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

EPSILON-SEMANTICS

Judea Pearl

July 1991
CSD-910029

Written for Encyclopedia of Artificial Intelligence, (2nd Ed).

TECHNICAL REPORT
R-159
January, 1991

Epsilon-Semantics

Judea Pearl
Computer Science Department
Cognitive Systems Laboratory
University of California Los Angeles, California, 90024

1. Introduction: Infinitesimal Probabilities

Epsilon-Semantics (ε-semantics, for short) is a formal framework for belief revision in which belief statements are interpreted as statements of high probability, infinitesimally close to one, and where belief revision takes place by conditioning current beliefs on newly available evidence. The conditionalization of extreme probabilities yields non-numeric belief revision, as if propositions were assigned qualitative truth values from the set {TRUE, BELIEVED, POSSIBLE, DISBELIEVED, FALSE}. The basic idea of ε-semantics can be traced back to the conditional logic (qv) of Adams (1966, 1975) and the Ordinal Condition Functions (OCF) of Spohn (1988). Potential applications in nonmonotonic logic and default reasoning (qv) were noted in McCarthy (1986) and Pearl (1988) and further developed by Lehmann (1989), Geffner (1989), Pearl (1990) and Goldszmidt et al (1990).

A simple way of viewing ε-semantics is to consider an ordinary probability function P defined over a set W of possible worlds (or states of the world), and to image that the probability $P(w)$ assigned to each world w is a polynomial function of some small positive parameter ε, for example, α, $\beta\varepsilon$, $\gamma\varepsilon^2$, ..., etc. Accordingly, the probabilities assigned to any subset A of W, as well as all conditional probabilities $P(A|B)$, will be rational functions of ε. Now define the ranking function [1] $\kappa(A|B)$ as the power of the most significant term in the expansion of $P(A|B)$ into a power series of ε,

$$\kappa(A|B) = \text{lowest } n \text{ such that } \lim_{\varepsilon \to 0} P(A|B)/\varepsilon^n \text{ is non-zero} . \tag{1}$$

In other words, $\kappa(A|B) = n$ iff $P(A|B)$ is of the same order of magnitude as ε^n, or equivalently, $\kappa(A|B)$ is of the same order-of-magnitude as $[P(A|B)]^{-1}$.

If we think of n for which $P(w) = \alpha\, \varepsilon^n$ as measuring the degree to which the world w is disbelieved (or the degree of surprise were we to observe w), then $\kappa(A|B)$ can be thought of as the degree of disbelief (or surprise) in A, given that B is true. Parameterizing a probability measure by ε, and extracting the lowest exponent of ε as the measure of (dis)belief is a way of capturing the process by which people abstract qualitative beliefs from numerical probabilities and accept them as tentative truths, until rejected by future evidence. For other formalizations of belief acceptance see [Kyburg 1961, Pearl 1987b].

It is easy to verify (see Spohn 1987) that κ satisfies the following properties:

1. $\kappa(A) = \min \{\kappa(w) \mid w \in A\}$

2. $\kappa(A) = 0$ or $\kappa(\neg A) = 0$, or both

3. $\kappa(A \cup B) = \min\{\kappa(A), \kappa(B)\}$

4. $\kappa(A \cap B) = \kappa(A|B) + \kappa(B)$

---

(1) Spohn (1987) called this function "non-probabilistic" Ordinal Conditional Function (OCF).

These reflect (on a logarithmic scale) the usual properties of probability functions, with *min* replacing addition, and addition replacing multiplication:

1'. $P(A) = \sum_{w \in A} P(w)$

2'. $P(A) + P(-A) = 1$

3'. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

4'. $P(A \cap B) = P(A|B)\,P(B)$.

The result is a probabilistically sound calculus, employing integer addition, for manipulating order-of-magnitudes of disbeliefs. For example, if we make the following correspondence between linguistic quantifiers and ε^n :

$P(A) = \varepsilon^0$	A is believable	$\kappa(A) = 0$
$P(A) = \varepsilon^1$	A is unlikely	$\kappa(A) = 1$
$P(A) = \varepsilon^2$	A is very unlikely	$\kappa(A) = 2$
$P(A) = \varepsilon^3$ ----	A is extremely unlikely ------------------------	$\kappa(A) = 3$ ----

then the infenitesimal approximation yields a nonmonotonic logic to reason about likelihood. It takes sentences in the form of quantified conditional sentences, e.g., "Birds are likely to fly", (written $\kappa(\neg f \mid b) = 1$), "Penguins are most likely birds", (written $\kappa(\neg\, b \mid p) = 2$), "Penguins are extremely unlikely to fly," (written $\kappa\,(f \mid p) = 3$) and returns quantified conclusions in the form of "If x is a penguin-bird then x is extremely unlikely to fly" (written $\kappa\,(f \mid p \wedge b) = 3$).

The basic κ ranking system, as described in Spohn (1987) requires the specification of a complete probability model before reasoning can commence. In other words, the knowledge base must be sufficiently rich to define the κ associated with every world w. In practice, such specification might require knowledge that is not readily available in common discourse. For example, we might be given the information that birds fly (written $\kappa(\neg f \mid b) = 1$) and no information at all about properties of non-birds, thus leaving $\kappa(f \wedge \neg b)$ unspecified. Hence, inferential machinery is required for drawing conclusions from partially specified models, like those associating a κ with isolated default statements. Such machinery is provided by the conditional logic of Adams [1975], which forms the basis of ε-semantics.

