Sound and Efficient Non-monotonic Inference

Hector Geffner
Depto de Computacion
Umversidad Simon Bolivar
Aptdo 89000, Caracas 1080
Venezuela

Jimena Llopis
Depto de Matematicas
Umversidad Simon Bolivar
Aptdo 89000, Caracas 1080
Venezuela

Gisela Mendez
Depto de Matematicas
Univ Central de Venezuela
Aptdo 47002, Caracas 1041
Venezuela

Abstract

We analyze conditions that allow for sound and efficient non-monotonic inference For that we consider theories comprised of rule6 and observations and a semantic framework developed elsewhere that allows us to view such theories as dynamic systems systems with a *transition function f* that maps states to sets of possi ble successor states and a *plausibility funr tion* that determines the relative likelihood of those transitions In this framework the transition function f is determined by the rules and the plausibility function is provided independently In this work we aim to identify plausibility functions that have good semantical and computational properties We do so b> identifying a vet *of tore predictions* to be accounted for that can be computed in polynomial time, can be justified in simple terms and are not tied to either Horn theories or closed world assumptions The resulting functions allow us to handle an interesting class of theories in a justifiable and efficient manner

1 Introduction

Non-monotonic reasoning is widely perceived to be inefficient and sevtral theoretical studies have confirmed that suspicion for a large class of theories (e g , [Kautz and Selman, 1989]) Yet, nianj practical *systems* including inheritance reasoners time map managers and logic programs with negation as failure suggest that non-monotonic reasoning can be both practical and effi cient in many cases of interest Stratified logic programs, for example, are tractable when negation is interpreted as negation as failure [Van Gelder *et al* 1988] and intractable when interpreted classically

In this work, we aim to understand conditions that allow for sound and efficient non-nonmonotonic inference For that we appeal to the semantic frampwork recently introduced in [Geffner 1995, Geffner and Bonet 1995] in which systems of strict and defeasible rules are viewed as dynamic systems with a transition function / that maps states pie, to sets *f(s,*) of possible successors and a plausibility function pie that determines the relative plausibility of those possible transitions

In this framework, the transition function $/(s_1,)$ is given by the collection of states $*_{I+1}$ such that the Iran sitions from *s,* to $s_{1+}j$ violate a minimal set of default rules Ihe plausibility function pie, on the other hand is provided independently and it's role is to determine the assumptions one is willing to make about missing information In particular when the plausibility function *it* is uniform, 1 e it assigns the same plausibility number to all stales, no assumptions are made and the resulting semantics is *monotonic* in the set of observations while *non-monotonii* in the set of rules * On the other hand when JT is compatible with the C losed World Assumption (CWA) namelv $ir_{,UJ1}(s,)$ stands for the number of atoms true m state s, (the higher the number the less plausible the state), the resulting semantics is n on-mono tonic in the observations and corresponds to the interpretations of *negation as failure* found in logic programs In neither case, however, Ihe semantics is adequate n the first case is too weak and in the second is too strong Interestingly however while in the first case the semantics is intractable, in the second case the semantics is tractable

In this work we will look for plausibility functions that havt good semantical and computational properties W will do so by identifying a set of core *predictions* to be accounted for that can be computed in polynomial time can bejuslified in various wavs and are not tied to either Horn theories or closed world assumptions Ihe resulting functions will enable us to handIt an interesting class of theories in *a. justifiable* and *efficient* manner

The plan of the paper is as follows First we review the semantic framework laid out in [Geffner 1995] and [Geffner and Bonet, 1995] (Section 2) fh en we con sider the virtues and limitations of the CWA interpreta tion and define the *core predictions* in three ways by means of a procedure a 3-valued logical interpretation and an epistemi. interpretation (Section 3) We then characterize other tractable conclusions that follow from the admissible plausibility functions (presuppositions) stud} semantic and syntactic conditions that guarantee the existence of such functions and assess the admissi-

