In Theoretical Aspects of Reasoning About Knowledge, (TARK-III),
R. Parikh (ed.), Morgan Kaufmann, 1990, pp. 121-135.

TECHNICAL REPORT
R-131
6/14/91

SYSTEM Z: A NATURAL ORDERING OF DEFAULTS WITH TRACTABLE APPLICATIONS TO NONMONOTONIC REASONING(*)

Judea Pearl

Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles

Abstract

Recent progress towards unifying the probabilistic and model preference semantics for nonmonotonic reasoning has led to a remarkable observation: Any consistent system of default rules imposes an unambiguous and natural ordering on these rules which, to emphasize its simple and basic character, we term "**Z**-ordering." This ordering can be used with various levels of refinement, to prioritize conflicting arguments, to rank the degree of abnormality of states of the world, and to define plausible consequence relationships. This paper defines the **Z**-ordering, briefly mentions its semantical origins, and illustrates two simple entailment relationships induced by the ordering. Two extensions are then described, maximum-entropy and conditional entailment, which trade in computational simplicity for semantic refinements.

1. Description

We begin with a set of rules $R = \{r: \alpha_r \rightarrow \beta_r\}$ where α_r and β_r are propositional formulas over a finite alphabet of literals, and $\rightarrow$ denotes a new connective to be given default interpretations later on. A truth valuation of the literals in the language will be called a *model*. A model M is said to *verify* a rule $\alpha \rightarrow \beta$ if $M \models \alpha \wedge \beta$ (i.e., α and β are both true in M), and to *falsify* $\alpha \rightarrow \beta$ if $M \models \alpha \wedge \neg \beta$.

Given a set R of such rules, we first define the relation of *toleration*.

(*) This work was supported in part by National Science Foundation grant #IRI-86-10155 and Naval Research Laboratory grant #N00014-89-J-2007.

Definition 1: A set of rules $R' \subseteq R$ is said to *tolerate* an individual rule r, denoted $T(r \mid R')$, if the set of formulas $(\alpha_r \wedge \beta_r) \bigcup_{r' \in R'} (\alpha_{r'} \supset \beta_{r'})$ is satisfiable, i.e., if there exists a model that verifies r and does not falsify any of the rules in R'.

To facilitate the construction of the desired ordering, we now define the notion of *consistency*.

Definition 2: A set R of rules is said to be *consistent* if for every non-empty subset $R' \subseteq R$ there is at least one rule that is tolerated by all the others, i.e.,

$$\forall R' \subseteq R, \exists\ r' \in R', \text{ such that } T(r' \mid R' - r') \tag{1}$$

This definition, named p-consistent in [Adams 1975] and ε-consistent in [Pearl 1988], assures the existence of an *admissible* probability assignment when rules are given a probabilistic interpretation. In other words, if each rule $\alpha \rightarrow \beta$ is interpreted as a statement of high conditional probability, $P(\beta \mid \alpha) \geq 1 - \varepsilon$, consistency assures that for every $\varepsilon > 0$ there will be a probability assignment P (to models of the language) that satisfies all these statements simultaneously. An identical criterion of consistency also assures the existence of an *admissible* preference ranking on models, when each rule $\alpha \rightarrow \beta$ is given a model-preference interpretation, namely, β is true in all the most preferred models of α [Lehmann and Magidor 1988].

A slightly more elaborate definition of consistency applies to databases containing mixtures of defeasible and nondefeasible rules [Goldszmidt and Pearl 1989a]. Note that the condition of consistency is stronger than that of mere satisfiability. For example, the two rules $a \rightarrow b$ and $a \rightarrow \neg b$ are satisfiable (if a is false) but not consistent. Intuitively, consistency requires that in addition to satisfying the constraint associated with the rule $a \rightarrow b$, the truth of a should not be ruled out as an impossibility. This reflects the common understanding that a conditional sentence "if a then b" is not fully satisfied by merely making a false; it requires that both a and b be true in at least one possible world, however unlikely.

The condition of consistency, Eq. (1), leads to a natural ordering of the rules in R. Given a consistent R, we first identify every rule that is tolerated by all the other rules of R, assign to each such rule the label 0, and remove it from R. Next, we attach a label 1 to every rule that is tolerated by all the remaining ones, and so on. Continuing in this way, we form an ordered partition of $R = (R_0, R_1, R_2, \cdots R_K)$, where

$$R_i = \{r : T(r \mid R - R_0 - R_1 - \cdots R_{i-1})\} \tag{2}$$

The label attached to each rule in the partition defines the Z-ranking or $\mathbf{Z}$-ordering. The process of constructing this partition also amounts to testing the consistency of R, because it terminates with a full partition iff R is consistent [Goldszmidt and Pearl 1989a].

