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

SYSTEM-Z+: A FORMALISM FOR REASONING WITH VARIABLE STRENGTH DEFAULTS

Moises Goldszmidt
Judea Pearl

July 1991
CSD-910031

To appear in Proceedings, AAAI-91, Anaheim, CA., July 1991.

TECHNICAL REPORT
R-161
April, 1991

System-Z^+: A formalism for reasoning with variable-strength defaults.*

Moisés Goldszmidt† **Judea Pearl**
< *moises@cs.ucla.edu* > < *judea@cs.ucla.edu* >
Cognitive Systems Laboratory, Computer Science Department,
University of California, Los Angeles, CA 90024

Abstract

We develop a formalism for reasoning with defaults that are expressed with different levels of firmness. Necessary and sufficient conditions for consistency are established, and a unique ranking of the rules is found, called Z^+, which renders models as normal as possible subject to the consistency conditions. We provide the necessary machinery for testing consistency, computing the Z^+ ranking and drawing the set of plausible conclusions it entails.

1 Introduction: Not All Defaults Were Created Equal

Regardless of how we choose to interpret default statements, it is generally acknowledged that some defaults are stated with greater firmness than others. For example, the action-response default "if Fred is shot with a loaded gun Fred is dead" is issued with a greater conviction than persistence defaults of the type "If Fred is alive at time t, he is alive at $t+1$." Moreover, the degree of conviction in this last statement should clearly depend on whether t is measured in years or in seconds. In diagnosis applications, likewise, the analyst may feel strongly that failures are more likely to occur in one type of devices (e.g., multipliers) than in another (e.g., adders). A language must be devised for expressing this valuable knowledge. Numerical probabilities or degrees of certainty have been suggested for this purpose, but if one is not concerned with the full precision provided by numerical calculi, an intermediate qualitative language might be more suitable.

Priorities among defaults have been proposed in many non-monotonic reasoning systems. For example, given a set of conflicting defaults, prioritized circumscription ([Lifschitz, 1988]) permits the user to identify which statement should override the other. The statement "penguins do not fly" for instance, can be given a higher priority over "birds fly", in order to enforce preferences toward the *more specific* classes. In certain systems, specificity preferences can be extracted automatically from the databases itself (e.g. [Geffner, 1989], [Kraus *et al.*, 1990]), when such information is available, say from the statement "all penguins are birds". However, certain priorities are not specificity based. For example, to reflect our intuitions that religious beliefs are stronger than political affiliations, we would like the default "typically Quakers are pacifists" to override the default "typically Republicans are not pacifists", when the two are found to conflicts with one another (say when Nixon is found to be a Quaker and a member of the Republican party). To resolve such conflict we need to encode these priorities on a rule-by-rule basis.

This paper proposes and analyzes a formalism to include priority information in the form of integers assigned to default rules, each integer signifying the degree of firmness with which the corresponding rule is stated or, alternatively, the degree of surprise (or abnormality) associated with finding the rule violated. These integers may encode linguistic quantifiers such as "typical", "highly typical", "extremely typical", etc.. They can also be viewed as powers of infinitesimals in the probabilistic interpretation of defaults, in the spirit of ϵ-semantics [Pearl, 1988], OCF [Spohn, 1987], and Kraus *et al.* [1990]. Our formalism takes after, and extends system-Z [Pearl, 1990], which proposes a conditional-preferential interpretation of defaults $\varphi \rightarrow \psi$ as saying that ψ holds in all most preferred models of φ (see [Shoham, 1987], [Kraus *et al.*, 1990], [Geffner, 1989]), but permits only one level of firmness for all defaults.

The paper is organized as follows: Section 2 introduces the concept of ranking functions on models and establishes the necessary and sufficient conditions for the existence of admissible rankings. Section 3 is concerned with the precise characterization of a privileged ranking κ^+ on models and its relation with the ranking Z^+ on rules. Its main properties, minimality, uniqueness, and the procedures necessary to compute

*This work was supported in part by National Science Foundation grant #IRI-88-21444 and State of California MICRO 90-127.

†Supported by an IBM graduate fellowship 1990-91.

the ranking and its set of plausible conclusions are presented. Section 4 provides some examples which illustrate the use of the κ^+-ranking to express belief strength, to enforce priorities among defaults and how specificity relations are maintained by system-Z^+. Finally, Section 5 discusses and evaluates the main results. All proofs are given in Appendix A.

2 Admissible Rankings

We consider a set of rules $\Delta = \{\varphi_i \xrightarrow{\delta_i} \psi_i\}$ where φ_i and ψ_i are propositional formulas over a finite alphabet of literals, "$\rightarrow$" denotes a new connective (to be given a default interpretation later on), and δ_i is a non-negative integer that measures the relative strength of the rule[1]. A truth valuation of the literals in the language will be called a *model*. A model M is said to *verify* a rule $\varphi \rightarrow \psi$ if $M \models \varphi \wedge \psi$, to *falsify* $\varphi \rightarrow \psi$ if $M \models \varphi \wedge \neg\psi$, and to *satisfy* $\varphi \rightarrow \psi$ if $M \models \varphi \supset \psi$.