Adams' logic can be regarded as a bi-valued infinitesimal analysis, with input sentences specifying κ values of only 0 and 1, corresponding to "likely" and "unlikely" associations. However, instead of insisting on a complete specification of $\kappa(w)$, the logic admits fragmentary sets of conditional sentences, treats them as constraints over the distribution of $\kappa(w)$, and infers only such statements that are compelled to acquire high likelihood in every distribution $\kappa(w)$ satisfying these constraints.

2. ε -Semantics

2.1 Two levels of knowledge

ε-semantics, like epistemic probabilities and many conditional logics (qv), distinguishes between two types of sentences, those that convey knowledge about necessary truths and the general tendency of things to happen, e.g. "Birds fly", "Birds are animals", and those that describe findings or observations specific to a given object or a situation, e.g., "Tim is a bird", "all blocks on this table are green". The first set of sentences, denoted K (for knowledge), corresponds to *nomic* (or law-like) assertions, and may include both defeasible (defaults) and strict sentences (denoted Δ and S, respectively). The second set of sentences, denoted E (for evidence), corresponds to incidental or transitory findings. This useful distinction is reflected in natural language by the selective usage of the word "If", especially in counterfactual forms. For example, it is legitimate to say "If I were a bird I would fly" but not "If this block were on this table it would be green". Accordingly, the sentence "Birds fly" will reside in K, while "all blocks on this table are green" will reside in E. Another distinguishing characteristic is that sentences in E accept the preemption "happen to", such as "Tim happened to be a bird". In contrast, those in K accept the preemption "always" or "almost always".

For simplicity of exposition, we shall first consider default theories in the form $T = <E, \Delta>$, void of strict conditionals. The evidence sentences (E) will assign properties to specific individuals; for example, $p(a)$ asserts that individual a has the property p. The default statements (Δ) are of the type "p's are typically q's", written $p(x) \rightarrow q(x)$ or simply $p \rightarrow q$, which is short for saying "any individual x having property p typically has property q". The properties $p, q, r \cdots$ can be compound boolean formulas of some atomic predicates $p_1, p_2, \ldots p_n$, with x as their only free variable. However, no ground defaults (e.g., $p(a) \rightarrow q(a)$) are allowed in E and no compound defaults (e.g., $p \rightarrow (q \rightarrow r)$) are allowed in Δ. The default statement $d' : p \rightarrow \neg q$ will be called the *denial* of $d : p \rightarrow q$.

2.2 Basic Definitions

Let L be the language of propositional formulas, and let a *truth-valuation* for L be a function t, that maps the sentences in L to the set $\{1,0\}$, (1 for *TRUE* and 0 for *FALSE,)* such that t respects the usual Boolean connectives. To define a probability assignment over the sentences in L, we regard each truth valuation t as a world w and define $P(w)$ such that $\sum_w P(w) = 1$. This assigns a probability measure to each sentence l of L via $P(l) = \sum_w P(w)\, w(l)$.

ε-semantics interprets Δ as a set of restrictions on P, in the form of *extreme* conditional probabilities, infinitesimally removed from either 0 or 1. For example, the sentence $Bird(x) \rightarrow Fly(x)$ is interpreted as $P(Fly(x) \mid Bird(x)) \geq 1 - \varepsilon$, where ε is understood to stand for an infinitesimal quantity that can be made arbitrarily small, short of actually being zero. Accordingly, ε-semantics qualifies a propositional formula r as a *plausible conclusion* of $T = <E, \Delta>$, written $E \mathrel{|\!\!\sim_{\Delta}} r$, whenever the restrictions of Δ force P to satisfy $\lim_{\varepsilon \rightarrow 0} P(r \mid E) = 1$.

It is convenient to characterize the set of conclusions sanctioned by this semantics in terms of the set of facts-conclusion pairs that are entailed by a given Δ. We call this relation ε-*entailment* (2) formally defined as follows:

Definition: Let $\mathcal{P}_{\Delta, \varepsilon}$ stand for the set of distributions licensed by Δ for any given ε, i.e.,

$$\mathcal{P}_{\Delta, \varepsilon} = \{P : P(v \mid u) \geq 1 - \varepsilon \quad \text{and} \quad P(u) > 0 \quad \text{whenever} \quad u \rightarrow v \in \Delta\} \tag{4}$$

A conditional statement $d : p \rightarrow q$ is said to be ε-*entailed* by Δ, if every distribution $P \in \mathcal{P}_{\Delta, \varepsilon}$ satisfies $P(q \mid p) = 1 - O(\varepsilon)$, (i.e., for every $\delta > 0$ there exists a $\varepsilon > 0$ such that every $P \in \mathcal{P}_{\Delta, \varepsilon}$ would satisfy $P(q \mid p) \geq 1 - \delta$).

2.3 Axiomatic Characterization

The conditional logic developed by Adams [1975] faithfully represents this semantics by qualitative inference rules, thus facilitating the derivation of new sound sentences by direct symbolic manipula-

(2) Adams (1975) named this p-entailment. However, ε-entailment better serves to distinguish this from other forms of probabilistic entailment, Section 4.

tions on Δ. The essence of Adams' logic is summarized in the following inference rules, restated for default theories in [Geffner 1988] (see also Lehmann and Magidor 1988, and Geffner and Pearl 1990).