Theorem 1: The complexity of testing the consistency of a set of rules is $O[PS(n)N^2]$, where N is the number of rules, n the number of literals in R and $PS(n)$ the complexity of propositional satisfiability in the sublanguage characterizing the rules (e.g., $PS(n) = O(n)$ for Horn expressions).

Proof: Identifying R_0 takes $N \cdot PS(n)$ steps, identifying R_1 takes $(N - |R_0|)PS(n)$ steps, and so on. Thus, the total time it takes to complete the labeling is

$$PS(n)[N + (N - |R_0|) + (N - |R_0| - |R_1|) + \cdots] \leq PS(n)[N + (N-1) + \cdots]$$

$$= PS(n)\frac{N^2}{2} \tag{3}$$

In order to define the notions of entailment and consequence it is useful to translate the ranking among rules into preferences among models. The reason is that we wish to proclaim a formula g to be a plausible consequence of f, written $f \vdash g$, only if the constraints imposed by R would force the models of $f \wedge g$ to stand in some preference relation over those of $f \wedge \neg g$. For example, the traditional preferential criterion for g to be a rational consequence of f requires that all the most preferred models of f satisfy g, i.e., that all the most preferred models of f reside in $f \wedge g$ and none resides in $f \wedge \neg g$ [Shoham 1987]. We shall initially limit ourselves to such preference criteria that do not require substantial enumeration of models, i.e., that the preference between $f \wedge g$ and $f \wedge \neg g$ be readily tested using the partition defined in Eq. (2). To that purpose, we propose the following ranking on models. Using $\mathbf{Z}(r)$ to denote the label assigned to rule r,

$$\mathbf{Z}(r) = i \quad \textit{iff} \quad r \in R_i \ , \tag{4}$$

we define the rank associated with a particular model M as the lowest integer n such that all rules having $\mathbf{Z}(r) \geq n$ are satisfied by M,

$$\mathbf{Z}(M) = \min \{n : M \models (\alpha_r \supset \beta_r) \quad \mathbf{Z}(r) \geq n\} \tag{5}$$

In other words, the rank of a model is equal to 1 plus the rank of the highest-ranked rule falsified by the model. The rank associated with a given formula f is now defined as the lowest $\mathbf{Z}$ of all models satisfy-

ing f,

$$\mathbf{Z}(f) = \min\{\mathbf{Z}(M): M \models f\} \tag{6}$$

Note that, once we establish the ranking of the rules, the complexity of determining the $\mathbf{Z}$ value of any given M is $O(N)$; we simply identify the highest $\mathbf{Z}$ rule that is falsified by M and add 1 to its $\mathbf{Z}$. More significantly, determining the $\mathbf{Z}$ value of an arbitrary formula f requires at most N satisfiability tests; we search for the lowest i such that all rules having $\mathbf{Z}(r) \geq i$ tolerate $f \rightarrow true$, i.e.,

$$\mathbf{Z}(f) = \min\{i: T(f \rightarrow true \mid R_i, R_{i+1}, ...)\} \tag{7}$$

Eq. (5) defines a total order on models, with those receiving a lower $\mathbf{Z}$ interpreted as being more normal or more preferred. This ordering satisfies the constraints that for each rule $\alpha_r \rightarrow \beta_r$, β_r holds true in all the most-preferred models of α_r, namely, the usual preferential model interpretation of default rules. It can be shown (see Appendix I) that the rankings defined by Eqs. (4) and (5) correspond to a special kind of a preferential structure; out of all rankings satisfying the rule constraints, the assignment defined in Eq. (5) is the only one that is *minimal*, in the sense of assigning to each model the lowest possible ranking (or highest normality) permitted by the rules in R.

2. Consequence Relations

We are now ready to define two notions of nonmonotonic entailment. Given a knowledge base in the form of a consistent set R of rules, and some factual information f, we wish to define the conditions under which f can be said to entail a conclusion g, in the context of R.

Definition 3 (0-entailment): g is said to be *0-entailed* by f in the context R, written $f \vdash_0 g$, if the augmented set of rules $R \cup f \rightarrow \neg g$ is inconsistent.

Theorem 2: 0-entailment is semi-monotonic, i.e., if $R' \subseteq R$ then

$$f \vdash_0 g \text{ under } R \textbf{ whenever } f \vdash_0 g \text{ under } R'.$$

The proof is immediate, from the fact that if $R' \cup f \rightarrow \neg g$ is inconsistent, then $R \cup f \rightarrow \neg g$ must be inconsistent as well. Semi-monotonicity reflects a strategy of extreme caution; no consequence will ever be issued if it is possible to add rules to R (consistently) in such a way as to render the conclusion no longer valid. Thus, 0-entailment generates the maximal set of "safe" conclusions that can be drawn from R, and hence, was proposed in [Pearl 1989] as a *conservative core* that ought to be common to all non-

monotonic formalisms.

