 1975.

[Adams, 1978] E. Adams. A note comparing probabilistic and modal logics of conditionals. *Theoria*, 43:186–194, 1978.

[Baker and Ginsberg, 1989] A. Baker and M. Ginsberg. A theorem prover for prioritized circumscription. *Proceedings IJCAI-89*, pages 463–467, Detroit, MI., 1989.

[de Kleer, 1986] J. de Kleer. An assumption-based truth maintenance system. *Artificial Intelligence*, 28:280–297, 1986.

[Delgrande, 1987] J. Delgrande. An approach to default reasoning based on a first-order conditional logic. *Proceedings AAAI-87*, pages 340–345, Seattle, 1987.

[Geffner and Pearl, 1990] H. Geffner and J. Pearl. A framework for reasoning with defaults. In H. Kyburg, R. Loui, and G. Carlson, editors, *Knowledge Representation and Defeasible Inference*, pages 69–87. Kluwer, The Netherlands, 1990.

[Geffner, 1988] H. Geffner. On the logic of defaults. *Proceedings AAAI-88*, pages 449–454, St. Paul, MN, 1988.

[Geffner, 1989] H. Geffner. *Default Reasoning: Causal and Conditional Theories.* PhD thesis, Computer Science Department, UCLA, Los Angeles, CA, November 1989.

[Geffner, 1990] H. Geffner. Causal theories for nonmonotonic reasoning. *Proceedings AAAI-90*, pages 524–530, Boston, MA, 1990.

[Gelfond and Przymusinska, 1986] M. Gelfond and H. Przymusinska. Negation as failure: careful closure procedure. *Artificial Intelligence*, 30:273–287, 1986.

[Goldszmidt and Pearl, 1989] M. Goldszmidt and J. Pearl. Deciding consistency of databases containing defeasible and strict information. *Proceedings Workshop on Uncertainty in AI*, pages 134–141, 1989.

[Goldzmidt *et al.*, 1990] M. Goldzmidt, P. Morris, and J. Pearl. A maximum entropy approach to nonmonotonic reasoning. *Proceedings AAAI-90*, pages 646–652, Boston, MA, 1990.

[Hanks and McDermott, 1987] S. Hanks and D. McDermott. Non-monotonic logics and temporal projection. *Artificial Intelligence*, 33:379–412, 1987.

[Kraus *et al.*, 1989] S. Kraus, D. Lehmann, and M. Magidor. Preferential models and cumulative logics. *Artificial Intelligence*, ?, 1989.

[Lehmann and Magidor, 1988] D. Lehmann and M. Magidor. Rational logics and their models: a study in cumulative logic. Technical report, Dept. of Computer Science, Hebrew University, Jerusalem 91904, Israel, November 1988.

[Lehmann, 1989] D. Lehmann. What does a conditional knowledge base entail? *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 212–222, Toronto, Ontario, 1989. Morgan Kaufmann.

[Lewis, 1973] D. Lewis. *Counterfactuals.* Harvard University Press, Cambridge, MA, 1973.

[Lifschitz, 1985] V. Lifschitz. Computing circumscription. *Proceedings IJCAI-85*, pages 121–127, Los Angeles, CA, 1985.

[Lifschitz, 1987] V. Lifschitz. Formal theories of action. *Proceedings of the 1987 Workshop on the Frame Problem in AI*, pages 35–57, Kansas, 1987.

[Lifschitz, 1988] V. Lifschitz. Circumscriptive theories: a logic-based framework for knowledge representation. *Journal of Philosophical Logic*, 17:391–441, 1988.

[Loui, 1987] R. Loui. Defeat among arguments: A system of defeasible inference. *Computational Intelligence*, 1987.

[Makinson, 1989] D. Makinson. General theory of cumulative inference. In M. Reinfrank *et al.*, editor, *Proceedings of the Second International Workshop on Non-Monotonic Reasoning*, pages 1–18, Berlin, Germany, 1989. Springer Lecture Notes on Computer Science.

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

[McCarthy, 1986] J. McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28:89–116, 1986.

[McDermott and Doyle, 1980] D. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13:41–72, 1980.