Definition 1 *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 Δ, if it satisfies*

$$\min\{\kappa(M) : M \models \varphi_i \wedge \psi_i\} + \delta_i < \min\{\kappa(M) : M \models \varphi_i \wedge \neg\psi_i\} \quad (1)$$

for every rule $\varphi_i \xrightarrow{\delta_i} \psi_i \in \Delta$. A model M^+ is said to be a characteristic model *for rule $\varphi \rightarrow \psi$ relative to ranking κ, if $\kappa(M^+) = \min\{\kappa(M) : M \models \varphi \wedge \psi\}$.*

Equivalently, among the lowest ranked models satisfying the antecedent φ_i, a rule $\varphi_i \xrightarrow{\delta_i} \psi_i$ forces any model satisfying $\neg\psi_i$ to rank at least δ_i units higher than those satisfying ψ_i. This echoes the usual interpretation of defaults ([Shoham, 1987]) according to which ψ_i holds in all *minimal* models satisfying φ_i. In our case minimality is reflected in having the lowest possible ranking. The new parameter δ_i can be interpreted as the minimal *cost* or *penalty* charged to models violating rule $\varphi_i \xrightarrow{\delta_i} \psi_i$. Along the same vein we can define consequence relations as follows:

Definition 2 *Given a set Δ, $\phi \mid\!\sim \sigma$ is in the consequence relation defined by an admissible ranking κ iff every κ-minimal model for ϕ is also a model for σ.*

The next couple of definitions and Theorem 1, characterize and provide a decision procedure for testing the consistency of a set Δ, namely its ability to accommodate at least one admissible ranking.

Definition 3 *A rule $\varphi \rightarrow \psi$ is* tolerated *by Δ iff there exists a model M such that M* verifies *$\varphi \rightarrow \psi$ and* satisfies *all the sentences in Δ.*

Definition 4 *A set Δ is said to be* consistent *if there exists an admissible ranking κ for Δ.*

[1] Whenever δ is not relevant we will simply write $\varphi \rightarrow \psi$ to identify a rule.

Theorem 1 *A set Δ is consistent iff there exists a tolerated rule in every nonempty subset of Δ.*

Fortunately it is not necessary to test tolerance in every subset of Δ. A procedure for deciding consistency will need only to continuously remove tolerated sentences in Δ until Δ becomes empty. If at any point a tolerated sentence cannot be found, Δ is inconsistent. The proof of the correctness of this procedure along with the proofs of Theorem 1 and Corollary 1 can be found in Appendix A.

Corollary 1 *Deciding the consistency of a set Δ requires at most $|\Delta|^2$ propositional satisfiability tests.*

3 Plausible Conclusions and Minimal Rankings

So far we were concerned with the conditions for consistency, and we have seen that these conditions make no reference to the cost δ associated with the rules. It is reassuring to know that once a database is consistent for one set of costs assignments, it will be consistent with respect to any such assignment which means that the rule author has the freedom to modify the costs without fear of forming an inconsistent database. Our main aim however, is to draw plausible conclusions from the database and this calls for further examinations of Def. 2.

According to Def. 2 each ranking would give raise to its own consequence relation. The requirement on these rankings to be admissible, is too loose, since many vastly different rankings are capable of satisfying the constraints. A straightforward way of standardizing the conclusion set would be to require the conditions of Def. 2 to hold in all admissible rankings. This leads to an entailment relation called ε-semantics in [Pearl, 1989], 0-entailment in [Pearl, 1990], and r-entailment in [Kraus *et al.*, 1990] which is recognized as being too conservative. A more reasonable approach would be to select a distinguished admissible ranking which best reflects the spirit of default reasoning. If a model with lower κ reflects a more normal world, it is reasonable that we attempt to assign to each world the lowest possible κ permitted by the constraints. Such an attempt can be interpreted as a tendency to believe that, unless forced otherwise, each world is assumed as normal as possible. The question we need to answer is whether making one world as normal as possible would not force other worlds to become more exceptional than otherwise. This would render the set of minimal rankings non unique and the entailment conditions rather complex, reminiscent of multiple extensions in default logic ([Reiter, 1980]). Remarkably we will show that there is a unique minimal ranking, i.e., that lowering the ranking of one world does not come at an expense of another. Moreover, we provide an effective procedure for computing this minimal ranking[2].

[2] Such uniqueness condition was previously shown to

Definition 5 *Let κ^+ be a ranking function on a consistent set Δ, such that $\kappa^+(M) = 0$ if M does not falsify any rule in Δ, and otherwise,*

$$\kappa^+(M) = \max\{Z^+(r_i) : M \models \varphi_i \wedge \neg\psi_i\} + 1 \qquad (2)$$

where

$$Z^+(r_i) = \min\{\kappa^+(M) : M \models \varphi_i \wedge \psi_i\} + \delta_i \qquad (3)$$

Note that the apparent circularity between κ^+ and Z^+ is benign. Both functions can be computed recursively in an interleaved fashion. Each time we assign a κ^+ to a model according to Eq. 2, it permits us to assign a Z^+ to some rules in Eq. 3 and vice versa. This can be illustrated by tracing the first few steps: Given that Δ is consistent, there must exists at least one rule r' tolerated by Δ, i.e., at least one model M' must satisfy Δ and verify r'. By Def. 5 we can set $\kappa^+(M') = 0$, for all those models and for all such rules set $Z^+(r') = \delta'$ in accordance with Eq. 3. The κ^+ of models falsifying these rules can now be computed using Eq. 2 and so on. The details of this recursive assignment can be found in Procedure Z_rank below. Another view of Eqs. 2 and 3 can be obtained by substituting Eq. 3 into Eq. 2. Define $\mathcal{V}[M]$ to be the set of rules verified by model M, and $\mathcal{F}[M]$ be the set of rules falsified by model M, then

$$\kappa^+(M) = \max_{r_i \in \mathcal{F}[M]} [\min_{M' : r_i \in \mathcal{V}[M']} [\kappa^+(M')] + \delta_i] + 1 \qquad (4)$$