0-entailment was named p-entailment by Adams [1975], ε-entailment by Pearl [1988] and r-entailment by Lehmann and Magidor [1988]. Probabilistically, 0-entailment guarantees that conclusions will receive arbitrarily high probabilities (i.e., $P(g \mid f) \rightarrow 1$) whenever the premises receive arbitrarily high probabilities (i.e., $P(\beta_r \mid \alpha_r) \rightarrow 1 \; \forall \, r \in R$). In the preferential model interpretation, 0-entailment guarantees that $\kappa(f \wedge g) < \kappa(f \wedge \neg g)$ holds in *all* admissible ranking functions κ, namely, in all ranking functions $\kappa(M)$ that satisfy the rule constraints

$$\kappa(\alpha_r \wedge \beta_r) < \kappa(\alpha_r \wedge \neg \beta_r) \quad \forall \, r \in R \tag{8}$$

where, for every formula α,

$$\kappa(\alpha) = \min\{\kappa(M) : M \models \alpha\} \, . \tag{9}$$

Due to its extremely conservative nature, 0-entailment does not properly handle irrelevant features, e.g., from $a \rightarrow c$ we cannot conclude $a \wedge b \rightarrow c$ even in cases where R makes no mention of b. To sanction such inferences we now define a more adventurous type of entailment.

Definition 4: (1-entailment). A formula g is said to be 1-*entailed* by f, in the context R, (written $f \mid\!\sim_1 g$), if

$$\mathbf{Z}(f \wedge g) < \mathbf{Z}(f \wedge \neg g) \, . \tag{10}$$

Namely, there exists an integer k such that the set of rules ranked higher or equal to k tolerates $f \rightarrow g$ but does not tolerate $f \rightarrow \neg g$. Note that, once we have the **Z**-rank of all rules, deciding 1-entailment for a given query requires at most $2(1 + \log |R|)$ satisfiability tests (using a binary-search strategy). 1-entailment can be given a clear motivation in preferential model semantics. Instead of insisting that $\kappa(f \wedge g) < \kappa(f \wedge \neg g)$ hold in *all* admissible ranking functions κ, as was done in 0-entailment, we only require that it holds in the unique admissible ranking that is minimal, namely, the **Z**-ranking (see Appendix I).

Lehmann [1989] has extended 0-entailment in a slightly different way, introducing a consequence relation called *rational closure*. Rational closure is defined in terms of a relation called *more exceptional*, where a formula α is said to be more exceptional than β if

$$\alpha \vee \beta \mid\!\sim_0 \neg \alpha \, .$$

Based on this relation, Lehmann then used an inductive definition to assign a *degree* to each formula α in

the language: *degree* $(\alpha) = i$ if *degree* (α) is not less than i and every β that is less exceptional than α has *degree* $(\beta) < i$. Finally, a sentence $\alpha \rightarrow \beta$ was defined to be in the rational closure of R iff *degree* $(\alpha) <$ *degree* $(\alpha \wedge \neg \beta)$.

Goldszmidt and Pearl [1989b] have recently shown that *degree* (α) is identical to $\mathbf{Z}(\alpha)$ and, hence, rational closure is equivalent to 1-entailment. This endows the **Z**-ranking with an additional motivation in terms of exceptionality; $\mathbf{Z}(\alpha) > \mathbf{Z}(\beta)$ if α is more exceptional than β. Additionally, the computational procedure developed for 1-entailment renders membership in the rational closure decidable in at most $2(1 + \log |R|)$ satisfiability tests.

Lehmann [1989] has also shown that the rational closure can be obtained by syntactically closing the relation of 0-entailment under a rule suggested by Makinson called *rational monotony*. Rational monotony permits us to conclude $a \wedge b \vdash c$ from $a \vdash c$ as long as the consequence relation does not contain $a \vdash \neg b$. Rational monotony is induced by any admissible ranking function, not necessarily the minimal one defined by system-Z (see Appendix II). Thus, 1-entailment can be thought of as an extension of 0-entailment to acquire properties that are sound in any individual (admissible) ranking function.

1-entailment, though more adventurous than 0-entailment, still does not go far enough, as is illustrated in the next section.