[Moore, 1985] R. Moore. Semantical considerations on non-monotonic logics. *Artificial Intelligence*, 25:75–94, 1985.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, Los Altos, CA., 1988.

[Pearl, 1989] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. *Proceedings of the First Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 505–516, Toronto, Canada, 1989.

[Pearl, 1990] J. Pearl. System Z: A natural ordering of defaults with tractable applications to non-monotonic reasoning. In R. Parikh, editor, *Theoretical Aspects of Reasoning about Knowledge*, pages 121–135, San Mateo, CA, 1990. Morgan Kaufmann.

[Pollock, 1987] J. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[Poole, 1988] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[Przymusinski, 1987] T. Przymusinski. On the declarative semantics of stratified deductive databases and logic programs. In J. Minker, editor, *Foundations of Deductive Databases and Logic Programming*, pages 193–216. Morgan Kaufmann, Los Altos, CA, 1987.

[Reiter and Criscuolo, 1983] R. Reiter and G. Criscuolo. Some representational issues in default reasoning. *Int. J. of Computers and Mathematics*, 9:1–13, 1983.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 12:81–132, 1980.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[Shoham, 1988] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*. MIT Press, Cambridge, Mass., 1988.

A Proofs

Lemma 1 *If the quadruple $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$ is a prioritized preferential structure, then the pair $\langle \mathcal{I}_{\mathcal{L}}, < \rangle$ is a preferential structure.*

Proof From Definition 6, the relation $M < M'$ holds iff $\Delta[M] \neq \Delta[M']$, and for every δ in $\Delta[M] - \Delta[M']$ there exists a δ' in $\Delta[M'] - \Delta[M]$, such that $\delta \prec \delta'$, where '$\prec$' is an irreflexive and transitive relation which does not contain infinite ascending chains. First, note that the relation '$<$' must too be irreflexive. We next show that '$<$' is also transitive. Let M_1, M_2, and M_3 be three interpretations such that $M_1 < M_2$ and $M_2 < M_3$, and let $\Delta_1 = \Delta[M_1]$, $\Delta_2 = \Delta[M_2]$, and $\Delta_3 = \Delta[M_3]$. We will use the notation $\overline{\Delta}$ to denote the complement of a set Δ , i.e. $\overline{\Delta} = \Delta_{\mathcal{L}} - \Delta$, and will write $\Delta_{i_1,i_2,\ldots,i_n}$ to denote the intersection of the sets $\Delta_{i_1}, \ldots, \Delta_{i_n}$. Furthermore, when one of the indices i is preceded by a minus sign, Δ_i is to be replaced by its complement $\overline{\Delta_i}$. Thus, for example, $\Delta_{1,-2,3}$ stands for the intersection of the sets Δ_1, Δ_3 and the complement $\overline{\Delta_2}$ of Δ_2. Similarly, $\Delta_{-1,2}$ stands for the intersection of $\overline{\Delta_1}$ and Δ_2.

To prove transitivity we thus need to show that for every assumption δ in $\Delta_{1,-3}$, there is an assumption δ' in $\Delta_{-1,3}$ such that $\delta \prec \delta'$.[13] First note that since '$\prec$' does not contain infinite ascending chains, it is sufficient to prove this for every *maximal* element δ in $\Delta_{1,-3}$. Let then δ_1 be an arbitrary maximal element in $\Delta_{1,-3}$. We need to consider two cases:

1. if δ_1 belongs to $\Delta_{1,-2,-3}$, then δ_1 must also belong to $\Delta_{1,-2}$. Thus, since $M_1 < M_2$, there must be an assumption δ_2 in $\Delta_{-1,2}$ such that $\delta_1 \prec \delta_2$. Furthermore, let δ_2 be the maximal such element. If $\delta_2 \in \Delta_{-1,2,3}$ we are done. Otherwise, $\delta_2 \in \Delta_{-1,2,-3}$, and then since $M_2 < M_3$, there must be an assumption $\delta_3 \in \Delta_{-2,3}$ such that $\delta_2 \prec \delta_3$. Now, if $\delta_3 \in \Delta_{1,-2,3}$, then from $M_1 < M_2$, there must be a δ_2' in $\Delta_{-1,2}$ such that $\delta_3 < \delta_2'$, and

[13]A similar proof can be found in [Przymusinski, 1987].

therefore, $\delta_2 \prec \delta_2'$, in contradiction with the maximality of δ_2. Thus, $\delta_3 \in \Delta_{-1,3}$, and $\delta_1 \prec \delta_3$, by the transitivity of '$\prec$'.