Eq. 4 illustrates that the value of $\kappa^+(M)$ is set just above the value of the characteristic model of the rules that M violates, thus "pushing down" the ranking of models to be as normal as possible. The reason that the function Z^+ is introduced in Def. 5 is that it provides an economical and convenient way of storing and manipulating the ranking κ^+. The amount of space required by the Z^+-ranking is linear on the number of default rules in the database, and once Z^+ is known, the ranking $\kappa^+(M)$ for any M can be obtained from Eq. 2 in at most $|\Delta|$ steps. To show that any function κ^+ satisfying Eqs. 2 and 3 is admissible, we re-write the conditions for admissibility (Eq. 1) as $Z^+(r_i) < \min\{\kappa^+(M) : M \models \varphi_i \wedge \neg\psi_i\}$ (using Eq. 3). Since $\kappa^+(M) = \max\{Z^+(r_i) : M \models \varphi_i \wedge \neg\psi_i\} + 1$, it follows that κ^+ is indeed admissible. The following is a step by step effective procedure for computing Z^+:

Procedure Z_rank

Input: A consistent set Δ. **Output:** Z^+-ranking on rules.

1. Let Δ_0 be the set of rules tolerated by Δ; and let $\mathcal{R}Z^+$ be an empty set.
2. For each rule $r_i : \varphi_i \xrightarrow{\delta_i} \psi_i \in \Delta_0$ do: set $Z(r_i) = \delta_i$; and $\mathcal{R}Z^+ = \mathcal{R}Z^+ \cup \{r_i\}$.
3. While $\mathcal{R}Z^+ \neq \Delta$ do:
   (a) Let Ω be the set of models M, such that M falsifies rules only in $\mathcal{R}Z^+$, and verifies at least one rule outside of $\mathcal{R}Z^+$.
   (b) For each M compute:
   $$\kappa(M) = \max\{Z(r_i) : M \models \varphi_i \wedge \neg\psi_i\} + 1 \qquad (5)$$
   (c) Let M^* be the model in Ω with minimum κ; For each rule $r_i : \varphi_i \xrightarrow{\delta_i} \psi_i \notin \mathcal{R}Z^+$ that M^* verifies do:
   $$Z(r_i) = \kappa(M^*) + \delta_i \qquad (6)$$
   $\mathcal{R}Z^+ = \mathcal{R}Z^+ \cup \{r_i\}$.

End Procedure

Theorem 2 *The function Z computed by* Procedure Z_rank *complies with Def. 5, i.e. $Z = Z^+$.*

We turn our attention to the main results of this section, namely the minimality and uniqueness of κ^+:

Definition 6 *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.*

Definition 7 *An admissible ranking κ is said to be* compact *if, for every M' any ranking κ' satisfying*

$$\begin{array}{lcll} \kappa'(M) & = & \kappa(M) & M \neq M' \\ \kappa'(M) & < & \kappa(M) & M = M' \end{array}$$

is not *admissible.*

Theorem 3 (Main.) *Every consistent Δ has a* unique compact ranking *given by κ^+ (see Def. 5).*

Corollary 2 (Main.) *Every consistent Δ has a* unique minimal ranking *given by κ^+ (see Def. 5).*

As mentioned before, once the Z^+ ranking on rules is found, the κ^+ of any given model can be readily computed using Eq. 2. Moreover, the Z^+ ranking also provides effective means for deriving new conclusions from Δ: To test whether σ is a plausible conclusion of ϕ [3] we need to compare the minimal $\kappa^+(M^+)$ such that $M^+ \models \phi \wedge \sigma$, against the minimal $\kappa^+(M^-)$ such that $M^- \models \phi \wedge \neg\sigma$ (see Def. 2). Fortunately this minimization does not require an enumerative search on models; it can be systemized using the ordering imposed by Z^+. Let M be a *witness* for $\phi \mid\!\sim \sigma$ with respect to a set Δ', if $M \models \phi \wedge \sigma$ and M satisfies Δ'. We start by testing whether there is a witness M for $\phi \mid\!\sim \sigma$ with respect to the set Δ. If one is found, then $\kappa^+(M)$ must be 0: M does not violate any rule in Δ (see Def. 5). If no witness is found, we remove from Δ all rules r' such that $Z^+(r')$ is minimal, and call the remaining set Δ'. We start a new iteration by testing the existence of a witness M' (for $\phi \mid\!\sim \sigma$) this time with respect to Δ'. If M' is found, $\kappa^+(M')$ must be $Z^+(r') + 1$, since M' must violate a rule removed

hold for uniform databases, in which all rules are assigned $\delta = 0$ ([Pearl, 1990]). The consequence relation emerging from the unique preferred ranking of uniform databases was called 1-entailment in [Pearl, 1990] and was shown to be equivalent to Lehmann's [1989] rational closure ([Goldszmidt and Pearl, 1990]).

[3] i.e., whether $\phi \mid\!\sim \sigma$ is in the consequence relation defined by κ^+.

in the previous iteration. If no witness is found we remove the rules r'' with minimal $Z^+(r'')$ and so on. The question of whether $\phi \mid\!\sim \sigma$, $\phi \mid\!\sim \neg\sigma$ or neither is in the consequence relation defined by κ^+ is reduced to whether we find a witness for $\phi \mid\!\sim \sigma$, before we find a witness for $\phi \mid\!\sim \neg\sigma$, the completely symmetrical case, or whether these witnesses are found in the same iteration. The steps just described are formalized in Procedure Z^+-consequences below, where cases 3.(a)-3.(c) correspond to $\phi \mid\!\sim \sigma$, $\phi \mid\!\sim \neg\sigma$ or neither. Case 3.(d) selects the rules r with minimal $Z^+(r)$ and modifies the current set Δ for the next iteration (in case no witness is found). Note that each iteration (i.e., the test of whether a witness for $\phi \mid\!\sim \sigma$ with respect to some subset of Δ exists) involves a satisfiability test for Δ, and there can be at most $|\Delta|$ iterations before a witness is found[4]. Therefore, the complexity of Procedure Z^+-consequences is bounded by $|\Delta|$ propositional satisfiability tests in the worst case.