3. Illustrations

Consider the following collection of rules R:

r_1: "Penguins are birds"	$p \rightarrow b$
r_2: "Birds fly"	$b \rightarrow f$
r_3: "Penguins do not fly"	$p \rightarrow \neg f$
r_4: "Penguins live in the antarctic"	$p \rightarrow a$
r_5: "Birds have wings"	$b \rightarrow w$
r_6: "Animals that fly are mobile"	$f \rightarrow m$

It can be readily verified that r_6, r_5, and r_2 are each tolerated by all the other five rules in R. For example, the truth assignment $(p = 0, a = 0, f = 1, b = 1, w = 1, m = 1)$ satisfies both

$$b \wedge w \wedge (p \supset b) \wedge (b \supset f) \wedge (p \supset \neg f) \wedge (p \supset a) \wedge (f \supset m)$$

and

$$b \wedge f \wedge (p \supset b) \wedge (b \supset w) \wedge (b \supset \neg f) \wedge (p \supset a) \wedge (f \supset m) .$$

Thus, r_6, r_5 and r_2 are each assigned a label 0 indicating that these rules pertain to the most normal state of affairs. No other rule can be labeled 0 because, once we assign p the truth value 1, we must assign 1 to b and 0 to f, which is inconsistent with $b \supset f$. The remaining three rules can now be labeled 1, because each of the three is tolerated by the other two. A network describing the six rules and their Z-labels is shown in Figure 1.

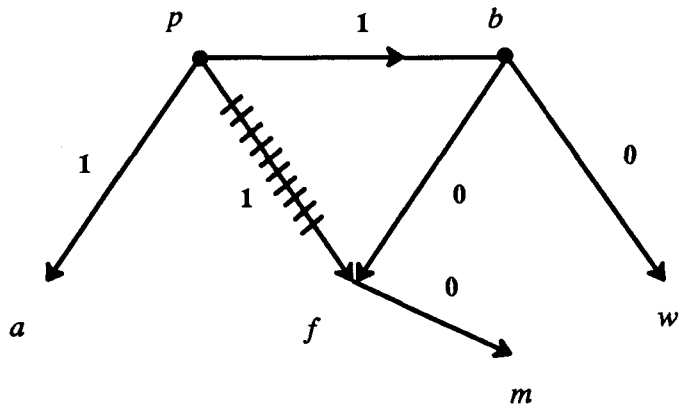

Figure 1.

The following are examples of plausible consequences one would expect to draw from R:

0-*entailed*	1-entailed	not-entailed
$b \wedge p \vdash \neg f$	$\neg b \vdash \neg p$	$p \vdash w$
$f \vdash \neg p$	$\neg f \vdash \neg b$	$p \wedge \neg a \vdash \neg f$
$b \vdash \neg p$	$b \vdash m$	$p \wedge \neg a \vdash w$
$p \wedge a \vdash b$	$\neg m \vdash \neg b$	
	$p \wedge \neg w \vdash b$	

For example, to test the validity of $b \wedge p \vdash_0 \neg f$ we add the rule r_6: $b \wedge p \rightarrow f$ to R, and realize that the augmented set becomes inconsistent; no rule in the set $\{b \wedge p \rightarrow f, p \rightarrow b, p \rightarrow \neg f\}$ can be tolerated by the other two.

1-entailment sanctions plausible inference patterns that are not 0-entailed, among them rule chaining, contraposition and the discounting of irrelevant features. For example, we cannot conclude by 0-entailment that birds are mobile, $b \vdash m$, because neither $b \rightarrow m$ nor $b \rightarrow \neg m$ would render R inconsistent. However, m is 1-entailed by b, because the rule $b \rightarrow m$ is tolerated by all rules in R while $b \rightarrow \neg m$ is tolerated by only those labeled 1. Thus,

$$\mathbf{Z}(b \wedge m) < \mathbf{Z}(b \wedge \neg m),$$

confirming Eq. (10). Similarly, if c is an irrelevant feature (i.e., not appearing in R), we obtain $b \wedge c \vdash_1 f$ but not $b \wedge c \vdash_0 f$.

On the other hand, 1-entailment does not permit us to conclude that flying objects are birds ($f \vdash b$) or that penguins who do not live in the antarctic are still birds ($p \wedge \neg a \vdash b$). This is because negating these consequences will not change their **Z**-ratings — in testing $f \vdash_1 b$ we have $\mathbf{Z}(f \wedge b) = \mathbf{Z}(f \wedge \neg b) = 0$, while in testing $p \wedge \neg a \vdash_1 b$ we have $\mathbf{Z}(p \wedge \neg a \wedge b) = \mathbf{Z}(p \wedge \neg a \wedge \neg b) = 2$.