2. if δ_1 belongs to $\Delta_{1,2,-3}$, then, since $M_2 < M_3$, there must be a δ_3 in $\Delta_{-2,3}$ such that $\delta_1 \prec \delta_3$. Moreover, if $\delta_3 \in \Delta_{-1,-2,3}$ we are done. Otherwise, $\delta_3 \in \Delta_{1,-2,3}$, and therefore, as a result of $M_1 < M_2$, there must be a δ_2 in $\Delta_{-1,2}$ such that $\delta_3 \prec \delta_2$. Let δ_2 be a maximal such element. Then if δ_2 belongs to Δ_3 we are done. Otherwise, $\delta_2 \in \Delta_{-1,2,-3}$, and therefore, there must be a δ_3' in $\Delta_{-2,3}$ such that $\delta_2 \prec \delta_3'$. Furthermore, δ_3' cannot belong to Δ_1; otherwise, there should be another element δ_2' in $\Delta_{-1,2}$, such that $\delta_3' \prec \delta_2'$, contradicting the maximality of δ_2. So, $\delta_3' \in \Delta_{-1,3}$ and $\delta_1 \prec \delta_3'$ by transitivity of '$\prec$.' ∎

Lemma 2 *In a given prioritized preferential structure, if M is a preferred model of a theory T, then M is model of T* minimal *in $\Delta[M]$, i.e. there is no model M' of T such that $\Delta[M'] \subset \Delta[M]$.*

Proof If $\Delta[M'] \subset \Delta[M]$, then $\Delta[M'] - \Delta[M] = \emptyset$, and $M' < M$ would trivially hold in every prioritized preferential structure preventing M from being preferred. ∎

Theorem 2 *If an assumption based default theory $T = \langle K, E\rangle$ preferentially entails a proposition p, then T also conditionally entails p.*

Proof If T is logically inconsistent, the result is trivial. So let us assume that T is logically consistent. We will show that if $\xi = \langle \mathcal{I}_\mathcal{L}, <, \Delta_\mathcal{L}, \prec\rangle$ is a prioritized structure admissible with $K = \langle L, D\rangle$, then $\pi = \langle \mathcal{I}, <\rangle$, where $\mathcal{I} \subseteq \mathcal{I}_\mathcal{L}$ stands for the collection of models of L, will be a preferential structure admissible with K. Thus if T does not conditionally entail p, T will not preferentially entail either. Note that since we are assuming $\mathcal{L}$ to be a finite propositional language, the preferential structure π must be well-founded.

We need to show that for every default $p \rightarrow \delta$ in D, δ holds in all the preferred models of the theory $T' = \langle K, \{p\} \rangle$ in π. Again the result is trivial if T' is logically inconsistent, so we will assume otherwise. Let then M' be a model of T' in which δ does not hold, i.e. $\delta \in \Delta[M']$. We will now construct a model M preferred to M' in which δ holds. Since the preference order '$<$' is well-founded, this is sufficient to prove that δ holds in all preferred models of T'. Let C stand for the collection of all minimal conflict sets in T' (i.e., minimal sets of assumptions logically inconsistent with T), and let C' stand for the collection of all minimal conflict sets Δ in T' such that $\Delta \cap \Delta[M'] = \{\delta\}$. Since the priority ordering '$\prec$' is admissible, any such set Δ must contain an assumption δ' such that $\delta' \prec \delta$. Let Δ' stand for the collection of all such assumptions δ', and let us select M as an interpretation which satisfies T', with a gap $\Delta[M] = \Delta[M'] + \Delta' - \{\delta\}$. There must be one such interpretation as $\Delta[M]$ is a *hitting set* for C (i.e., $\Delta[M]$ contains at least one assumption for every conflict set in C; see [Reiter, 1987]). Indeed, any conflict set in C not 'hit' by assumptions from $\Delta[M'] - \{\delta\}$ will be certainly 'hit' by assumptions from Δ'. Furthermore, $M < M'$ must hold, as $\Delta[M] - \Delta[M'] = \Delta'$, $\Delta[M'] = \Delta[M] = \{\delta\}$, and for every δ' in Δ, $\delta' \prec \delta$ holds. ∎