[1] For temporal theories this eemantirs in very much like Gelfond and Linschitz's [1993] semantic for actions and both are equivalent to Sandewalls [1991] form of chronolog ical minimization See [Geffner,]995]

bility of some plausibility orderings commonly used and *show how such orderings can be constructed* (Section 4) We end with a summary of the main ideas (Section 5)

2 The Basic Framework A review

2 1 Language

We deal with theories $T = (D, O)$ comprised of a finite set of rules D and a finite set of observations O *Rules* can be either strict or defeasible and are represented by expressions of the form $\alpha_1 \wedge \alpha_2 \wedge \quad \wedge \alpha_n \Rightarrow \beta$ and $\alpha_1 \wedge \alpha_2 \wedge \quad \wedge \alpha_n \rightarrow \beta$ respectively with $n \geq 1$ The α_i s β s and the *observations* are assumed to be literals i e atoms or their negation

We often speak of *variables* rather than propositions Thus, two literals a and $\neg a$ will be associated with the values t and f of an abstract boolean variable A We say that a formula involves or refers to the variable A when the formula involves one of the literals a or $\neg a$

Each defeasible rule has a *priority* number represented by a non-negative integer Intuitively, among a set of conflicting rules, i e rules with complementary consequents those with higher priority are preferred In this work, we assume that the theories are *weakly deterministic* in the sense that whenever two rules are in conflict one rule must have priority over the other [2] These theories are common in certain domains and theories which don't comply with this condition can normally be mapped into a family of theories that do (by breaking the ties among the rules that are in conflict)

We also assume that rules are *acyclic* For that we define the *causal graph* associated with a set of rules as the directed graph whose nodes are the variables and where for every rule there is a link connecting every variable referred to in the antecedent to the variable referred to in the consequent The rules are *acyclic* when the causal graph does not contain cycles Acyclic logic programs are an special case of these theories in which both defeasible rules and rules with negative heads are excluded Other inheritance and temporal theories can be represented in this framework in a straightforward way (see the references)

2 2 Semantics

The semantics of the theories $T = (D, O)$ is given by dynamic-systems type of structures (σ, f, π, O), where σ is a ranking function that assigns a non-negative integer $\sigma(A)$ to every variable A, f is a state transition function π is a plausibility function and O stands for the observations

In temporal theories the σ-ranking of a variable is simply the time-point associated with the variable [Geffner 1995] In more general causal theories there is no explicit notion of time and the role of the ranking σ is to make it up σ assigns a time point to every variable with the only restriction that causes must precede their effects (i e variables in rule antecedents must have a smaller σ-rank than variables in the consequent) There are of course many rankings σ that satisfy this condition yet given certain conditions on the transition and plausibility functions f and π the resulting semantics is insensitive to the particular σ chosen [Geffner and Bonet 1995] Here for simplicity we will choose one particular ranking function σ_0 throughout the paper the one that corresponds to the topological sort of the causal graph Thus the rank of the variables that have no parents in the causal graph will be zero, while the rank of the rest of the variables will be the length of the longest paths in the graph leading to the variable

Given the ranking σ_0, the *states* s_i at time i are defined as the truth-valuations of the variables A assigned to time i (i e with $\sigma_0(A) = i$) The *state transition function* f maps states s_i into a non-empty set $f(s_i)$ of possible successor states s_{i+1} Given a weakly-deterministic domain theory, $f(s_i)$ is defined as the collection of states s_{i+1} that satisfy the consequents of all the rules that are *applicable* given s_i where a rule $A \rightarrow \beta$ is applicable given s_i when A is true in s_i and no conflicting rule with higher priority is applicable in s_i [3]