There are cases, however, where 1-entailment produces conclusions whose plausibility may be subject to dispute. For example,(1) if we add to Figure 1 the rule $c \rightarrow f$ we obtain $\mathbf{Z}(c \rightarrow f) = 0$, which yields $c \vdash_1 \neg p$ and $c \wedge p \vdash_1 \neg f$. In other words, 1-entailment ranks the new class c to be as normal as birds, and penguins, by virtue of being exceptional kind of birds (relative to flying) are also treated as exceptional c's. Were the database to contain no information relative to birds, penguins and c's would be treated as equal status classes and the conclusion $p \wedge c \vdash \neg f$ would not be inferred. Thus, merely mentioning a property (f) by which a class (p) differs from its superclass (b) automatically brands that class (p) exceptional relative to any neutral class (c).

The main weakness of the system described so far is its inability to sanction property inheritance from classes to exceptional sub-classes. For example, neither of the two types of entailments can sanction the conclusion that penguins have wings ($p \rightarrow w$) by virtue of being birds (albeit exceptional birds). The reason is that the label 1 assigned to all rules emanating from p amounts to proclaiming penguins an exceptional type of birds in *all* respects, barred from inheriting *any* bird-like properties (e.g., laying eggs, having beaks, etc.). This is a drawback that cannot be remedied by methods based solely on the **Z**-ordering of defaults. The fact that $p \rightarrow w$ is tolerated by two extra rules ($p \rightarrow b$, and $b \rightarrow w$) on top of those tolerating $p \rightarrow \neg w$, remains undetected.

(1) This observation is due to Hector Geffner.

To sanction property inheritance, a more refined ordering is required which also takes into account the *number* of rules tolerating a formula, not merely their rank orders. One such refinement is provided by the maximum-entropy approach [Goldszmidt and Pearl 1989c] where each model is ranked by the sum of weights on the rules falsified by that model. Another refinement is provided by Geffner's conditional entailment [Geffner 1989], where the priority of rules induces a *partial* order on models. These two refinements will be summarized next.

4. The Maximum Entropy Approach

The maximum-entropy (ME) approach [Pearl 1988] is motivated by the convention that, unless mentioned explicitly, properties are presumed to be independent of one another; such presumptions are normally embedded in probability distributions that attain the maximum entropy subject to a set of constraints. Given a set R of rules and a family of probability distributions that are admissible relative the constraints conveyed by R (i.e., $P(\beta_r \rightarrow \alpha_r) \geq 1-\varepsilon \;\; \forall\, r \in R$), we can single out a distinguished distribution $P^*_{\varepsilon,R}$ having the greatest entropy $-\sum_M P(M)\log P(M)$, and define entailment relative to this distribution by

$$f \mid\!\sim_{ME} g \quad \text{iff} \quad P^*_{\varepsilon,R}(g \mid f) \underset{\varepsilon \rightarrow 0}{\rightarrow} 1 . \tag{11}$$

An infinitessimal analysis of the ME approach also yields a ranking function κ on models, where $\kappa(M)$ now corresponds to the lowest exponent of ε in the expansion of $P^*_{\varepsilon,R}(M)$ into a power series in ε. Moreover, this ranking function can be encoded parsimoniously by assigning an integer weight w_r to each rule $r \in R$ and letting $\kappa(M)$ be the sum of the weights associated with the rules falsified by M. The weight w_r, in turn, reflects the "cost" we must add to each model M that falsifies rule r, so that the resulting ranking function would satisfy the constraint conveyed by R, namely,

$$\min\{\kappa(M): M \models \alpha_r \wedge \beta_r\} < \min\{\kappa(M): M \models \alpha_r \wedge \neg\beta_r\}\,, \; r \in R$$

These considerations lead to a set of $|R|$ non-linear equations for the weights w_r which, under certain conditions, can be solved by iterative methods. Once the rule weights are established, ME-entailment is determined by the criterion of Eq. (11), translated to

$$f \mid\!\sim_{ME} g \quad \text{iff} \quad \min\{\kappa(M): M \models f \wedge g\} < \min\{\kappa(M): M \models f \wedge \neg g\}\,.$$

where

$$\kappa(M) = \sum_{r: M \not\models \alpha_r \wedge \neg \beta_r} w_r$$

We see that ME-entailment requires minimization over models, a task that may take exponential time. In practice, however, this minimization is accomplished quite effectively in databases of Horn expressions, yielding a reasonable set of inference patterns. For example, in the database of Figure 1, ME-entailment will sanction the desired consequences $p \vdash w$, $p \wedge \neg a \vdash \neg f$ and $p \wedge \neg a \vdash w$ and, moreover, it will avoid the undersirable pattern of concluding $c \wedge p \vdash \neg f$ from $R \cup \{c \rightarrow f\}$.