Lemma 3 *A proposition q is* conditionally entailed *by a default theory $T = \langle K, E \rangle$ iff q holds in all preferred models of T of every* minimal *prioritized preferential structure admissible with K.*

Proof It is sufficient to show that if $\prec'$ is an admissible priority ordering that properly contains all the tuples in a *minimal* admissible priority ordering '$\prec$', then the preferred models in the structure $\xi' = \langle \mathcal{I}_{\mathcal{L}}, <', \Delta_{\mathcal{L}}, \prec' \rangle$ will be a subset of the preferred models of the structure $\xi = \langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$. Assume otherwise, that M is a preferred model in ξ' but not in ξ. Then there must be a model M' such that $M' < M$. This means that for every assumption δ in $\Delta[M'] - \Delta[M]$ there is an assumption δ' in $\Delta[M] - \Delta[M']$ such that

$\delta \prec \delta'$. However, since $\delta \prec \delta'$ implies $\delta \prec' \delta'$, then M' would be preferred to M in ξ' as well, in contradiction with the minimality M in ξ'. ∎

Lemmas 4 and 5 are special cases of lemma 6, and the latter is a special case of the 'if part' of Theorem 4 (Δ is a stable argument iff $C = \{\Delta\}$ is a stable cover). Let us recall that we say that an assumption is *free* in a default theory T if it does not belong to any minimal conflict set, and is *bound* otherwise. Furthermore, we also said that a ***default theory*** T is ***bound*** when it gives rise to a finite number of bound assumptions. The proof of Theorem 4 appeals to the well-foundedness of bound default theories captured in the following lemma:

Lemma 7 *If $T = \langle K, E \rangle$ is a bound default theory and $\xi = \langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$ is a prioritized structure, then for every non-preferred model M of T there is a preferred model M' of T such that $M' < M$.*

Proof If T is logically inconsistent, the lemma follows trivially. So let us assume that T is logically consistent and let C stand for the collection of all minimal conflict sets that T gives rise to. It is easy to show that for every hitting set Δ for C (see above) there is a model M of T such that $\Delta[M] = \Delta$, and vice versa, that if M is a model of T, then $\Delta[M]$ must include a hitting set Δ for C. Furthermore, since T is bound, there must be a finite number of minimal hitting sets Δ_i, $i = 1, \ldots, n$. Thus let M_i, stand for n models of T such that $\Delta[M_i] = \Delta_i$ and let $\mathcal{M}$ stand for the collection of all such models. Furthermore, let $\mathcal{M}_p$ denote the *minimal* collection of models in $\mathcal{M}$ such that if $M \in \mathcal{M} - \mathcal{M}_p$ then $\mathcal{M}_p$ contains a model M_i such that $M_i < M$. It is simple to show that such collection of models $\mathcal{M}_p$ is unique. We show furthermore that they are all preferred models of T. Assume otherwise that there is a model M' of T preferred to some M_i in $\mathcal{M}_p$. This implies that the gap $\Delta[M']$ of M' contains some hitting set Δ_j, and thus, that if M' is preferred to M_i, so will be M_j, in contradiction with the selection of M_i. We are thus left to show that for every non-preferred model M of T there is a model in $\mathcal{M}_p$ preferred to M. Two cases need to considered. If $\Delta[M] = \Delta_i$, $1 \leq i \leq n$, then a model will be preferred to M if and only if it is preferred to M_i above. Since M is not a preferred model of T, then M_i must belong to $\mathcal{M} - \mathcal{M}_p$, and thus, there must be a model M_j in $\mathcal{M}_p$ preferred to M_i, and therefore, to M. If for no i, $1 \leq i \leq n$, $\Delta[M] = \Delta_i$, then there must be one such i for which $\Delta[M] \supset \Delta_i$. In that case, $M_i < M$, and since $\mathcal{M}_p$ must contain a model M_j preferred to M_i, by transitivity, $M_j < M$. ∎