Inference Rules: Let $T = <E, \Delta>$ be a default theory where E is a set of ground proposition formulas and Δ is a set of default rules. r is a plausible conclusion of F in the context of Δ, written $F \vdash_{\Delta} r$, iff r is derivable from F using the following rules of inference:

Rule 1 (Conditionals) $(p \rightarrow q) \in \Delta \implies p \vdash_{\Delta} q$

Rule 2 (Deduction) $p \vdash q \implies p \vdash_{\Delta} q$

Rule 3 (Cumulativity) $p \vdash_{\Delta} q, p \vdash_{\Delta} r \implies (p \wedge q) \vdash_{\Delta} r$

Rule 4 (Contraction) $p \vdash_{\Delta} q, (p \wedge q) \vdash_{\Delta} r \implies p \vdash_{\Delta} r$

Rule 5 (Disjunction) $p \vdash_{\Delta} r, q \vdash_{\Delta} r \implies (p \vee q) \vdash_{\Delta} r$

Rule 1 permits us to conclude the consequent of a default when its antecedent is all that has been learned, and this permission is granted regardless of other information that Δ may contain. Rule 2 states that theorems that logically follow from a set of formulas can be concluded in any theory containing those formulas. Rule 3 (called *triangularity* in [Pearl 1988] and *cautious monotony* in [Lehmann and Magidor 1988]) permits the attachment of any established conclusion (q) to the current set of findings (p), without affecting the status of any other derived conclusion (r). Rule 4 says that any conclusion (r) that follows from an evidence set (p) augmented by a derived conclusion (q) also follows from the original evidence set alone. Finally, rule 5 says that a conclusion that follows from two findings also follows from their disjunction.

Some Meta-Theorems

T-1 (Logical Closure) $p \models_{\Delta} q, p \wedge q \supset r \implies p \models_{\Delta} r$

T-2 (Equivalent Contexts) $p \equiv q, p \models_{\Delta} r \implies q \models_{\Delta} r$

T-3 (Exceptions) $p \wedge q \models_{\Delta} r, p \models_{\Delta} \neg r \implies p \models_{\Delta} \neg q$

T-4 (Right Conjunction) $p \models_{\Delta} r, p \models_{\Delta} q \implies p \models_{\Delta} q \wedge r$

Some Non-Theorems:

(Irrelevance) $p \models_{\Delta} r \implies p \wedge q \models_{\Delta} r$

(Transitivity) $p \models_{\Delta} q , q \models_{\Delta} r \implies p \models_{\Delta} r$

(Left Conjunction) $p \models_{\Delta} r , q \models_{\Delta} r \implies p \wedge q \models_{\Delta} r$

(Contraposition) $p \models_{\Delta} r \implies \neg r \models_{\Delta} \neg p$

(Rational Monotony)

$$p \models_{\Delta} r, \text{ NOT}(p \models_{\Delta} \neg q) \implies p \wedge q \models_{\Delta} r \tag{6}$$

This last property (similar to CV of conditional logic (qv)) has one of its antecedents negated, hence, its consequences cannot be derived from Δ using the five rules of ε-semantics. It is, nevertheless, a desirable feature of a consequence relation, and can be restored within ε-semantics using the extensions described in Section 3.

ε-semantics does not sanction transitivity, left conjunction, and contraposition as absolute inference rules, because there are possible worlds in which these rules fail. For instance, transitivity fails in the penguin example — all penguins are birds, birds typically fly, yet penguins do not. Left conjunction fails when p and q create a new condition unshared by either p or q. For example, if you marry Ann (p) you will be happy (r), if you marry Nancy (q) you will be happy as well (r), but if you marry both ($p \wedge q$), you will be miserable ($\neg r$). Contraposition fails in situations where $\neg p$ is incompatible with $\neg r$. For example, let $p \rightarrow r$ stand for $Birds \rightarrow Fly$. Now imagine a world in which the only nonflying objects are a few sick birds. Clearly, $Bird \rightarrow Fly$ holds, yet if we observe a nonflying object we can safely conclude that it is a bird, hence $\neg r \rightarrow p$, defying contraposition.

Semi-monotonicity: The consequence relation defined by ε-semantics is monotonic relative to the addition of default rules, i.e.,

$$\text{if} \quad p \mid\!\!\approx_{\Delta} r \quad \text{and} \quad \Delta \subseteq \Delta', \text{ then } \; p \mid\!\!\approx_{\Delta'} r \tag{7}$$

This follows directly from the fact that $\mathcal{P}_{\Delta', \varepsilon} \subseteq \mathcal{P}_{\Delta, \varepsilon}$ because each default statement imposes a new constraint on $\mathcal{P}_{\Delta, \varepsilon}$. Thus, ε-entailment is nonmonotonic relative to the addition of new findings (in E) and monotonic relative to the addition of new defaults (in Δ). Full nonmonotonicity will be exhibited in Section 3, where stronger forms of entailment are considered.

The cautious, semi-monotonic character of ε-semantics, and especially it's failure to accommodate arguments based on "irrelevance", (e.g., to conclude a red bird flies from "birds fly") clearly show the ε-semantics is not complete for default reasoning. Nevertheless, the set of conclusions that are derived by this semantics constitutes a core of plausible conclusions that should clearly be accommodated by every system of default reasoning (qv). Interestingly, it is this very core which more conventional approaches to default reasoning find hardest to accommodate.