Procedure Z^+-consequences

Input: A consistent set Δ, the function Z^+ on Δ, and a pair of consistent formulas ϕ and σ. **Output:** answer YES/NO/AMBIGUOUS depending on whether $\phi \mid\!\sim \sigma$, $\phi \mid\!\sim \neg\sigma$ or neither.

1. TEST1 whether there is a model M such that $M \models \phi \wedge \sigma$ and M satisfies Δ.
2. TEST2 whether there is a model M such that $M \models \phi \wedge \neg\sigma$ and M satisfies Δ.
3. CASES indexed by the results from TEST1-TEST2:
   (a) IF YES-NO then return($\phi \mid\!\sim \sigma$)
   (b) IF NO-YES then return($\phi \mid\!\sim \neg\sigma$)
   (c) IF YES-YES then return(AMBIGUOUS)
   (d) IF NO-NO then let MIN_Z be the set of rules in Δ with minimum Z^+. Set $\Delta' = \Delta -$ MIN_Z. Set $\Delta = \Delta'$ and goto Step 1. [5]

End Procedure

It is natural to define the strength with which Δ endorses the validity of $\phi \mid\!\sim \sigma$ as the difference between the minimal $\kappa^+(M^+)$ such that $M^+ \models \phi \wedge \sigma$, and the minimal $\kappa^+(M^-)$ such that $M^- \models \phi \wedge \neg\sigma$. The Procedure Z^+-consequences can be easily modified to return this value, by simply computing the difference between the Z^+ levels at which each of the two witnesses are found.

4 Examples

The following examples illustrate properties of the κ^+-ranking and the use of δ to impose priorities among defaults. Example 1 shows how specificity-based preferences are established and maintained by the κ^+-ranking, freeing the rule-encoder from such considerations. A general formalization of this behavior is given in the next section (Theorem 4). In the second example, the priorities δ are used to establish preferences when specificity relations are not available. Example 3 constitutes a combination of the previous two.

Example 1: Specificity. Consider $\Delta_P = \{b \xrightarrow{\delta_1} f, p \xrightarrow{\delta_2} b, p \xrightarrow{\delta_3} \neg f\}$ which stands for r_1:"birds fly", r_2:"penguins are birds", and r_3:"penguins don't fly". The Z^+-ranking is computed as follows: Since r_1 is tolerated by Δ_P, $Z^+(r_1) = \delta_1$. Any κ^+-minimal model verifying r_2 and r_3 must violate r_1, therefore, following Procedure Z_rank, $Z^+(r_2) = \delta_1 + \delta_2 + 1$ and $Z^+(r_3) = \delta_1 + \delta_3 + 1$. According to Def. 2, in order to decide whether $p \wedge b \mid\!\sim \neg f$ ("penguin-birds don't fly") we must test whether "$\neg f$" is satisfied in all κ^+-minimal models of "$p \wedge b$" or, equivalently, whether $\kappa^+(p \wedge b \wedge \neg f) < \kappa^+(p \wedge b \wedge f)$. This test is performed mechanically by Procedure Z^+-consequences, yielding the expected conclusion: $p \wedge b \mid\!\sim \neg f$. The reason is as follows: Any model for "$p \wedge b$" will violate either r_1 ("birds fly") or r_3 ("penguins don't fly"). Since $Z^+(r_3) = Z^+(r_1) + \delta_3 + 1$, models violating r_1 (including those satisfying "$p \wedge b \wedge \neg f$") will have a lower κ^+-ranking and will thus be preferred to those violating r_3 (including those satisfying "$p \wedge b \wedge f$"); it follows that $\kappa^+(p \wedge b \wedge \neg f) < \kappa^+(p \wedge b \wedge f)$. Note that the preference of r_3 over r_1 is established independently of the initial priorities δ assigned to these rules.

Example 2: Belief strength. Consider a database containing two conflicting default rules: $\Delta_N = \{q \xrightarrow{\delta_1} p, r \xrightarrow{\delta_2} \neg p\}$, standing for r_1:"typically Quakers are pacifists", and r_2:"typically Republicans are not pacifists" (a version of the "Nixon-diamond"). Since each rule is tolerated by the other, the Z^+ of each rule is equal to its associated δ: $Z^+(r_1) = \delta_1$ and $Z^+(r_2) = \delta_2$. Given an individual, say Nixon, who is both a Republican and a Quaker, the decision of whether Nixon is a pacifist will depend on whether δ_1 is bigger, less or equal than δ_2. This is so because any model M_{rqp} for Quakers, Republicans and pacifists must violate r_2, and consequently $\kappa^+(M_{rqp}) = \delta_2$, while any model $M_{rq\neg p}$ for Quakers, Republicans and non-pacifists must violate r_1, i.e., $\kappa^+(M_{rq\neg p}) = \delta_1$. Note that in this case the decision to prefer one model over the other does not depend on specificity considerations but, rather, on whether the rule encoder believes that religious convictions bear more strength than political affiliations. This kind of preferences cannot be expressed in system-Z or in Lehmann's rational closure [1989].