Theorem 4 *A proposition p is conditionally entailed if and only if p is supported by a stable cover.*

Proof (if part) Since we are assuming that the theory $T = \langle K, E \rangle$ under consideration is bound, by lemma 7 above, it is sufficient to show that for any model M which violates assumptions from every set Δ_i, $i = 1, \ldots, n$ in the cover, and any structure $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$ admissible with K, there is model M', $M' < M$, such that one of the assumption sets Δ_i is satisfied. Without loss of generality we can select M to be a minimal model, so that the set Δ' of assumptions validated by M is maximal. If there is no such a minimal model M, we are done, because as stated in the previous lemma, there would be a minimal model M' that satisfies some Δ_i, such that $\Delta[M'] \subset \Delta[M]$, and thus, $M' < M$. We assume thus that Δ' is maximal and in conflict with every set Δ_i in the cover. Since the cover is stable though, it must then contain a set Δ_i strongly protected from Δ'. That is, for every subset Δ'_j of Δ' in conflict with Δ_i, there is a subset Δ^j_i of Δ_i in conflict with Δ', such that $\Delta'_j \prec \Delta^j_i$. That means that every set Δ'_j in Δ' in conflict with Δ_i contains an assumption δ'_j such that $\delta'_j \prec \delta^j_i$, for some assumption δ^j_i in $\Delta[M]$. Let Δ'' stand for the collection of all those assumptions δ'_j in Δ'. Then, it is possible to build a model M' of T that satisfies Δ_i such that $\Delta[M'] - \Delta[M] \subseteq \Delta''$. Thus for every assumption δ'_j in $\Delta[M'] - \Delta[M]$ there is an assumption δ^j_i in $\Delta[M] - \Delta[M']$ such that $\delta'_j \prec \delta^j_i$, and therefore, $M' < M$. ∎

Proof (only if part) If $T = \langle K, E \rangle$ is a bound theory, then T is well-founded and for any structure $\langle \mathcal{I}_{\mathcal{L}}, <, \Delta_{\mathcal{L}}, \prec \rangle$ T gives rise to a finite number of preferred classes of models. Let $\Delta_1, \Delta_2, \ldots, \Delta_n$, be the *maximal* sets of assumptions validated by the preferred classes of T. Since p is conditionally entailed this means that every such set supports p. We will show now that the collection C of sets $\Delta_1, \ldots, \Delta_n$ constitutes a stable cover, i.e. that for any definite conflicting argument Δ', the cover contains an argument Δ_i strongly protected from Δ'. For that purpose, let M be a model of T satisfying Δ' and let M_i be a preferred model of T satisfying Δ_i, $1 \leq i \leq n$, such that $M_i < M$. From the previous lemma we know that there must be one such model. We show now that Δ_i is strongly protected from Δ'. Assume otherwise, i.e. there is a subset Δ'_j of Δ' in conflict with Δ_i such that for every set $\Delta_i^j \subseteq \Delta_i$ in conflict with Δ', $\Delta'_j \not\prec \Delta_i^j$. This implies that the set Δ_A of assumptions δ in Δ such that $\Delta'_j \prec \delta$ is consistent with Δ'. Furthermore, since Δ' is a *definite* conflicting argument this means that $\Delta_A \subset \Delta'$, and therefore, that for every assumption δ' in $\Delta[M] - \Delta[M_i]$, $\Delta'_j \not\prec \delta'$. However, this contradicts $M_i < M$; indeed, since Δ'_j is inconsistent with Δ_i, one of the assumptions δ'_j in Δ'_j must belong to $\Delta[M_i] - \Delta[M]$, and for $M_i < M$ to be true, another assumption δ_i, such that $\delta'_j \prec \delta_i$ must belong to $\Delta[M] - \Delta[M_i]$. ∎

Theorem 5 *For two sets of assumptions Δ and Δ', the relation $\Delta' \prec \Delta$ holds in every priority ordering '$\prec$' admissible with a* consistent background *$K = \langle L, D \rangle$ if and only if Δ is part of a set Δ'' that* dominates *Δ' in K.*