2.4 Consistency and Ambiguity

An important and unique feature of ε-semantics is its ability to distinguish theories portraying inconsistencies (e.g., $<p \rightarrow q, p \rightarrow \neg q>$), from those conveying ambiguity (e.g.,

$<p(a) \wedge q(a), p \to r, q \to \neg r>$, and those conveying exceptions (e.g., $<p(a) \wedge q(a), p \to \neg q>$).

Definition: Δ is said to be ε-*consistent* if $\mathcal{P}_{\Delta, \varepsilon}$ is non-empty for every $\varepsilon > 0$, else, Δ *is* ε-*inconsistent*. A default statement $d: p \to q$ is said to be *ambiguous*, given Δ, if both $\{p \to q\} \cup \Delta$ and $\{p \to \neg >q\} \cup \Delta$ are ε-consistent.

Consistency-Entailment Symmetry: ε-entailment and ε-consistency are connected by a symmetrical relation, reminiscent of that in classical logic (Adams, 1975). If Δ is ε-consistent, then a statement $d : p \to q$ is ε-entailed by Δ iff its denial $d' : p \to \neg q$ is ε-inconsistent with Δ.

In addition to Rules 1-5, the logic also possesses a systematic procedure for testing ε-consistency (hence, ε-entailment), involving a moderate number of propositional satisfiability tests. The test is based on the notion of *toleration*:

Definition (Toleration): Given a truth-valuation t, a default statement $p \to q$ is said to be *verified* under t if t assigns the value 1 to both p and q. $p \to q$ is said to be *falsified* under t if p is assigned a 1 and q is assigned a 0. A default statement $d: p \to q$ is said to be *tolerated* by a set Δ' of such statements if there is a t that verifies d and does not falsify any statement in Δ'.

It can be shown (Adams, 1975) that a finite set Δ of default statements is ε-consistent iff in every non-empty subset Δ' of Δ there exists at least one statement that is tolerated by Δ'. This leads to a simple procedure of testing the consistency of defeasible databases:

Consistency testing procedure

1. Find a default statement that is tolerated by Δ,

2. remove it from Δ,

3. repeat the process on the remaining set of statements, until there are no more default statements left.

4. If this process leads to an empty set then Δ is ε-consistent, else it is ε-inconsistent.

This procedure requires $\frac{|\Delta|^2}{2}$ propositional satisfiability tests. Hence, if the material counterpart of $p \supset q$ of each statement $p \rightarrow q$ in Δ is a Horn expression, then consistency (hence entailment) can be tested in time quadratic with the number of literals in Δ (Goldszmidt and Pearl, 1989a). When Δ can be represented as a *default network,* (i.e., a set of default statements $p \rightarrow q$ where both p and q are atomic propositions (or negation thereof)), consistency can be established by a simple graphical criterion (Pearl 1987a), generalizing that of Touretzky (1986):

Network Consistency: Δ is consistent iff every pair of conflicting arcs $p_1 \rightarrow q$ and $p_2 \rightarrow \neg q$

1. p_1 and p_2 are distinct, and

2. There is no cycle of positive arcs that embraces both p_1 and p_2.

These tests are valid only when K consists of purely defeasible conditionals. For mixtures $K = <\Delta, S>$ of defeasible and non-defeasible statements, consistency and entailment require a slightly modified procedure (Goldszmidt and Pearl 1989a): After removing all tolerated sentences from Δ, each sentence in S should be tolerated by S. This procedure attributes a special meaning to strict conditional statement $s: a \rightarrow b$, different than the material implication $a \supset b$. For example, conforming to common usage of conditionals, it will proclaim $S = \{a \Rightarrow b, a \Rightarrow \neg b\}$ as inconsistent and will entail $a \Rightarrow b$ from $\neg b \Rightarrow \neg a$ but not from $\neg a$.

2.5 Illustrations

To illustrate the syntactical and graphical derivations facilitate by ε-semantics, consider the celebrated "Penguin triangle" of Figure 1.

x x x x x x x x x x Figure 1 x x x x x x x x x

T comprises the sentences:

$$F = \{Penguin\,(Tweety),\, Bird\,(Tweety)\}, \tag{8}$$

$$\Delta = \{Penguin \rightarrow \neg fly,\, Bird \rightarrow Fly,\, Penguin \rightarrow Bird\}; \tag{9}$$

Although Δ does not specify explicitly whether penguin-birds fly, the desired conclusion is derived in three steps, using Rule 1 and 3.

1. $Penguin\ (Tweety) \vdash_{\Delta} \neg Fly\ (Tweety)$ (from Rule 1)
2. $Penguin\ (Tweety) \vdash_{\Delta} Bird\ (Tweety)$ (from Rule 1)
3. $Penguin\ (Tweety), Bird(Tweety) \vdash_{\Delta} \neg Fly\ (Tweety)$ (Applying Rule 3 to lines 1, 2)

Note that preference toward subclass specificity is maintained despite the defeasible nature of the rule $Penguin \rightarrow Bird$, which admits exceptional penguins in the form of non-birds.