The weaknesses of the ME approach are two-fold. First, it does not properly handle causal relationships and, second, it is sensitive to the format in which the rules are expressed. This latter sensitivity is illustrated in the following example. From R = {Swedes are blond, Swedes are well-mannered}, ME will conclude that dark-haired Swedes are still well-mannered, while no such conclusion will be drawn from R = {Swedes are blond and well-mannered}. This sensitivity might sometimes be useful for distinguishing fine nuances in natural discourse, concluding, for example, that mannerisms and hair color are two independent qualities. However, it stands at variance with one of the basic conventions of formal logic, which treats $a \rightarrow b \wedge c$ as a shorthand notation of $a \rightarrow b$ and $a \rightarrow c$ and, moreover, unlike 1-entailment it will conclude $c \wedge p \vdash_{ME} \neg f$ from $\Delta \cup \{c \rightarrow f\}$, where c is an irrelevant property.

The failure to respond to causal information (see Pearl [1988, pp. 463, 519] and Hunter [1989]) prevents the ME approach from properly handling tasks such as the Yale shooting problem [Hanks and McDermott 1986], where rules of causal character are given priority over other rules. This weakness may perhaps be overcome by introducing causal operators into the ME formulation, similar to the way causal operators are incorporated within other formalisms of nonmonotonic reasoning (e.g., Shoham [1986], Geffner [1989]).

5. Conditional Entailment

Geffner [1989] has overcome the weaknesses of 1-entailment by introducing two new refinements. First, rather than letting rule priorities dictate a ranking function on models, a partial order on models is induced instead. To determine the preference between two models, M and M', we examine the highest priority rules that distinguish between the two, i.e., that are falsified by one and not by the other. If all such rules remain unfalsified in one of the two models, then this model is the preferred one. Formally, if $\Delta[M]$ and $\Delta[M']$ stand for the set of rules falsified by M and M', respectively, then M is preferred to M' (written $M < M'$) iff $\Delta[M] \neq \Delta[M']$ and for every rule r in $\Delta[M] - \Delta[M']$ there exists a rule r' in $\Delta[M'] - \Delta[M]$ such that r' has a higher priority than r (written $r \prec r'$). Using this criterion, a model M

will always be preferred to M' if it falsifies a proper subset of the rules falsified by M'. Lacking this feature in the Z-ordering has prevented 1-entailment from concluding $p \mid\!\sim w$ in the example of Section 3.

The second refinement introduced by Geffner is allowing the rule-priority relation, $\prec$, to become a partial order as well. This partial order is determined by the following interpretation of the rule $\alpha \rightarrow \beta$; if α is all that we know, then, regardless of other rules that R may contain, we are authorized to assert β. This means that $r: \alpha \rightarrow \beta$ should get a higher priority than any argument (a chain of rules) leading from α to $\neg \beta$ and, more generally, if a set of rules $R' \subset R$ does not tolerate r, then at least one rule in R' ought to have a lower priority than r. In Figure 1, for example, the rule $r_3: p \rightarrow \neg f$ is not tolerated by the set $\{r_1: p \rightarrow b, r_2: b \rightarrow f\}$, hence, we must have $r_1 \prec r_3$ or $r_2 \prec r_3$. Similarly, the rule $r_1: p \rightarrow b$ is not tolerated by $\{r_2, r_3\}$, hence, we also have $r_2 \prec r_1$ or $r_3 \prec r_1$. From the asymmetry and transitivity of $\prec$, these two conditions yield $r_2 \prec r_3$ and $r_2 \prec r_1$. It is clear, then, that this priority on rules will induce the preference $M < M'$, whenever M validates $p \wedge b \wedge \neg f$ and M' validates $p \wedge b \wedge f$; the former falsifies r_2, while the latter falsifies the higher priority rule r_3. In general, we say that a proposition g is conditionally entailed by f (in the context of R) if g holds in all the preferred models of f induced by every priority ordering admissible with R.

Conditional entailment rectifies many of the shortcomings of 1-entailment as well as some weaknesses of ME-entailment. However, having been based on model minimization as well as on enumeration of subsets of rules, its computational complexity might be overbearing. A proof theory for conditional entailment can be found in Geffner [1989].

Conclusions

The central theme in this paper has been the realization that underlying any consistent system of default rules there is a natural ranking of these defaults and that this ranking can be used to induce preferences on models and plausible consequence relationships. We have seen that the Z-ranking emerges from both the probabilistic interpretation of defaults and their preferential model interpretation, and that two of its immediate entailment relations are decidable in $O(N^2)$ satisfiability tests. The major weakness of these entailment relationships has been the blockage of property inheritance across exceptional subclasses. Two refinements were described, maximum-entropy and conditional entailment, which properly overcome this weakness at the cost of a higher complexity. An open problem remains whether there exists a tractable approximation to the maximum entropy or the conditional entailment schemes which permits inheritance across exceptional subclasses and, at the same time, retains a proper handling of specificity-based priority.