The *plausibility function* π determines the relative likelihood of the legal *transitions* by assigning to each state s_{i+1} a non-negative integer $\pi(s_{i+1})$ [4] Better states, i e , states with higher plausibilities have smaller numbers and vice versa Formally the plausibility $\kappa_\sigma(s_{i+1}|s_i)$ of a *transition* from a state s_i to a state $s_{i+1} \in f(s_i)$ that satisfies the observations O at time $i+1$ is defined as

$$\kappa_\sigma(s_{i+1}|s_i) \stackrel{\text{def}}{=} \pi(s_{i+1}) - \pi^*_\sigma(f(s_i)) \qquad (1)$$

where $\pi^*_\sigma(f(s_i))$ is a normalizing constant that refers to the plausibility of the best states $s'_{i+1} \in f(s_i)$ that satisfy the observations at $i+1$ We also define $\kappa(s_0)$ the plausibility of the transition to s_0 as

$$\kappa_\sigma(s_0) \stackrel{\text{def}}{=} \pi(s_0) - \pi^*_\sigma(S_0) \qquad (2)$$

where $\pi^*_\sigma(S_0)$ is another normalizing constant that refers to the plausibility of the best initial states s'_0 that satisfy the observations at 0

A *legal* trajectory τ is a sequence of states $s_0\ s_1, \quad, s_n$ where n is the highest σ_0-rank such that each s_i satisfies the observations at i and $s_i \in f(s_{i-1})$ for $i > 0$ The *plausibility* of a legal trajectory given the observations O is simply the sum of the plausibilities associated with its transitions i e $\kappa_\sigma(\tau) = \kappa_\sigma(s_0) + \sum_{1\ N} \kappa_\sigma(s_i|s_{i-1})$ The plausibility of illegal trajectories is assumed to be infinite (for the probabilistic interpretation of this model, see the references)

The semantics of a theory $T = (D, O)$ given a plausibility function π is identified with its most plausible trajectories (recall that the transition function f is determined by the domain theory D) We refer to the trajectories τ that have a plausibility $\pi(\tau) = 0$ as *normal trajectories* When normal trajectories exist, they

[2] A rule r has priority over a rule r' when the two rules are defeasible and r has higher priority or when r' is defeasible and r is not

[3] This definition of f is not general and assumes that the theory is markovian See below

[4] The notation $f(s_i)$ and $\pi(s_{i+1})$ is actually an abbreviation for $f(i, s_i)$ and $\pi(i\ s_i)$ as both functions depend not only on the state s_i but also on the time point i

are the most plausible trajectories Yet such trajectories may fail to exist when predictions are refuted by observations We call the theories that have normal trajectories *predictive theories* To a large extent they will be the focus of this work

Example Let us consider a domain theory composed of a defeasible rule $p \rightarrow q$ and a strict rule $r \Rightarrow \neg q$ The ranking σ_0 assigns the variables P and R to the layer 0 and variable Q to layer 1, i e , $\sigma_0(P) = \sigma_0(R) = 0$ and $\sigma_0(Q) = 1$ The set S_0 of states s_0 are the four possible valuations of the variables P and Q, and the set S_1 of states s_1 are the two possible valuations of the variable Q For any state $s_0 \in S_0$ $f(s_0)$ denotes the states in S_1 that satisfy the consequents of all the rules applicable in s_0 Hence, when s_0 does not satisfy either p or r, $f(s_0) = \{s_q, s_{\neg q}\}$ when s_0 satisfies $p \wedge \neg r$, $f(s_0) = \{s_q\}$ and when s_0 satisfies r $f(s_0) = \{s_{\neg q}\}$ Given the observation p and the uniform plausibility function π_u i e the function that assigns the same plausibility number to all states, we get two normal trajectories one corresponds to p, $\neg r$ and q, and the other to p r and $\neg q$ On the other hand given the CWA-plausibility function instead where $\pi_{cwa}(s_i)$ measures the number of atoms satisfied by s_i, we get only one normal trajectory that corresponds to p $\neg r$ and q

