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

WHERE DO DEFAULT PRIORITIES COME FROM?

Moises Goldszmidt
Judea Pearl

July 1991
CSD-910057

TECHNICAL REPORT
R-154
November, 1990

Where Do Default Priorities Come From?*

- Extended Abstract -

Moisés Goldszmidt
moises@cs.ucla.edu

Judea Pearl
judea@cs.ucla.edu

Cognitive Systems Lab.
Dept. of Computer Science, UCLA
Los Angeles, CA 90024

1 Background and motivation

The notion of priorities among defeasible sentences has proven necessary for resolving conflicting defeasible arguments. Proposals for incorporating these priorities fall into one of two main approaches: the first, exemplified by prioritized circumscription ([McCarthy, 1986, Lifschitz, 1988]) and non-normal defaults ([Etherington and Reiter, 1983]), rely on the user to input these priorities. The second, exemplified by ε-semantics ([Pearl, 1988]) and the preferential logic of [Kraus *et al.*, 1990], derive some of these priorities from the knowledge base. This advantage of the latter approach stems from attributing a conditional reading to defeasible sentences: a default $\varphi \rightarrow \psi$ stands for "in the most normal situations where φ holds, ψ should be expected to be true". This interpretation imposes ordering constraints on worlds (or states of affairs) which are immediately translated into priorities on defaults. However, these conditional-based approaches are too "conservative": desirable conclusions are often quenched by irrelevant information (e.g., "typically red birds fly" will not follow from "typically birds fly").

Formalisms aimed at solving this deficiency have been proposed, based on restricting the set of admissible models with respect to which conclusions are to be entailed. These include the rational closure of [Lehmann, 1989], 1-entailment in system-Z [Pearl, 1990], and the maximum entropy approach of [Goldszmidt *et al.*, 1990]. These systems, while extending the power of ε-semantics and preferential logic (e.g., sanctioning transitivity or contraposition),

*This work was supported in part by National Science Foundation grant #IRI-88-21444 and Naval Research Laboratory grant #N00014-89-J-2007. The first author is supported by an IBM graduate fellowship 1990-91.

are also prone to yield undesired conclusions in some cases, and are over-conservative in others (see [Goldszmidt and Pearl, 1990b] for limitations of rational closure and 1-entailment). As it is argued in [Geffner, 1989], part of the blame for these weaknesses is the commitment to a structure in which the worlds are ranked in a total order. In [Geffner, 1989], a partial order on possible worlds is proposed based on examining the priorities of defaults violated in these worlds. The priorities themselves are obtained from argument-based considerations; a single default always gets a higher priority over at least one default in any conflicting argument. This system achieves the power of prioritized circumscription (in the propositional case) with the added advantage that priorities are automatically derived from the knowledge base (see [Geffner, 1989] for details).

This paper shows that the basic scheme of coupling priorities on defaults with preferences on worlds follows naturally from two probabilistic principles: First, each default imposes a constraint on the ranking of worlds[1], forcing worlds that violate the default to attain a higher rank (lower probability) than those that confirm the default. Second, the rank of each world is equal to the sum of the priorities of those defaults violated in that world. The summation originates with the principle of maximum entropy as shown in [Goldszmidt *et al.*, 1990]. These principles induce two partial orders: One on worlds (called *preference*) and one on defaults (called *dominance*). We show that the plausible conclusions of a given theory, (i.e., those that attain higher likelihood than their denial), are precisely those formulas that are true in all preferred worlds. We believe this formulation provides a probabilistic basis for both conditional entailment (cd-entailment [Geffner, 1989]) and prioritized circumscription ([Lifschitz, 1988]).

2 Formal description of the proposed system

The basic language will be a closed set $\mathcal{L}$ of well formed propositional formulas built in the usual way from a *finite* set of propositional variables and the connectives "$\neg$" and "$\supset$" (where "$\supset$" denotes material implication). We will use ω to denote truth assignments to the propositional variables in $\mathcal{L}$. The satisfaction of a formula φ of $\mathcal{L}$ by a truth assignment ω is defined in the usual way and written $\omega \models \varphi$.

We extend this language with a new binary connective "$\rightarrow$" to construct conditional sentences of the form $\varphi \rightarrow \psi$, where φ and ψ (the antecedent and consequent of the conditional sentence) are formulas from $\mathcal{L}$. A conditional sentence $\varphi \rightarrow \psi$ is *verified* by a truth assignment ω if $\omega \models \varphi \wedge \psi$. It is *falsified* by ω if $\omega \models \varphi \wedge \neg\psi$ and it is *satisfied* if $\omega \models \varphi \supset \psi$.