The conclusion $p \wedge b \rightarrow f$ can also be established by showing that its denial, $p \wedge b \rightarrow \neg f$, is ε-inconsistent with

$$\Delta = \{p \rightarrow \neg f, b \rightarrow f, p \rightarrow b\}. \tag{10}$$

Indeed, no truth-valuation of $\{p, b, f\}$ can verify any sentence in

$$\Delta' = \{p \rightarrow \neg f, p \rightarrow b, p \wedge b \rightarrow \neg f\} \tag{11}$$

without falsifying at least one other sentence.

Applying theorem T-3 to the network of Figure 1 yields another plausible conclusion, $Bird \rightarrow \neg Penguin$, stating that when one talks about birds one does not have penguins in mind, i.e., penguins are exceptional kind of birds. It is a valid conclusion of Δ because every P in $\mathcal{P}_{\Delta, \varepsilon}$ must yield $P(p \mid b) = O(\varepsilon)$. Of course, if the statement $Bird \rightarrow Penguin$ is artificially added to Δ, inconsistency results; as ε diminishes below a certain level (1/3 in our case), $\mathcal{P}_{\Delta, \varepsilon}$ becomes empty. This can be predicted from purely topological considerations by testing whether the denial of the conclusion renders the network inconsistent. Adding the arc $Bird \rightarrow Penguin$ would create a cycle of positive arcs embracing "Bird" and "Penguin", and these sprout two conflicting arcs toward "Fly", which establishes inconsistency. Hence, the network of Figure 1 ε-entails $Bird \rightarrow \neg Penguin$. By the same graphical method one can show that the network also ε-entails the natural conclusion, $Fly \rightarrow \neg Penguin$. This contraposition of $Penguin \rightarrow \neg Fly$ is sanctioned only because the existence of flying non-penguins (i.e., normal birds) is guaranteed by the other rules in Δ.

3. Recent Extensions

Summarizing the preceding discussion, ε-semantics yields a system of defeasible inference with the following features:

1. The system provides a formal distinction between exceptions, ambiguities and inconsistencies and effective procedures for testing and maintaining consistency.

2. Multiple extensions do not arise and preferences among arguments (e.g., toward higher specificity) are respected by natural deduction.

3. There is no need to specify abnormality relations in advance (as in circumscription (qv); such relations (e.g., that penguin are abnormal birds) are automatically inferred from the knowledge base.

However, default reasoning requires two facilities: one which forces conclusions to be retractable in the light of new refuting evidence; the second which protects conclusions from retraction in the light of new but irrelevant evidence. ε-semantics excel on the first requirement but fails on the second. For instance, in the example of Fig. 1, if we are told that Tweety is also a blue penguin, the system would retract all previous conclusions (as ambiguous), even though there is no rule which in any way connects color to flying. (The opposite is true in default logic [Reiter 1987] and circumscription [McCarthy 1986] - they excel on the second requirement but do not retract conclusions refuted by more specific information, unless exceptions are enumerated in advance.)

The reason for this conservative behavior lies in the insistence that any issued conclusion attains high probability in *all* probability models licensed by Δ and one such model reflects a world in which blue penguins do fly. In order to respect the communication convention that, unless stated explicitly, properties are presumed to be *irrelevant* to each other, additional restrictions must be imposed on the family of probability models relative to which a given conclusion is checked for soundness. The restricted probabilities should embody only dependencies that are implied by Δ, but no others. Several such extensions to ε-semantics will be described next.

3.1 System Z

One way of suppressing irrelevant properties is to restrict our attention to the "most normal" or "least surprising" probability models that comply with the constraints in Δ. This can be most conveniently done within the infinitesimal analysis of Spohn (see Section 1), where the ranking function $\kappa(w)$ represents the degree of surprise associated with world w. The "least surprising" probability corresponds to assigning each world w the lowest possible ranking $\kappa(w)$ permitted by the constraints in Δ. To ratify a sentence $p \rightarrow q$ within this paradigm, we must first find this *minimal* ranking function κ and, then, test whether $\kappa(q \mid p) < \kappa(\neg q \mid p)$ holds in this ranking.

Translating the constraints of Eq. (4) to the language of infinitesimals, yields

$$\kappa(v \wedge u) < \kappa(\neg v \wedge u) \quad \text{if } u \rightarrow u \in \Delta \tag{12.a}$$

where κ of a formula f is given by

$$\kappa(f) = \min_{w} \{\kappa(w) : w \models f\} \tag{12.b}$$

Remarkably, if Δ is ε-consistent, such constraints admit a unique minimal κ distribution which was named **Z**-ranking in (Pearl 1990). Moreover, finding this minimal distribution for a given world w, requires no more computation than testing for ε-consistency according to the procedure of Section 2.4. We first identify all default statements in Δ that are tolerated by Δ, assign to them a **Z**-rank of 0, and remove them from Δ. Next we assign a **Z**-rank of 1 to every default statement that is tolerated by the remaining set, and so on. Continuing in this way, we form an ordered partition of $\Delta = (\Delta_0, \Delta_1, \Delta_2, \cdots, \Delta_K)$, where Δ_i consists of all statements tolerated by $\Delta - \Delta_0 - \Delta_1$ - ... Δ_{i-1}. This partition uncovers a natural priority among the default rules in Δ, and represents the relative "cost" associated with violating any of these defaults, with preference given to the more specific classes.