Acknowledgement

I am indebted to Daniel Lehmann for sharing his thoughts on the relations between r-entailment, ε-entailment, rational closure, and maximum-entropy. Hector Geffner and Moises Goldszmidt have contributed many ideas, and are responsible for the developments described in Sections 4 and 5.

References

[Adams 1975] Adams, E. 1975. *The logic of conditionals.* Dordrecht, The Netherlands: D. Reidel.

[Geffner 1989] Geffner, H. 1989. Default reasoning: causal and conditional theories. UCLA Cognitive Systems Laboratory *Technical Report (R-137)*, December 1989. PhD. dissertation.

[Goldszmidt and Pearl 1989a] Goldszmidt, M. and Pearl, J. 1989. On the consistency of defeasible databases. *Proc. 5th Workshop on Uncertainty in AI*, Windsor, Ontario, Canada, pp. 134-141.

[Goldszmidt and Pearl 1989b] Goldszmidt, M. and Pearl, J. 1989. On the relation between rational closure and System-Z. UCLA Cognitive Systems Laboratory, *Technical Report (R-139)*, December 1989. Submitted.

[Goldszmidt and Pearl 1989c] Goldszmidt, M. and Pearl, J. 1989. A maximum entropy approach to nonmonotonic reasoning. UCLA Cognitive Systems Laboratory, Technical Report R-132, in preparation.

[Hanks and McDermott 1986] Hanks, S. and McDermott, D. V. 1986. Default reasoning, nonmonotonic logics, and the frame problem. *Proc., 5th Natl. Conf. on AI (AAAI-86)*, Philadelphia, pp. 328-33.

[Hunter 1989] Hunter, D. 1989. Causality and maximum entropy updating. *Intl. Journal of Approximate Reasoning.* 3 (no. 1) pp. 87-114.

[Lehmann 1989] Lehmann, D. 1989. What does a conditional knowledge base entail? *Proc. 1st Intl. Conf. on Principals of Knowledge Representation and Reasoning (KR'89)*, Toronto, May 1989, pp. 212-222, San Mateo: Morgan Kaufmann Publishers.

[Lehmann and Magidor 1988] Lehmann, D. and Magidor, M. 1988. Rational logics and their models: a study in cumulative logics. Dept. of Computer Science, Hebrew University, Jerusalem, Israel, Technical Report #TR-88-16.

[Pearl 1988] Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo: Morgan Kaufmann Publishers.

[Pearl 1989] Pearl, J. 1989. Probabilistic semantics for nonmonotonic reasoning: a survey. *Proc. 1st Intl. Conf. on Principles of Knowledge Representation and Reasoning (KR'89)*, Toronto, May 1989, pp. 505-516, San Mateo: Morgan Kaufmann Publishers.

[Shoham 1986] Shoham, Y. 1986. Chronological ignorance: Time, nonmonotonicity, necessity, and causal theories. *Proc., 5th Natl. Conf. on AI (AAAI-86)*, Philadelphia, pp. 389-93.

[Shoham 1987] Shoham, Y. 1987. Nonmonotonic logics: meaning and utility. *Proc. Intl. Joint Conf. on AI (IJCAI-87)*, Milan, pp. 388-393.

APPENDIX I: Uniqueness of The Minimal Ranking Function

Definition: A *ranking function* is an assignment of non-negative integers to the models of the language. A ranking function κ is said to be *admissible* relative to database R, if it satisfies

$$\min \{\kappa(M): M \models \alpha_r \wedge \beta_r\} < \min \{\kappa(M): M \models \alpha_r \wedge \neg \beta_r\} \tag{I-1}$$

for every rule $r: \alpha_r \rightarrow \beta_r$ in R.

Let W stand for the set of models considered.

Definition: A ranking function κ is said to be *minimal* if every other admissible ranking κ' satisfies $\kappa'(M) > \kappa(M')$ for at least one model $M' \in W$.

Clearly, every minimal ranking has the property of "local compactness," namely, it is not possible to lower the rank of one model while keeping the ranks of all other models constant. Every such attempt will result in violating the constraint imposed by at least one rule in R. We will now show that local compactness is also a sufficient property for minimality, because there is in fact only one unique ranking that is locally compact.

Definition: An admissible ranking function κ is said to be *compact* if, for every $M' \in W$, any ranking κ' satisfying

$$\kappa'(M) = \kappa(M) \quad M \neq M'$$

$$\kappa'(M) < \kappa(M) \quad M = M'$$

is inadmissible.

Theorem (uniqueness): Every consistent R has a unique compact ranking $\mathbf{Z}(M)$ given by Eq. (5).

Corollary: Every consistent R has a unique minimal ranking given by the compact ranking $\mathbf{Z}(M)$ of Eq. (5).