Non-Markovian Theories The definition of f above presumes that all the variables in the antecedent of a rule belong to the same σ_0 layer and that the variable in the consequent of the rule belongs to immediately succeeding σ_0-layer We call theories where this holds *markovian* Many temporal theories are markovian in this sense yet most non-temporal theories are not There are two general ways for dealing with non-markovian theories One is to reformulate the notion of state in the general case a state s_i is no longer a valuation over the variables in layer i alone but over *all* the variables in layer j, $j \leq i$ (see [Geffner and Bonet 1995] for details) A second way is to transform the theories into equivalent markovian theories This can be easily done by introducing surrogate variables For example if a variable A belongs to layer i and some rule requires A in layer $i+1$ we just introduce a new variable A' and relate it to A by means of two strict rules $a \Rightarrow a'$ and $\neg a \Rightarrow \neg a'$ The two methods are general and can be applied to any non-markovian theory For this reason, and to keep things simple we focus in this paper on theories that are markovian All the results easily carry to non-markovian theories as well

3 Tractable Inference without the CWA

In the semantic framework laid out above the defaults get compiled into the transition function f while the plausibility functions encodes the assumptions one is willing to make to 'complete' partially specified states The CWA plausibility function, for example, presumes that variables whose values are not known are false This is adequate sometimes, but it is often too strong For example, if a rule $\neg w \Rightarrow \neg q$ is added above, the CWA assumption would make the *exception* $\neg w$ more plausible than it's negation, not only failing to conclude q from p, but actually concluding $\neg q$ This limitation of the CWA ordering is not surprising though the CWA ordering is determined without considering the rules in the theory A better approach is to define the plausibility ordering b\ looking first at such rules as done in the interpretation that view defaults *at,* conditionals (e g [Lehmann 1989 Pearl 1990 Geffner 1992]) Howe\er since extracting a reasonable plausibility ordering from the rules remains too difficult, what we will do instead is to define some *core predictions* that can be justified in simple terms and can be computed in polynomial time, ind then focus on the plausibility functions that make those predictions sound The resulting semantics will combine the properties of causal and conditional interpretations of defaults [Geffner 1992], and will have a significant core of infer-(nces that can be computed efficiently

3 1 The Core Predictions

The core predictions are conclusions that can be justified in simple meaningful terms and which can be computed in polynomial time We will provide three different but equivalent characterization of them The first is a tractable inference procedure very similar to other procedures used in non-monotonic systems (e g [Horty *et al* 1987]) Basically the procedure applies all the rules $A \rightarrow \beta$ whose antecedents have been either observed or derived as long no higher priority rule $B \rightarrow \sim\beta$ with a conflicting consequent also has its antecedents observed or derived

In the procedure, the rules with *highest precedence* in a set S of rules $A \rightarrow [\neg]p$ refer to the rules in S whose consequents are ranked lowest (i e have a minimum $\sigma_0(P)$) The symbol L represents the conclusions (literals) computed by the algorithm We say that a rule $A \rightarrow \beta$ is applicable given L when $A \subseteq$ L The function new-applicable-rules(p S) returns the rules that become applicable when S is known and p is derived i e the rules $A \rightarrow p$ such that A holds in $\{p\} + S$ but not in S The complement of a literal x is denoted as $\sim x$ For simplicity strict rules are handled as defeasible rules of higher priority [5]

Procedure A

```
L = O   RULES = rules applicable given L
while RULES not empty do
  SELECT = rules in RULES with highest precedence
  RULES = RULES − SELECT
  while SELECT not empty do
    pick a highest priority rule A → β from SELECT
    L = L + {β}
    RULES = RULES + new-applicable-rules(β L)
    SELECT = SELECT − {rules B → β & B → ∼β}
  end while
end while
```

Proposition 1 *The complexity of* A *is linear in the set of rules*

[5] For weakly deterministic predictive theories that are predictive strict rules can actually be *replaced* by defeasible rules with higher priority without affecting the meaning of the theory