Once we establish the **Z**-ranking on defaults, the minimal ranking on worlds is given by:

$$\mathbf{Z}(w) = \min \{ n : w \models (v \supset u), \quad \mathbf{Z}(v \rightarrow u) \geq n \} \tag{13}$$

In other words, $\mathbf{Z}(w)$ is equal to 1 plus the rank of the highest-ranked default statement falsified in w.

Given $Z(w)$, we can now define a useful extension of ε-entailment, which was called *1-entailment* in (Pearl 1990).

Definition (1-entailment): A formula g is said to be *1-entailed* by f, in the context Δ, (written $f \vdash_1 g$), if g holds in all minimal-**Z** worlds satisfying f. In other words,

$$f \vdash_1 g \quad \textit{iff} \quad \mathbf{Z}(f \wedge g) < \mathbf{Z}(f \wedge \neg g) \tag{14}$$

Note that ε-entailment is clearly a subset of 1-entailment since, using the language of Z-ranking, it corresponds to: $f \,|\!\!\!\sim_{\Delta}\, g$ iff $\mathbf{Z}(f \wedge \neg g) = \infty$

Lehmann (1989) has extended ε-entailment by closing it under the *rational monotony* rule of Eq. (6), thus obtaining a new consequence relation which he called *rational closure*. Goldszmidt and Pearl [1989b] have shown that 1-entailment and rational closure are identical whenever Δ is ε-consistent. Thus, the procedure for testing ε-consistency also provides a $O(|\Delta^2|)$ procedure for testing entailment in rational closure.

3.2 Illustrations

Figure 2 represents a knowledge base formed by adding three rules to that of Figure 1:

1. "Penguins live in the arctic" $p \rightarrow a$

2. "Birds have wings" $b \rightarrow w$

3. "Animals that fly are mobile" $f \rightarrow m$

The numerical labels on the arcs stand for the **Z**-ranking of the corresponding rules.

x x x x x x x x x x Figure 2 x x x x x x x x x

The following are examples of plausible consequences that can be drawn from Δ by the various systems discussed in this section (ME will be discussed in Section 3.3):

ε-entailed	1-entailed	*ME*-entailed
$b \wedge p \vdash_\Delta \neg f$	$\neg b \vdash_1 \neg p$	$p \vdash_{ME} w$
$f \vdash_\Delta \neg p$	$\neg f \vdash_1 \neg b$	$p \wedge \neg a \vdash_{ME} \neg f$
$b \vdash_\Delta \neg p$	$\neg f \vdash_1 m$	$p \wedge \neg a \vdash_{ME} w$
$p \wedge a \vdash_\Delta b$	$\neg m \vdash_1 \neg b$	
	$p \neg w \vdash_1 b$	

1-entailment sanctions many plausible inference patterns that are not ε-entailed, among them chaining, contraposition and discounting irrelevant features. For example, from the knowledge base of Figure 2 we can now conclude that birds are mobile, $b \vdash_1 m$, and that immobile objects are non-birds, $\neg m \vdash_1 \neg b$, and that green birds still fly. On the other hand, 1-entailment does not permit us to conclude that penguins who do not live in the arctic still do not fly, $p \wedge \neg a \rightarrow \neg f$. In general, from $\Delta = \{ a \rightarrow b, c \rightarrow d \}$ we cannot conclude $a \wedge \neg b \wedge c \rightarrow d$.

This inability to sanction property inheritance from classes to exceptional sub-classes represents the main weakness of 1-entailment. For example, given the knowledge base of Figure 2, 1-entailment will not sanction the conclusion that penguins have wings ($p \rightarrow w$) by virtue of being birds (albeit exceptional birds). The reason is that according to the **Z**-ranking procedure all statements conditioned on p should obtain a rank of 1, and this 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.). To sanction property inheritance across exceptional classes, a more refined ordering is required which also takes into account the *number* of defaults falsified in a given world, not merely their rank orders. One such refinement is provided by the maximum-entropy approach [Goldszmidt et al 1990] where each world is ranked by the sum of weights on the defaults falsified by that world. Another refinement is provided by Geffner's conditional entailment [Geffner 1989], where the priority of defaults induces a *partial* order on worlds. These two refinements will be summarized next.

3.3 The Maximum Entropy Approach

The maximum-entropy (ME) approach attempts to capture 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 [Jaynes 1979]. Given a set Δ of default rules and a family of probability distributions that are admissible relative the constraints conveyed by Δ (i.e., $P(\beta_r \rightarrow \alpha_r) \geq 1 - \varepsilon \;\; \forall \, r \in \Delta$), we single out a distinguished distribution $P^*_{\varepsilon,\Delta}$ having the greatest entropy $-\sum_w P(w)\log(w)$, and define entailment relative to this distribution by

$$f \vdash_{ME} g \quad \text{iff} \quad P^*_{\varepsilon,\Delta}(g \mid f) \underset{\varepsilon \to 0}{\to} 1 . \tag{15}$$

An infinitesimal analysis of the ME approach also yields a ranking function κ on worlds, where $\kappa(w)$ corresponds to the lowest exponent of ε in the expansion of $P^*_{\varepsilon,\Delta}(w)$ into a power series in ε. It can be shown that this ranking function can be encoded parsimoniously by assigning an integer weight κ_r to each default rule $r \in \Delta$ and letting $\kappa(w)$ be the sum of the weights associated with the rules falsified by w. The weight κ_r, in turn, reflects the "cost" we must add to each world w that falsifies rule r, so that the resulting ranking function would satisfy the constraints conveyed by Δ, namely,

$$\min \{ \kappa(w) : w \models \alpha_r \wedge \beta_r \} < \min \{ \kappa(w) : w \models \alpha_r \wedge \neg \beta_r \} , \qquad r : \alpha_r \to \beta_r \in \Delta. \tag{16}$$