Proof: We will prove that the ranking function $\mathbf{Z}$ given in Eq. (5) is the unique compact ranking. First we show, by contradiction, that $\mathbf{Z}$ is indeed compact. Suppose it is possible to lower the rank $\mathbf{Z}(M')$ of

some model M'. Let $\mathbf{Z}(M') = I$. From Eq. (5) we know that M' falsifies some rule $r: \alpha \rightarrow \beta$ of rank $\mathbf{Z}(r) = I - 1$, namely, $M' \models \alpha \wedge \neg \beta$, and there exists $\hat{M} \models \alpha \wedge \beta$ having $\mathbf{Z}(\hat{M}) = I - 1$. Lowering the rank of M' below I, while keeping $\mathbf{Z}(\hat{M}) = I - 1$ would clearly violate the constraint imposed by the rule $\alpha \rightarrow \beta$ (see Eq. (I-1)). Thus, $\mathbf{Z}$ is compact.

We now prove that $\mathbf{Z}$ is unique. Suppose there exists some other compact ranking function κ, that differs from $\mathbf{Z}$ on at least one model. We shall show that if there exists an M' such that $\kappa(M') < \mathbf{Z}(M')$ then κ could not be admissible, while if there exists an M' such that $\kappa(M') > \mathbf{Z}(M')$, then κ could not be compact. Assume $\kappa(M') < \mathbf{Z}(M')$, let I be the lowest κ value for which such inequality holds, and let $\mathbf{Z}(M') = J > I$. >From Eq. (5), M' falsifies some rule $\alpha \rightarrow \beta$ of rank $J - 1$, namely, $M' \models \alpha \wedge \neg \beta$ and every model M validating $\alpha \wedge \beta$ must obtain $\mathbf{Z}(M) \geq J - 1$. By our assumption, $\kappa(M)$ must also assign to each such M a value not lower than $J - 1 \geq I$. But this is incompatible with the constraint $\alpha \rightarrow \beta$ (see Eq. (I-1)). Thus, κ is inadmissible.

Now assume there is a non-empty set of models for which $\kappa(M) > \mathbf{Z}(M)$, and let I be the lowest $\mathbf{Z}$ value in which $\kappa(M') > \mathbf{Z}(M')$ holds for some model M'. We will show that κ could not be compact, because it should be possible to reduce $\kappa(M')$ to $\mathbf{Z}(M')$ while keeping constant the κ of all other models. From $\mathbf{Z}(M') = I$ we know that M' does not falsify any rule $\alpha' \rightarrow \beta'$ whose $\mathbf{Z}$ rank is higher than $I - 1$. Hence, we only need to watch whether the reduction of κ can violate rules r for which $\mathbf{Z}(r) < I$. However, every such rule $r: \alpha \rightarrow \beta$ has a model $M \models \alpha \wedge \beta$ having $\mathbf{Z}(M) < I$, and every such model was assumed to obtain a κ rank equal to that assigned by $\mathbf{Z}$. Hence, none of these rules will be violated by lowering $\kappa(M')$ to $\mathbf{Z}(M')$. QED.

APPENDIX II: Rational Monotony of Admissible Rankings

Theorem: The consequence relation $\vdash$ defined by the criterion

$$f \vdash g \text{ iff } \kappa(f \wedge g) < \kappa(f \wedge \neg g)$$

is closed under rational monotony, for every admissible ranking function κ.

Proof: We need to show that for every three formulas a, b and c, if $a \vdash c$, then either $a \vdash \neg b$ or $a \wedge b \vdash c$. Assume $a \vdash c$ and $a \nvdash \neg b$, namely,

(i) $\kappa(a \wedge c) < \kappa(a \wedge \neg c)$

(ii) $\kappa(a \wedge \neg b) \geq \kappa(a \wedge b)$,

we must prove

(iii) $\kappa(a \wedge b \wedge c) < \kappa(a \wedge b \wedge \neg c)$.

Rewriting (i) as

$$\kappa(a \wedge c) = \min\{\kappa(a \wedge c \wedge b, \kappa(a \wedge c \wedge \neg b)\} < \min\{\kappa(a \wedge b \wedge \neg c), \kappa(a \wedge \neg b \wedge \neg c)\} = \kappa(a \wedge \neg c)$$

we need to show only that the min on the left hand side is obtained at the second term, i.e., that

$$\min\{\kappa(a \wedge c \wedge b), \kappa(a \wedge c \wedge \neg b)\} = \kappa(a \wedge c \wedge \neg b).$$

But this is guaranteed by (ii), because the alternative possibility:

$$\kappa(a \wedge c \wedge b) < \kappa(a \wedge c \wedge \neg b)$$

together with (ii), would violate (i). QED