Example 3: Combining priorities with specificity. For the final example, consider $\Delta_B = \{w \xrightarrow{\delta_1} b, w \wedge p \xrightarrow{\delta_2} \neg b\}$ encoding the information that r_1:"if it is Wednesday night I play basketball", and r_2:"if it is Wednesday night and I have a paper due, I don't play basketball" with the δ's reflecting the degree of firmness of these rules. Suppose we wish to inquire

[4] In each iteration the size of Δ decreases by at least one since at least one rule is removed.

[5] Note that since we are requiring that both ϕ and σ be consistent, Δ' cannot be empty.

whether "I'll play basketball on a Wednesday night when a paper is due" ($w \wedge p \mid\sim b$). On one hand, the answer to such query is explicitly contained in r_2. On the other hand, r_2 conflicts with r_1, and in many formalisms (e.g. circumscription [McCarthy, 1986], default logic [Reiter, 1980]) such a conflict would require extra information in order to give an unambiguous answer (the relation between the *ab* predicates associated with the defaults for circumscription, and some preference criteria among extensions or the use of seminormal defaults for default logic). System-Z^+ yields the expected result regardless of how strongly one believes in r_1. According to Procedure Z_rank the Z^+-ranking computes to $Z^+(r_1) = \delta_1$ and $Z^+(r_2) = \delta_1 + \delta_2 + 1$. Any model for $w \wedge p$ must violate either r_1 or r_2, and since $Z^+(r_2) = Z^+(r_1) + \delta_2 + 1$, the model violating r_1 will be preferred. Thus, $\kappa^+(w \wedge p \wedge b) < \kappa^+(w \wedge p \wedge \neg b)$ and we conclude (with firmness $\delta_2 + 1$) that "I won't be playing basketball that night". Now suppose Δ_B is part of a bigger database and that we wish to include information that takes precedence over all other so far mentioned. For example, "If I am sick I stay in bed" (assuming of course that "staying in bed" rules out "playing basketball" or other activities). In order to enforce this precedence we need only give to this new rule a sufficiently high δ without considering its relation to previous commitments.

5 Discussion

System-Z^+ provides the user with the power to explicitly set priorities among default rules, and simultaneously maintains a proper account for specificity relations. However, it inherits some of the deficiencies of system-Z (and the rational closure described in [Lehmann, 1989]) the main one being the inability to sanction inheritance across exceptional subclasses. For example if a fourth rule $b \xrightarrow{\delta_4} l$ ("birds have legs") is added to Δ_P (Example 1), we would normally conclude that "penguins have legs". However, system-Z will consider "penguins" exceptional "birds", (since they do not fly) with respect to *all* properties, including "having legs". The κ^+-ranking allows the rule author to partially bypass this obstacle by means of the δ's associated with the rules. If δ_4 is set to be bigger than δ_1 (to express perhaps the intuition that anatomic properties are more typical than developmental facilities) then the system will conclude that "typically penguins have legs"[6]. This solution however, is not entirely satisfactory. If we add to this new set of rules a class of "birds" which are "legless", system-Z^+ will conclude that either "penguins have legs" or "legless birds fly" but not both[7]. To overcome this difficulty *non-layered* priorities among rules must be imposed (see [Geffner, 1989], [Grosof, 1991]).

In a similar vein we remark that more refined selection functions (than the maximum rule violated) might be needed for certain domains. For example in circuit diagnosis the ranking of a given explanation should also reflect the *number* of faults it predicts, not merely the abnormality of the least likely fault or, better yet, the sum of the faults weighted by their abnormality ranking. A refinement such as the one proposed by maximum entropy approach ([Goldszmidt *et al.*, 1990]) embodies this cost function and may yield better results in such domains.

In some sense system-Z^+ can be viewed as a version of prioritized circumscription [Lifschitz, 1988], where default priorities are induced by means of partial order imposed on the abnormalities in the minimization process. However, in prioritized circumscription the relative ranking of abnormalities remains fixed at the level furnished by the user, and does not reflect interactions between adjacents rules. In system-Z^+ the input priorities undergo adjustments so as to take into account all related rules in the system. For example, in the database Δ_P above, the ranking of r_3 ("typically penguins do not fly") was adjusted from δ_3 to $\delta_1 + \delta_3 + 1$, so as to supercede δ_1, the priority of the conflicting rule "typically birds fly". As a result of such adjustments, the consistency of the rankings is maintained throughout the system, and compliance with specificity-type constrains is automatically preserved. This is made precise in the following theorem:

Theorem 4 *Let* $r1 : \varphi \xrightarrow{\delta_1} \psi$ *and* $r2 : \phi \xrightarrow{\delta_2} \sigma$ *be two rules in a consistent set* Δ *such that:*

1. $\varphi \mid\sim \phi$ *is in all consequence relations of admissible* κ*-rankings (i.e.,* φ *is more specific than* ϕ*).*
2. *There is no model satisfying* $\varphi \wedge \psi \wedge \phi \wedge \sigma$ *(i.e.,* r_1 *conflicts with* r_2*).*

then $Z^+(r1) > Z^+(r2)$ *independently of the values of* δ_1 *and* δ_2

In other words, the Z^+-ranking guarantees that features of more specific contexts override conflicting features of less specific contexts.

Note that although the computation of the adjusted ranking may be expensive (non-polynomial in the number of rules), once it is found, it constitutes an efficient encoding of κ^+ and facilitates an efficient procedure for answering queries about the plausibility of proposed conclusions: deciding whether $\phi \mid\sim \sigma$ requires only $O(|\Delta|)$ propositional satisfiability tests.

Acknowledgements

Sound criticisms by two anonymous reviewers helped in putting some of the results in perspective.

[6] Note that the fact that "penguins" are only exceptional with respect "flying" (and not necessarily with respect to "having legs") is automatically encoded in the Z^+ ranking by forcing $Z^+(r_3)$ to exceed $Z^+(r_1) + \delta_3$ independently of δ_4 (and $Z^+(r_4)$).