Proposition 2 *Given the* CWA *plausibility function, A is sound and atomic-complete for weakly deterministic theories that are predictive*

We will refer the literals in L for a theory T as the *core predictions* of T and denote them as $Pd(T)$

3.2 A 3-Valued Interpretation

Two different models will shed some light into what this procedure actually computes The first model is defined in terms a transition and a plausibility function but assumes a 3-valued logic The plausibility function π^3_{cwa}, unlike π_{cwa}, maximizes not 'falsehoods but uncertainties (don't know's) The transition function $f(s_i)$, as before, maps a state s_i into the set of states s_{i+1} that satisfies the consequents of all the rules applicable in s_i The only difference is that states are now 3-valued interpretations, rules continue being applicable when their antecedents are true and no conflicting rules with higher priority are applicable

The *normal* trajectories are defined as above They now stand for 3-valued models so some literals may be neither true or false We call this the 3-CWA interpretation It's very simple to check this interpretation makes the predictions computed by the procedure A *sound and complete*

Proposition 3 *Let $T = (D\ O)$ be weakly-deterministic theory which is predictive relative to the 3-*CWA *interpretation, namely T has a normal 3-*CWA *trajectory Then this trajectory is unique and the core predictions are exactly the literals that are true in that trajectory*

3.3 An Epistemic Interpretation

We consider now a different model where the rules are interpreted epistemically Given a domain theory D with strict and defeasible rules of the form if A then β, we will consider a new domain theory $\mathrm{K}(D)$ where those rules are replaced by their epistemic counterparts, i e epistemic rules of the form if K A then β', where K is an intensional knowledge operator We will interpret the resulting theories with a semantic structure consisting only of a transition function F_{K} that maps *sets* of states S_i into *sets* of states S_{i+1} The individual states s_i are, once again, classical interpretations over the variables occurring in layer i, and $F_{\mathrm{K}}(S_i)$, in analogy to $f(s_i)$ is defined as the set of states s_{i+1} that satisfy the consequents of all the rules applicable in the *set* of states S_i, where a rule if K A then β' is applicable in S_i when A holds *in all* the states $s_i \in S_i$ and no conflicting rule with higher priority is also applicable

By the *epistemic interpretation* of a theory $T = (D, O)$ we will mean the succession $S_0, S_1, \ldots, S_n$ of non-empty sets of states, such that S_0 is the set of states that satisfy the observations at time 0, and S_i, $i > 0$, is the set of states in $F_{\mathrm{K}}(S_{i-1})$ that satisfy the observations at time i The literals sanctioned by this interpretation are the literals that are not falsified by any such state

Proposition 4 *Let $T = (D\ O)$ be a weakly deterministic theory Then T has a 3* CWA *normal trajectory iff T has an epistemic interpretation In either case the literals that hold in both models correspond exactly to the core predictions*

4 Admissible Orderings

The core predictions are thus the conclusions that follow from an epistemic reading of the rules The *admissible* plausibility functions are as the functions that make those predictions sound

Definition 1 *π_a is an* admissible *plausibility function relative to a weakly-deterministic domain theory D with transition function f, if for any set of observations O, the core predictions $Pd(T)$ are true in all the π_a normal trajectories compatible with O*

The semantics that results from the admissible plausibility functions is stronger than many conditional interpretations of defaults, since given a rule $A \rightarrow \beta$ β will be a core prediction of A, except in the pathological cases in which there is a higher priority rule $A' \rightarrow \neg\beta$ with $A' \subseteq A$ While conditional interpretations read such rules as saying that 'if A is *all that is known* then β is true the proposed interpretation reads them as if A is known and no exception is known then β is true This is similar to the epistemic interpretation of temporal theories in [Shoham 1988]

We say that a domain theory D is *admissible* when it admits an admissible plausibility ordering We also say that a theory $T = (D, O)$ is *admissible* when there is normal trajectory compatible with one such admissible ordering We call such trajectories the admissible normal trajectories of T A basic result for positive theories is [6]