Proof (if part) Let us recall, that we use the notation $\Delta' \prec \Delta$ to state that for every assumption δ in Δ there exists an assumption δ' in Δ' such that $\delta' \prec \delta$. Moreover, the relation '$\prec$' among *sets* of assumptions remains irreflexive and transitive, and therefore, asymmetric. That is, for every priority ordering $\Delta \not\prec \Delta$, and if $\Delta_1 \prec \Delta_2$ and $\Delta_2 \prec \Delta_3$ hold, so does $\Delta_1 \prec \Delta_3$.

Let Δ stand for a collection of assumptions δ_i, $i = 1, \ldots, n$. We will use the notation $\Delta_{i,j}$, for $i \leq j$, to stand for the set $\{\delta_i, \delta_{i+1}, \ldots, \delta_j\}$. If $j > n$, the notation $\Delta_{i,j}$ is to be understood as $\Delta_{i,n}$, and if $i > n$, $\Delta_{i,j}$ as the empty set. We show first that if Δ dominates a set Δ' then the relation $\Delta' \prec \Delta$

must hold for any priority ordering '$\prec$' admissible with K. We show this by induction; the base case $\Delta_{2,n} + \Delta' \prec \Delta_{1,1}$ first. Clearly, if Δ dominates Δ', the assumption δ_1 must dominate $\Delta_{2,n} + \Delta'$, and thus, $\Delta_{2,n} + \Delta' \prec \delta_1$ must hold. Thus, if $n = 1$, we are done. So let us assume that n is greater than one. Furthermore, let us assume as inductive hypothesis that $\Delta_{i+1,n} + \Delta' \prec \Delta_{1,i}$ holds for every i, $1 \leq i < n$. We need to show the same relation for $i = n$, for which $\Delta_{n+1,n} = \emptyset$ and $\Delta_{1,n} = \Delta$. By hypothesis, we have $\{\delta_n\} + \Delta' \prec \Delta_{1,n-1}$, since $\Delta_{n,n} = \{\delta_n\}$. Let Δ_A stand for the maximal set of assumptions in $\Delta_{1,n-1}$ for which the relation $\{\delta_n\} \prec \Delta_A$ holds, and let Δ_B stand for $\Delta_{1,n-1} - \Delta_A$. Then, since the assumption δ_n dominates $\Delta + \Delta'$, there must an assumption δ' in $\Delta + \Delta'$, such that $\delta' \prec \delta_n$. Furthermore, δ' cannot belong to Δ_A, because $\{\delta_n\} \prec \Delta_A$, and the relation $\prec$ is asymmetric. So there are two cases to consider. If δ' belongs to the set Δ', then by transitivity we would have $\Delta' \prec \Delta_A$, and therefore, $\Delta' \prec \Delta_{1,n}$. On the other hand, if $\delta' \in \Delta_B$, the relation $\Delta' \prec \delta_n$ must hold by transitivity on Δ_B, since the way Δ_B was selected guarantees $\Delta' \prec \Delta_B$. Furthermore, by transitivity on δ_n, $\Delta' \prec \Delta_A$ must hold as well, and therefore, $\Delta' \prec \Delta_{1,n}$, and thus $\Delta' \prec \Delta$.

Proof (only if part) This part of the proof is slightly more involved. We need to show that if the relation $\Delta' \prec \Delta$ holds for every admissible ordering with a (conditionally) consistent background context K, then Δ is part of a set that dominates Δ'. Let us first divide the assumptions in $\Delta_{\mathcal{L}}$ between those which participate in a set that dominates Δ', which we group into a set Δ_A, from those which do not participate in a set that dominates Δ'. Furthermore, let $\Delta_B = \Delta' - \Delta_A$, and $\Delta_C = \Delta_{\mathcal{L}} - \Delta_A - \Delta_B$. Note that Δ_B cannot be empty, otherwise Δ_A would dominate itself, precluding K from being consistent. Note also, that if two sets dominate Δ', so will their union. It follows then that Δ_A dominates Δ'. Our goal will be to show that Δ is included in Δ_A. For that we will show that there is a priority ordering '$\prec$' admissible with K, such that the relation $\Delta' \prec \delta$ holds only if $\delta \in \Delta_A$.