[7] This counterexample is due to Kurt Konolige.

A Appendix: Theorems and Proofs

Theorem 1 *A set Δ is consistent iff there exists a tolerated rule in every nonempty subset of Δ.*

Proof: We first show that if there exists a tolerated rule in every nonempty subset of Δ we can always produce an admissible ranking κ. Under the stated condition, we can construct the following ordered partition $(\Delta_0, \Delta_1, \ldots, \Delta_n)$ of Δ: Rules in Δ_0 are tolerated by Δ, rules in Δ_1 are tolerated by $\Delta - \Delta_0$ and so on. By Def. 3, for each one of these Δ_j, there must exist a nonempty subset Ω_j of Ω (the set of all possible models), such that for each rule $r_j \in \Delta_j$ there must exists a model $M_j \in \Omega_j$, where M_j verifies r_j and M_j satisfies Δ if $j = 0$ and $\Delta - \{\Delta_0 \cup \ldots \cup \Delta_{j-1}\}$ otherwise. Thus, using these models (the models actually required to effectively build the partition of Δ), we define a partition $(\Omega_0, \Omega_1, \ldots, \Omega_n, \Omega_{n+1})$ of Ω, where each Ω_j contains models with the characteristics mentioned above, and Ω_{n+1} contains the models necessary to complete the partition. Let δ_i^* denote the highest δ among rules in set Δ_i. We now build, in a recursive fashion, an admissible ranking κ based on these two partitions in the following manner: If $M_0 \in \Omega_0$, set $\kappa(M_0) = 0$. Else if $M_j \in \Omega_j$, set $\kappa(M_j) = \kappa(M_{j-1}) + \delta_{j-1}^* + 1$. Note that each model $M_j \in \Omega_j$ is a characteristic model[8] of the rule $r_j \in \Delta_j$ it verifies, and the κ-minimal model falsifying any rule $r_j \in \Delta_j$ must belong to the set Ω_{j+1}. Thus, in order to guarantee the admissibility of κ, it is enough to show that for an arbitrary pair of models $M_j \in \Omega_j$ and $M_{j+1} \in \Omega_{j+1}$ the following relation holds:

$$\kappa(M_j) + \delta_j < \kappa(M_{j+1}) \tag{7}$$

where δ_j can be any δ among the rules in Δ_j. But this relation is guaranteed by the construction of κ since $\kappa(M_j) + \delta_j^* + 1 = \kappa(M_{j+1})$, where δ_j^* is the highest δ among the rules in Δ_j. Therefore κ is admissible.

To show the converse we reason by contradiction: Assume that there is no tolerated rule in $\Delta' \subseteq \Delta$ and there is an admissible ranking κ' for Δ. Since there is no tolerated rule in Δ', we know that any characteristic model M_1 for rule $r_1 \in \Delta'$ must falsify another rule $r_2 \in \Delta'$. By the admissibility of κ' the following must hold

$$\kappa'(M_2) + \delta_2 < \kappa'(M_1) \tag{8}$$

where M_2 is a characteristic model for r_2. By the same token, M_2 must falsify another rule in Δ', say r_3, and we can insert $\kappa'(M_3)$ [9] in the chain of Eq. 8:

$$\kappa'(M_3) + \delta_3 < \kappa'(M_2) + \delta_2 < \kappa'(M_1) \tag{9}$$

We can continue to expand the chain in this fashion and get,

$$\kappa'(M_n) + \delta_n < \kappa'(M_{n-1}) + \delta_{n-1} < \ldots < \kappa'(M_2) + \delta_2 < \kappa'(M_1) \tag{10}$$

Note that if at any point in the construction of this chain, a model falsifies a rule that has a characteristic model in the chain, we arrive at a contradiction since by the admissibility of κ', $\kappa'(M') + \delta' < \kappa'(M'')$ but since both M' and M'' are characteristic models of the same rule it must be that $\kappa'(M') = \kappa'(M'')$. Moreover, given that Δ' is finite we are bound to encounter such contradiction. □

Corollary 1 *Deciding the consistency of a set Δ requires at most $|\Delta|^2$ propositional satisfiability tests.*

Proof: A procedure for deciding consistency would only need to construct the partition $(\Delta_0, \Delta_1, \ldots, \Delta_n)$, since if it succeeds, we know from the previous theorem that it is possible to build an admissible ranking. On the other hand, if it fails, there is a nonempty subset of Δ with no tolerated sentence. Identifying Δ_0 takes at most $|\Delta|$ satisfiability tests, identifying Δ_1 takes at most $(|\Delta| - |\Delta_0|)$ satisfiability tests, and so on. Thus, overall, it will take $[|\Delta| + |\Delta| - |\Delta_0| + |\Delta| - |\Delta_0| - |\Delta_1| \ldots]$ satisfiability steps which is bounded by $|\Delta|^2$. □

Proposition 1 *The ranking function κ^+ is admissible.*

Proof: Given that $Z^+(r_i) = \min\{\kappa^+(M) : M \models \phi_i \wedge \psi_i\} + \delta_i$, we can re-write the conditions for admissibility (Eq. 1) as

$$Z^+(r_i) < \min\{\kappa^+(M) : M \models \varphi_i \wedge \neg\psi_i\} \tag{11}$$

Since $\kappa^+(M) = \max\{Z^+(r_i) : M \models \varphi_i \wedge \neg\psi_i\} + 1$, it follows that κ^+ is admissible. □

Theorem 2 *The function Z computed by* Procedure Z_rank *complies with Def. 5, i.e. $Z = Z^+$.*

Proof: We first show that the relevant steps in Procedure Z_rank are well defined. By the assumption that Δ is consistent, Δ_0 cannot be an empty set (steps 1 and 2): There must be at least one rule tolerated by Δ. By similar reasons, Ω cannot be empty in each iteration of the loop in step 3. By consistency we must be able to find a tolerated sentence in each nonempty subset of Δ. Finally, in the computation of Eq. 5, since M only falsifies rules in $\mathcal{R}Z^+$, all Z for these rules are available.