Proposition 5 *The* CWA *plausibility function is admissible for positive domain theories Moreover a theory $T = (D\ O)$ with a positive domain theory D has an admissible normal trajectory iff T has a* CWA *normal trajectory*

Moreover since all the positive literals that can be obtained by the CWA interpretation for positive theories are core predictions we also get that

Corollary 1 *Let T be a positive theory with a* CWA *normal trajectory Then the* atoms *that are true in that trajectory correspond exactly to the atoms that are true in all the admissible normal trajectories of T (atomic soundness and completeness)*

The account in [Geffner, 1994] is characterized exactly by these atoms This does not mean however, that there are no *negative consequences* to be extracted from positive theories Actually, there may be plenty of them and their computation is tractable

4.1 Tractable Presuppositions

The epistemic and 3-CWA interpretations sanction all the conclusions $Pd(T)$ computed by the procedure A and nothing else Yet we can obtain more conclusions from A by 'reductio ad absurdum', i e , by assuming a value for literals whose truth has not been determined and checking its repercussions This operation leads to a sound and tractable extension of the set of core predictions $Pd(T)$ which in certain cases is complete

[6] Positive theories refer to theories with no rules with negated antecedents

Let us define $Pp_0(T)$ as the set of literals $\alpha \notin Pd(T)$ which if assumed to be false introduce new predictions β which are in conflict with the old ones, i.e. $\beta \in Pd(T + \{\sim\alpha\})$ and $\sim\beta \in Pd(T)$. Moreover, let us define $Pp_i(T)$ iteratively for $i > 0$, as the union of $Pp_{i-1}(T)$ and the literals $\alpha \notin Pd(T + Ps_{i-1}(T))$ such that for some literal β, $\beta \in Pd(T + Pp_{i-1}(T) + \{\sim\alpha\})$ and $\sim\beta \in Pd(T + Ps_{i-1}(T))$.

We will call the literals in $Pp_0(T)$ simple presuppositions and the other literals in $Pp_i(T)$, $i > 0$, iterated presuppositions. Let us also define $Pp(T)$ as the set of all simple and iterated presuppositions, i.e. $Pp(T) = Pp_i(T)$ for the smallest i such that $Pp_i(T) = Pp_{i+1}(T)$.

Proposition 6 *All presuppositions are true in all the admissible normal trajectories (soundness). Moreover computing all presuppositions is tractable.*

Proposition 7 *Let T be an admissible positive theory i.e. a positive theory with admissible normal trajectories. Every literal that is true in all such trajectories is either a core prediction or a simple presupposition (completeness).*

Corollary 2 *For a positive T as above, computing the positive and negative consequences of T is tractable.*

4.2 Constructing Admissible Orderings

Semantic Conditions

Below we provide two semantic conditions that allow us to determine when a given plausibility function π is admissible. S_i and S_{i+1} refer to the collection of states s_i and s_{i+1}, $i = 1, 2, \ldots$, that satisfy two arbitrary sets of literals A_i and A_{i+1} in layers i and $i+1$. S_i^* denotes the best states in S_i according to π and $\pi^*(S_i)$ denotes the plausibility of the best states in S_i. F_K is the transition function associated with the epistemic interpretation. The conditions are:

1. $\forall s_i \in S_i^*,\ f(s_i) \subseteq F_K(S_i)$
2. $\forall s_i \in S_i^*,\ \pi^*(f(s_i) \cap S_{i+1}) = \pi^*(F_K(S_i) \cap S_{i+1})$

Proposition 8 *If a plausibility function π satisfies conditions 1-2 for any two collections of states S_i and S_{i+1} as above, π is admissible.*

Syntactic Conditions

Let us say that one rule r *preempts* another rule r' when r has priority over r' and the two rules have conflicting consequents. We will call the literals that occur in the antecedent of r *exceptions* relative to the rule r' or simply exceptions. Let's also say that a literal α is *connected* to a literal β in a given domain theory when there is a chain of rules $r_1, r_2, \ldots, r_n$ (i.e. the consequent of each rule r_i occurs in the antecedent of the rule r_{i+1} for $i = 1, \ldots, n-1$), where the literal α occurs in antecedent of r_1 and β is the consequent of r_n. Each literal is also 'connected' to itself.