Let us say that a priority ordering '$\prec$' in a background context K is admissible within a *range* Δ and a *restriction* Δ' iff every set Δ'' dominated by an assumption δ in Δ contains an assumption δ' in Δ', such that $\delta' \prec \delta$ holds. The notions of *range* and *restriction* provide a finer measure of the admissibility of a priority ordering. In particular, an admissible priority ordering, must be admissible within a range $\Delta_{\mathcal{L}}$ and a restriction $\Delta_{\mathcal{L}}$. Furthermore, if a priority relation '$\prec$' is admissible within a range Δ_1 and a restriction Δ_2, for two sets Δ_1 and Δ_2 such that $\Delta_1 + \Delta_2 = \Delta_{\mathcal{L}}$, then there must be a priority relation '$\prec$' admissible within a range Δ_1 and restriction $\Delta_{\mathcal{L}}$, such that $\delta_2 \prec \delta_1$ holds only if $\delta_1 \in \Delta_1$ and $\delta_2 \in \Delta_2$. Indeed, if '$\prec'$' is a priority relation admissible within a range Δ_1 and a restriction Δ_2, the relation that results by deleting all pairs $\langle \delta_1 \notin \Delta_1, \delta_2 \notin \Delta_2 \rangle$ for which $\delta_1 \prec' \delta_2$ holds, remains irreflexive, transitive, and admissible.

Now, let us assume that there is no priority ordering admissible within a range Δ_C and a restriction Δ_C, for Δ_C as above. It is possible to show then, that there must be a non-empty subset Δ'_C of Δ_C such that each assumption $\delta' \in \Delta'_C$ dominates the set $\Delta'_C + \overline{\Delta_C}$, where $\overline{\Delta_C}$ stands for the set of assumptions not in Δ_C; in this case, $\Delta_A + \Delta_B$. This, however, amounts to say that Δ'_C *dominates* the set $\Delta_A + \Delta_B$, which by virtue of the dominance of Δ_A over Δ' and the inclusion of Δ_B in Δ', implies that Δ'_C dominates Δ' as well, in contradiction with the maximality of Δ_A. Thus, there must be a priority ordering '$\prec_C$' admissible within a range Δ_C and a restriction Δ_C, such that $\delta \prec_C \delta'$ holds only if δ and δ' both belong to Δ_C. Furthermore, since K is consistent, there must be a priority ordering '$\prec_A$' admissible within range Δ_A and restriction $\Delta_{\mathcal{L}}$, such that $\delta \prec_A \delta'$ holds only if δ' belongs to Δ_A. We can thus define a relation '$\prec$' such that $\delta \prec \delta'$ iff $[\delta \prec_A \delta']$ or $[\delta \prec_C \delta']$ or $[\delta \in \Delta_C$ and $\delta' \in \Delta_A + \Delta_B]$. It is simple to show that such a relation is a priority relation, and that it is admissible within a range $\Delta_A + \Delta_C$. Let us assume, on the other hand, that '$\prec$' is not admissible within a range Δ_B. That is, there is an assumption δ in Δ_B which dominates a set Δ'_B for which the relation $\Delta'_B \prec \delta$ fails to hold. Note that Δ'_B cannot contain elements from Δ_C; for, otherwise, the relation Δ'_B will certainly hold. Thus,

$\Delta'_B \subseteq \Delta_A + \Delta_B$, so that δ dominates $\Delta_A + \Delta_B$. That means, however, that the set $\Delta_A + \{\delta\}$ dominates the set Δ', in contradiction with the assumption that Δ_A is the maximal such set. So, the ordering '$\prec$' must be admissible within the range Δ_B as well, and so '$\prec$' must be a priority relation admissible with K. Since $\Delta' \prec \Delta$ holds by hypothesis, and $\Delta' \prec \delta$ holds only if $\delta \in \Delta_A$, it follows that Δ belongs to a set, Δ_A, which dominates Δ'. ∎