We now show that $Z = Z^+$ for rules $r_0 \in \Delta_0$. Since each r_0 is tolerated by Δ, there must be a model M_0 (for each one of these rules), such that M_0 verifies r_0 and M_0 satisfies Δ. Thus, each one of these models does not falsify any rules in Δ, and $\kappa^+(M_0) = 0$. According to Eq. 3 in Def. 5, $Z^+(r_0) = \delta_0$ for those rules and that is precisely what is computed in step 2.

The proof proceeds by induction on the iterations of loop 3, where we show that for every rule $r \in \mathcal{R}Z^+$, $Z(r) = Z^+(r)$ holds. For the basis of the induction consider the first iteration: Since $\mathcal{R}Z^+ = \Delta_0$, then for every $r_0 \in \Delta_0$, $Z(r_0) = Z^+(r_0)$ holds as shown above. Our objective is to show that this equality holds for the rules inserted into $\mathcal{R}Z^+$ at step 3.(c). We need the following preliminary result: for any model M' verifying a rule outside $\mathcal{R}Z^+$, either $\kappa^+(M')$ is set by

---

[8] Recall that a model M^+ is said to be a *characteristic model* for rule $\varphi \rightarrow \psi$ relative to ranking κ, if $\kappa(M^+) = \min\{\kappa(M) : M \models \varphi \wedge \psi\}$.

[9] M_3 is a characteristic model for r_3.

the $\kappa^+(M^+)$ of a model M^+ which is a characteristic model of a rule in $\mathcal{RZ}^+$, or it depends on another model which complies with these characteristics. Consider an arbitrary model M' verifying a rule outside $\mathcal{RZ}^+$. By Eqs. 2, 3 and 4, there must exist a model M'' such that $\kappa^+(M') = \kappa^+(M'') + \delta'' + 1$, and M'' is the characteristic model of some rule r''. If M'' does not falsify any rule, we are done, since $r'' \in \mathcal{RZ}^+$. Otherwise there must exists a model M''' such that $\kappa^+(M'') = \kappa^+(M''') + \delta''' + 1$, and M''' is the characteristic model of rule r'''. If M''' does not falsify any rule, $r''' \in \Delta_0$ and we've found that $M''' = M^+$. Note that since Δ is consistent and finite, we are guarantee that this process will stop with the desired M^+. What we have just shown is that for any model $M' \notin \Omega$ verifying a rule outside $\mathcal{RZ}^+$, there is a model $M'' \in \Omega$ such that $\kappa^+(M'') \leq \kappa^+(M')$. Since models in Ω falsify rules only in $\mathcal{RZ}^+$, $\kappa(M) = \kappa^+(M)$ for these models as computed by Eq. 5 in step 3.(b), and given that $\kappa^+(M^*)$ is the model with minimal κ^+ among those verifying rules outside of $\mathcal{RZ}^+$, M^* is a characteristic model for those rules r_i with respect to κ^+. We can then re-write Eq. 6 as

$$Z(r_i) = \min\{\kappa^+(M) : M \models \varphi_i \wedge \psi_i\} + \delta_i \qquad (12)$$

and for these rules inserted in $\mathcal{RZ}^+$ (step 3.(c)) $Z(r_i) = Z^+(r_i)$ holds. For the induction step consider the n^{th} iteration. By the induction hypothesis for all rules $r_j \in \mathcal{RZ}^+$, $Z(r_j) = Z^+(r_j)$, and therefore by Eq. 5 in step 3.(b) $\kappa(M) = \kappa^+(M)$ for all $M \in \Omega$. The claim is that M^* is a characteristic model for the rules r^* that it verifies (outside of $\mathcal{RZ}^+$). The arguments are essentially the same as in the basis: the κ^+ of any model verifying a rule outside $\mathcal{RZ}^+$ must depend on the κ^+ of some model in Ω. Thus, for rules r_i inserted in $\mathcal{RZ}^+$ during the n^th iteration, Eq. 6 can be re-written as Eq. 12 and $Z(r_i) = Z^+(r_i)$ holds. □

Lemma 1 *The ranking κ^+ is compact.*

Proof: By contradiction. Assume it is possible to lower $\kappa^+(M')$ of some model M', where $\kappa^+(M') > 0$. From the definition of κ^+ (Def. 5, there must be a rule $r : \varphi \xrightarrow{\delta} \psi$ such that $\kappa^+(M') = Z^+(r) + 1$ (see Eq. 2), which implies that

$$\kappa^+(M') = \min\{\kappa^+(M) : M \models \varphi \wedge \psi\} + \delta + 1 \quad (13)$$

Lowering the value of $\kappa^+(M')$ will violate Eq. 13 which will imply the violation of Eq. 1 for rule r. □

Theorem 3 (Main.) ***Every consistent Δ has a* unique compact ranking *given by κ^+ (see Def. 5).***