Then we say that a domain theory is *normal* when for every atom a, if a is connected to an exception, then $\neg a$ is not connected to any exception.

Proposition 9 *Normal domain theories are admissible.*

Below we show how to construct admissible orderings for theories that are normal. Although this is not needed for accepting the core predictions and presuppositions it will shed some light on the adequacy of some of the plausibility orderings that have been proposed.

Some Admissible Orderings

We will consider first a very simple class of normal theories where there are no chains of rules. In other words the ranking σ_0 assigns all the variables that occur in the antecedent of the rules to the 0-layer, and all the variables that occur in the consequent of the rules to the 1 layer. We call such theories, 2 layered theories.

For these theories, finding an admissible plausibility function reduces to finding an ordering on the states s_0 that satisfies Condition 1 above. It is interesting to see how some of the standard plausibility orderings fare in this regard. For example, the uniform plausibility ordering π_u will work only in theories in which there are no pairs of rules in conflict. The CWA ordering π_{cwa} on the other hand will work whenever there are no negative exceptions.

Let's consider now the plausibility function π_{min} that minimizes default violations, i.e., $\pi_{min}(s)$ measures the number of rules violated in s where a rule $A \rightarrow \beta$ is violated in s when s satisfies A but some state $s' \in f(s)$ fails to satisfy β. Since π_{min} does not reflect the rule priorities let's assume that all rules have the same priority. Does π_{min} comply with Condition 1 above? Interestingly the answer is not: for example, the theory comprised of the rules $a \rightarrow b$, $c \Rightarrow \neg b$, $\neg a \Rightarrow d$ and $f \rightarrow \neg d$ is normal and yet the best π_{min} states compatible with the literals c and f contain a state s_0 where $\neg a$ holds that leads to d in contradiction with F_K that yields $\neg d$.

A modification of π_{min}, however, yields a plausibility function that is admissible for all (2-layered) normal theories. Given a state s, let s_A denote the closest state to s where A holds, where the distance between two states is measured by the number of atoms that have different truth-values. Let us also say that a rule r: $A \rightarrow \beta$ *is preempted* in a state s when the rule r is *violated* in the state s_A (this is just the standard notion of violation when rules are understood as *conditionals*, see [Nute 1984]). Then the plausibility function $\pi_{pmi}(s)$ that measures the number of rules *preempted* in s is admissible relative to *all* normal 2-layered theories.

Let's finally consider the plausibility function π_z that corresponds Pearl's Z-ranking, which captures both the conditional reading of rules and certain independence assumptions (see [Pearl, 1990] and the equivalent ranking due to Lehmann [1989]). Basically the states s with Z-plausibility 0 are the states that don't violate any rule, while inductively, the states s' with Z-plausibility i are the states that violate rules $A \rightarrow \beta$ which have been *verified* in states s with Z-plausibility j, $j < i$ (a state s verifies a rule when s satisfies both the rule *and* its antecedent). The Z-plausibility function is *not* admissible yet an admissible function π_w can be defined by a slight modification of the above definition: the states s' with W-plausibility i need to be inductively defined as the states s' that violate rules $A \rightarrow \beta$ such that *all the states s that verify the rule $A \rightarrow \beta$ and that are closest*

to s' have a W-plausibility j, $j < i$ This W-plausibility ordering can be shown to exist for all 2-layered normal theories

Proposition 10 *The plausibility functions π_{pmi} and π_u are admissible relative to all 2-layered normal domain theories The plausibility functions π_u π_{cwa} π_{min} and π are not*