Definition 1 (Admissible structures.) *A structure $\mathcal{S}$ for a set Δ of defeasible sentences is a triple $\langle \Omega, \kappa, \delta \rangle$, where Ω is the set of truth assignments (or possible worlds ω) to the variables in $\mathcal{L}$, δ is a function from the set Δ of sentences to the positive set of natural*

[1]Given that the basic language of our formalism will consist of standard propositional logic, the terms "possible worlds", "interpretations" and "truth assignments" will be regarded as equivalent.

numbers, and κ is a ranking from Ω to the set of natural numbers. A structure is said to be admissible with respect to Δ, *written S_Δ, iff δ and κ satisfy the following two conditions:*

$$\kappa(\omega) = \begin{cases} \sum_{\omega \models \varphi_i \wedge \neg\psi_i} \delta(\phi_i \rightarrow \psi_i) \\ 0 \;\; otherwise \end{cases} \tag{1}$$

$$\min\{\kappa(\omega) : \omega \models \varphi_i \wedge \psi_i\} < \min\{\kappa(\omega) : \omega \models \varphi_i \wedge \neg\psi_i\} \;\; \text{for every} \;\; \phi_i \rightarrow \psi_i \in \Delta \tag{2}$$

Eq. 2 enforces the conditional reading of defeasible sentences as constraints on the ranking of worlds. If we interpret the κ-ranking of a world ω as the degree of surprise of finding such world, then each default imposes the constraint that the minimally ranked world which falsifies the default should be more surprising than the minimally ranked worlds verifying the default. The priority-function δ on defaults can be interpreted as the "cost" of falsifying these defaults (see Eq. 1).

Definition 2 (Consistency.) *A set Δ is said to be* consistent *iff there exists at least one admissible structure for Δ.*

The next theorem establishes the conditions for consistency. Note that these conditions are the same as the ones required for probabilistic consistency ([Adams, 1975]).

Theorem 1 *Let $\varphi \rightarrow \psi$ be* tolerated *by Δ if there exists a truth assignment which verifies $\varphi \rightarrow \psi$ and satisfies Δ. Then Δ is* consistent *iff there exists one tolerated sentence in each nonempty subset of Δ.*

It follows that the effective procedure developed in [Goldszmidt and Pearl, 1990a] is sufficient for deciding consistency as defined above.

Let Δ be a set of defaults representing a knowledge base, and let Γ be a set of formulas describing facts about a specific situation. It is natural to proclaim a formula ψ a *plausible conclusion* of $\langle\Delta, \Gamma\rangle$ if, in every admissible structure for Δ, ψ is true in all minimally ranked worlds satisfying Γ. Formally:

Definition 3 (Plausible conclusion) *A formula ψ is a plausible conclusion of Δ from a set of premisses Γ, written $\langle\Delta, \Gamma\rangle \mid\!\sim \psi$, iff*

$$\min\{\kappa(\omega) : \omega \models \Gamma \wedge \psi\} < \min\{\kappa(\omega) : \omega \models \Gamma \wedge \neg\psi\} \tag{3}$$

holds in all S_Δ.

Definition 4 (Sentence-dominance) $\varphi \rightarrow \psi$ dominates $\varphi' \rightarrow \psi'$ *in the context of Δ, written $\varphi \rightarrow \psi \succ_\Delta \varphi' \rightarrow \psi'$, iff $\delta(\varphi \rightarrow \psi) > \delta(\varphi' \rightarrow \psi')$ holds in every S_Δ.*

Definition 5 (World-preference) *Let $\mathcal{F}[\omega]$ be the set of sentences falsified by ω. A world ω is* preferred *to ω' in the context of Δ, written $\omega \sqsubset_\Delta \omega'$, iff $\mathcal{F}[\omega] \neq \mathcal{F}[\omega']$ and for every sentence $\varphi \rightarrow \psi \in \mathcal{F}[\omega] - \mathcal{F}[\omega']$ there is a sentence $\varphi' \rightarrow \psi' \in \mathcal{F}[\omega'] - \mathcal{F}[\omega]$ such that $\varphi' \rightarrow \psi' \succ_\Delta \varphi \rightarrow \psi$.*

Theorem 2 $\langle \Delta, \Gamma \rangle \mid\!\sim \psi$ iff *ψ is true in all the most preferred worlds (according to $\sqsubset_\Delta$) satisfying Γ.*

3 Remarks

We have formalized a proposal for defeasible reasoning centered on priorities among the defeasible sentences. We have shown that these priorities emerge naturally from a conditional reading of the defeasible information, plus the intuitive notion that the ranking of a world should be proportional to a weighted account of the defaults violated in that world. The resulting formalism extends the power of previous attempts including ε-semantics, preferential logic, rational closure, 1-entailment and maximum entropy, and seems to behave identically to cd-entailment ([Geffner, 1989]). Its power stems from allowing defaults to attain different "degrees of strength" in the context of ε-semantics and maximum entropy. Thus, instead of interpreting every default $\varphi \rightarrow \psi$ as $P(\psi|\varphi) \geq 1 - \varepsilon$, we now write $P(\psi|\varphi) \geq 1 - \varepsilon^\eta$, $\eta \geq 1$, permitting the exponent of ε to vary from default to default. Entailment is decided with respect to all possible maximum entropy probability distributions.

We believe this proposal promises to unify three different formalisms: probabilistic interpretation (augmented with maximum entropy), conditional entailment and prioritized circumscription.

References

[Adams, 1975] Ernest W. Adams. *The Logic of Conditionals.* D.Reidel, Dordrecht, 1975.

[Etherington and Reiter, 1983] D. Etherington and R. Reiter. On inheritance hierarchies with exceptions. In *Proceedings of American Association for Artificial Intelligence Conference*, pages 104–108, Washington D.C., 1983.

[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, 1990a] Moisés Goldszmidt and Judea Pearl. On the consistency of defeasible databases. Technical Report TR-122, University of California Los Angeles, Cognitive Systems Lab., Los Angeles, 1990.

[Goldszmidt and Pearl, 1990b] 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.

[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, 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.