Proof: By Lemma 1, κ^+ is compact. We show it is also unique. Suppose there exists some other compact ranking κ that differs from κ^+ in at least one model. We will show that if there exists an M' such that $\kappa(M') < \kappa^+(M')$ then κ cannot be admissible, where if $\kappa(M') > \kappa^+(M')$, then κ cannot be compact. Assume $\kappa(M') < \kappa^+(M')$, let I be the lowest κ value for which such inequality holds, and let $\kappa^+(M') = J > I$. By the definition of κ^+ (Def. 5), we know that there is a rule $r : \varphi \xrightarrow{\delta} \psi$ such that Eq. 13 holds, and as a consequence

$$\min\{\kappa^+(M) : M \models \varphi \wedge \psi\} = J - \delta - 1 \qquad (14)$$

Since κ is assumed to be admissible, the following must hold for rule r

$$\kappa(M') \geq \min\{\kappa(M) : M \models \varphi \wedge \psi\} + \delta + 1 \qquad (15)$$

Since $J > \kappa(M')$,

$$J > \min\{\kappa(M) : M \models \varphi \wedge \psi\} + \delta + 1 \qquad (16)$$

If we subtract $\delta + 1$ from both sides of this inequality and use Eq. 14 we get

$$\min\{\kappa^+(M) : M \models \varphi \wedge \psi\} > \min\{\kappa(M) : M \models \varphi \wedge \psi\} \qquad (17)$$

But this cannot be since I was assumed to be the minimal value of κ for which this inequality can occur, and if $\min\{\kappa(M) : M \models \varphi \wedge \psi\} > I$, then κ is not admissible (see Eq. 15).

Now assume that there is a non-empty set of models for which $\kappa(M) > \kappa^+(M)$, and let I be the lowest κ^+ value in which $\kappa(M') > \kappa^+(M')$ for some model M'. We will show that κ cannot be compact, since it will be possible to reduce $\kappa(M')$ to $\kappa^+(M')$ while keeping constant the κ of all other models. From $\kappa^+(M') = I$ we know that M' does not falsify any rule r with Z^+ rank higher than $I - 1$. Hence, we only need to watch whether the reduction of κ can violate rules r for which $Z^+(r) < I$. For every such rule there exists a model M, such that M verifies r and $\kappa^+(M) < I$. Since for all these models κ is assumed to be equal to κ^+ it follows that none of these models can be violated by reducing $\kappa(M')$ to $\kappa^+(M')$. □

Theorem 4 *Let $r1 : \varphi \xrightarrow{\delta_1} \psi$ and $r2 : \phi \xrightarrow{\delta_2} \sigma$ be two rules in a consistent set Δ such that:*

1. *$\varphi \mid\!\sim \phi$ is in all consequence relations of admissible κ-rankings (i.e., φ is more specific than ϕ).*
2. *There is no model satisfying $\varphi \wedge \psi \wedge \phi \wedge \sigma$ (i.e., r_1 conflicts with r_2).*

then $Z^+(r1) > Z^+(r2)$ independently of the values of δ_1 and δ_2

Proof: If $\varphi \mid\!\sim \phi$ is in every consequence relation of every κ admissible with Δ then (by Def. 2) the following constraint must hold in all these κ-rankings (including κ^+):

$$\kappa(\varphi \wedge \phi) < \kappa(\varphi \wedge \neg\phi) \qquad (18)$$

Thus, any characteristic model $M^+_{r_1}$ for r_1 must render ϕ (the antecedent for r_2) true, and since there is no model such that both rules are verified (condition 2 in the theorem above), all $M^+_{r_1}$ must also falsify r_2. From Def. 5 (Eqs. 2 and 3): $\kappa^+(M^+_{r_1}) \geq Z^+(r_2) + 1$, and $Z^+(r_1) = \kappa^+(M^+_{r_1}) + \delta_2$. It follows that $Z^+(r_1) > Z^+(r_2)$. Note that the characteristic model for r_2 cannot in turn falsify r_1 since this will preclude the existence of an admissible ranking κ and Δ was assumed to be consistent. □

References

[Geffner, 1989] Hector A. Geffner. Default reasoning: Causal and conditional theories. Technical Report TR-137, PhD dissertation, University of California Los Angeles, Cognitive Systems Lab., Los Angeles, 1989.

[Goldszmidt and Pearl, 1990] Moisés Goldszmidt and Judea Pearl. On the relation between rational closure and system Z. In *Third International Workshop on Nonmonotonic Reasoning*, pages 130–140, South Lake Tahoe, 1990.

[Goldszmidt *et al.*, 1990] Moisés Goldszmidt, Paul Morris, and Judea Pearl. A maximum entropy approach to nonmonotonic reasoning. In *Proceedings of American Association for Artificial Intelligence Conference*, pages 646–652, Boston, 1990.

[Grosof, 1991] Benjamin N. Grosof. Generalizing prioritation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, Boston, 1991.

[Kraus *et al.*, 1990] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Lehmann, 1989] Daniel Lehmann. What does a conditional knowledge base entail? In *Proceedings of Principles of Knowledge Representation and Reasoning*, pages 212–222, Toronto, 1989.

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

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

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

[Pearl, 1989] Judea Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proceedings of Principles of Knowledge Representation and Reasoning*, pages 505–516, Toronto, 1989.

[Pearl, 1990] Judea Pearl. System Z: A natural ordering of defaults with tractable applications to default reasoning. In M. Vardi, editor, *Proceedings of Theoretical Aspects of Reasoning about Knowledge*, pages 121–135. Morgan Kaufmann, San Mateo, 1990.

[Reiter, 1980] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Shoham, 1987] Yoav Shoham. Nonmonotonic logics: Meaning and utility. In *Proceedings of International Conference on AI (IJCAI-87)*, pages 388–393, Milan, Italy, 1987.

[Spohn, 1987] W. Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, pages 105–134. Dordrecht: Reidel, Holland, 1987.