General Normal Theories The plausibility functions π_{pmi} and π_u above can in principle be extended to n-layered normal theories but the construction is complex There is nonetheless a simple construction that applies to *all* normal theories Let us say that a literal α is *normal* when α is not an exception and is not connected to any exception Then the plausibility function $\pi_{ncp}(s)$ that measures the number of non-normal literals made true by s is admissible

Proposition 11 *The plausibility function π_{ncp} is admissible relative to all normal domain theories*

Note that in the case of positive domain theories, the only non-normal literals are positive literals In that case the CWA-plausibility function can be thought as a stronger version of the function π_{ncp} that assumes *all* positive literals to be non-normal

The plausibility function π_{ncp} has an additional useful property it's decomposable in the sense that the plausibility of a state is the sum of the plausibilities associated with each of the literals it satisfies (the plausibility of normal and non-normal literals is 0 and 1 respectively) As shown in [Geffner and Bonet, 1995], this guarantees that the behavior determined by π_{ncp} is insensitive to the ranking σ used and that like in Bayesian Networks [Pearl, 1988] every variable is independent of its non-descendants given the value of its parents

5 Summary

We have identified a basic core of inferences that can be justified in simple terms and have developed a semantics that makes those inferences sound The semantics has good computational properties combines elements from causal conditional and extensional interpretations of defaults, and sheds some light on the adequacy of various plausibility ordenngs and on the scope of interpretations that view default rules epistemically

Acknowledgments

I want to thank yoav Shoham for the chance to spend time in Stanford with my family and for many useful discussions that led to this work I also want to thank Nir Friedman and Joe Halpern (HG)

References

[Geffner and Bonet, 1995] H Geffner and B Bonet Causal systems as dynamic s>stems Submitted, 1995

[Geffner 1992] H Geffner *Default Reasoning Causa! and Conditional Theories* MIT Press Cambridge MA, 1992

[Geffner, 1994] H Geffner Causal default reasoning Principles and algorithms In *Proceedings AAAI-94* pages 245-250 Seattle, WA 1994 MIT Press

[Geffner, 1995] H Geffner Dynamic systems and qualitative markov processes in logic Submitted, 1995

[Gelfond and Lifschitz 1993] M Gelfond and V Lifschitz Representing action and change by logic programs *J of Logic Programming,* 17 301-322, 1993

[Hortv *el* a/, 1987] J Horty, R Thomason and D Touretzky A skeptical theory of inheritance In *Proceedings AAAI 87,* pages 358-363 Seattle WA , 1987

[kautz and Selman 1989] H Kautz and B Selman Hard problems for simple default logics In *Proceedings hR 89,* pages 189-197 Toronto, Ontario 1989

[Lehmann 1989] D Lehmann What does a conditional knowledge base entail[7] In *Proceedings hR 89,* pages 212 222, Toronto Ontario 1989 Morgan Kaufmann

[Nute J984] D Nute Conditional logic In D Gabbay and F Guenthner editors, *Handbook of Philosophical Logic,* pages 387-439 D Rcidel Dordrecht, 1984

[Pearl 1988] J Ptarl *Probabilistic Reasoning* in *Intelligent Systems* Morgan kaufmann, CA , 1988

[Pearl 1990] J Pearl System Z A natural ordering of defaults In R Pankh editor, *TARK 90* pages 121-135 CA, 1990 Morgan kaufmann

[Sandewal, 1991] E Sandewal Features and fluents Technical Report R-91-29, CS Department Linkoping University Linkoping, Sweden 1991

[Shoham 1988] \ Shoham *Reasoning about Change Time and Causation from the standpoint of Artificial Intelligence* MIT Press Cambridge, Mass , 1988

[Van C.elder *et at* , 1988] A Van Gelder, k Ross and J S Schlipf Unfounded sets and well-founded semantics for general logic programs In *Proceedings PODS'88* pages 221-230, 1988