These considerations lead to a set of $|\Delta|$ non-linear equations for the weights κ_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. (15), translated to

$$f \vdash_{ME} g \quad \text{iff} \quad \min \{ \kappa(w) : w \models f \wedge g \} < \min \{ \kappa(w) : w \models f \wedge \neg g \} . \tag{17}$$

where

$$\kappa(w) = \sum_{r : w \models \alpha_r \wedge \neg \beta_r} \kappa_r$$

We see that ME-entailment requires minimization over worlds, a task that is *NP*-hard even for Horn expressions (Ben-Eliyahu, 1990). In practice, however, this minimization is accomplished quite effectively in network type databases, yielding a reasonable set of inference patterns. For example, in the database of Figure 2, 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 an undesirable feature of 1-entailment which concludes $c \wedge p \vdash_1 \neg f$ from $\Delta \cup \{c \rightarrow f\}$, where c is an irrelevant property.

An interesting feature of the ME approach is its sensitivity to the format in which the rules are expressed. This is illustrated in the following example. From $\Delta =$ {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 $\Delta =$ {Swedes are blond and well-mannered}. This sensitivity might sometimes be useful for distinguishing fine nuances in natural discourse, indicating, for example, that behavior and hair color are two independent qualities. It stands at variance with most approaches to default reasoning, where $a \rightarrow b \wedge c$ is treated as a shorthand notation of $a \rightarrow b$ and $a \rightarrow c$.

The ME approach fails to respond to causal information (see Pearl [1988, pp. 463, 519] and Hunter [1989]). This prevents it from properly handling tasks such as the Yale shooting problem [Hanks and McDermott 1986], where rules of causal character should be 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]).

3.4 Conditional Entailment

Conditional entailment [Geffner 1989] overcomes the weaknesses of 1-entailment by introducing two refinements. First, rather than letting rule priorities dictate a ranking function on worlds, a partial order on worlds is induced instead. To determine the preference between two worlds, w and w', we examine the highest priority default 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 worlds, then this world is the preferred one. Formally, if $\Delta[w]$ and $\Delta[w']$ stand for the set of rules falsified by w and w', respectively, then w is preferred to w' (written $w < w'$) iff $\Delta[w'] \neq \Delta[w']$ and for every rule r in $\Delta[w] - \Delta[w']$ there exists a rule r' in $\Delta[w'] - \Delta[w]$ such that r' has a higher priority than r (written $r \prec r'$). Using this criterion, a world w will always be preferred to w' if it falsifies a proper subset of the rules falsified by w'. Lacking this feature in the **Z**-ordering has prevented 1-entailment from concluding $p \vdash w$ in the example of Figure 2.

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 $\alpha \to \beta$; if α is all that we know, then, regardless of other rules that Δ may contain, we are authorized to assert β. This means that $r: \alpha \to \beta$ 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 $\Delta' \subset \Delta$ does not tolerate r, then at least one rule in Δ' ought to have a lower priority than r. In Figure 2, for example, the rule $r_3: p \to \neg f$ is not tolerated by the set $\{r_1: p \to b, r_2: b \to f\}$, hence, we must have $r_1 \prec r_3$ or $r_2 \prec r_3$. Similarly, the rule $r_1: p \to b$ 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 $w < w'$, whenever w validates $p \wedge b \wedge \neg f$ and w' 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 Δ) if g holds in all the preferred worlds of f induced by every priority ordering admissible with Δ.

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].

References

[Adams 1966] Adams, E. 1966. Probability and the logic of conditionals. In *Aspects of inductive logic,* ed. J. Hintikka and P. Suppes. Amsterdam: North Holland.

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

[Ben-Eliyahu 1990] Ben-Eliyahu, R., 1990. Minimizing Horn-Clause violations is NP-hard. UCLA Cognitive Systems Laboratory, Computer Science Dept., Technical Report (R-158).

[DelGrande 1988] DelGrande, J.P. 1988. An approach to default reasoning based on first-order conditional logic; revised report. *Artificial Intelligence,* 36:63-90.

[Geffner 1988] Geffner, H. 1988. On the Logic of defaults. *Proc.of the Natl. Conf. on AI (AAAI-88),* St. Paul, 449-454.

[Geffner 1989] Geffner, H. 1989. "Default reasoning: causal and conditional theories. PhD. Dissertation. UCLA, Computer Science Department, November 1989.

[Geffner and Pearl 1990] Geffner, H., & Pearl, J. 1990. A framework for reasoning with defaults. In Kyburg et al (Eds.) *Knowledge Representation and Defeasible Inference,* Kluwer Academic Publishers, pp. 69-87, 1990.

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

[Goldszmidt and Pearl 1989b] Goldszmidt, M. and Pearl, J. 1989. On the relation between rational closure and System-Z. In *3rd International Workshop on Nonmonotonic Reasoning,* May

1990, pp. 130-140.

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

[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, 328-33.

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

[Jaynes 1979] Jaynes, E. T. 1979. Where do we stand on maximum entropy? In *The maximum entropy formalism,* ed. R. D. Levine and M. Tribus. Cambridge: MIT Press.

[Kyburg 1961] Kyburg, H. E. 1961. *Probability and the logic of rational belief.* Middleton, Conn.: Wesleyan University Press.

[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.

[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.

[McCarthy 1980] McCarthy, J. 1980, Circumscription - A form of non-monotonic Reasoning, *Artificial Intelligence,* vol. 13 (1), 1980, pp. 27-70.

[McCarthy 1986] McCarthy, J. 1986. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence* 28 (no. 1):89-116.

[Pearl 1987a] Pearl, J. 1987. Deciding consistency in Inheritance Networks. UCLA Cognitive Systems Laboratory, Technical Report 870053 (R-96).

[Pearl 1987b] Pearl, J. 1987. Distributed revision of composite beliefs. *Artificial Intelligence* 33 (no. 2):173-215.

[Pearl 1988] *Probabilistic reasoning in intelligent systems: networks of plausible inference.* 1988. San Mateo: Morgan Kaufmann.

[Pearl 1990] Pearl, J., "System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning, *Proceedings* Theoretical Aspects of Reasoning about Knowledge (TARK-III), Asilomar, CA, March 1990, Morgan Kaufmann, pp. 121-135.

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

[Reiter 1987] Reiter, R. 1987. Nonmonotonic reasoning. In *Annual review of computer science,* vol. 2, 147-86. Palo Alto, Calif.: Annual Reviews.

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

[Spohn 1987] Spohn, W. 1987. Ordinal Conditional Functions: A Dynamic Theory of Epistemic States, in *Causation in Decision, Belief Change, and Statistics,* W.L. Harper and B. Skyrms, eds. Dordrecht: Reidel, vol. 2, 105-134.

[Touretzky 1986] Touretzky, D. 1986. *The mathematics of inheritance systems.* Los Altos, Calif.: Morgan Kaufmann.

This work was supported in part by National Science Foundation Grant #IRI-8821444 and Naval Research Laboratory Grant #N00014-87-K-2029. An earlier version of this paper was presented at the First International Conference on Principles of Knowledge Representation and Reasoning (KR '89), Toronto, May 1989.

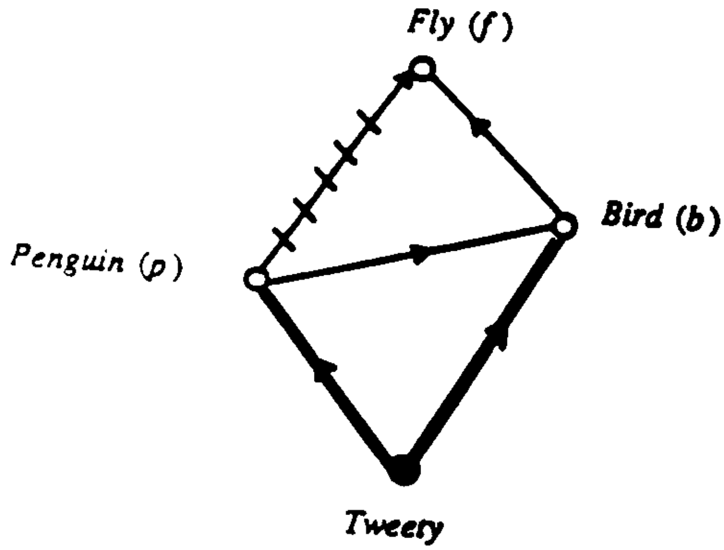

Figure 1

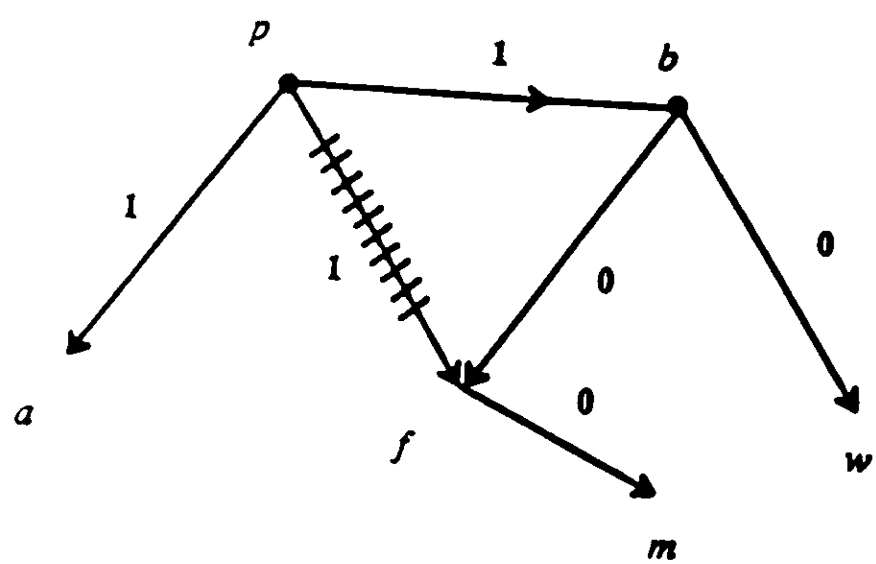

Figure 2