ELSEVIER

Artificial Intelligence 84 (1996) 57-112

Artificial Intelligence

Qualitative probabilities for default reasoning, belief revision, and causal modeling

Moisés Goldszmidt [a,1], Judea Pearl [b,*]

[a] SRI International, Menlo Park, CA 94025, USA

[b] University of California (UCLA), Computer Science Department, Los Angeles, CA 90024, USA

Received June 1993; revised June 1995

Abstract

This paper presents a formalism that combines useful properties of both logic and probabilities. Like logic, the formalism admits qualitative sentences and provides symbolic machinery for deriving deductively closed beliefs and, like probability, it permits us to express if-then rules with different levels of firmness and to retract beliefs in response to changing observations. Rules are interpreted as order-of-magnitude approximations of conditional probabilities which impose constraints over the rankings of worlds. Inferences are supported by a unique priority ordering on rules which is syntactically derived from the knowledge base. This ordering accounts for rule interactions, respects specificity considerations and facilitates the construction of coherent states of beliefs. Practical algorithms are developed and analyzed for testing consistency, computing rule ordering, and answering queries. Imprecise observations are incorporated using qualitative versions of Jeffrey's rule and Bayesian updating, with the result that coherent belief revision is embodied naturally and tractably. Finally, causal rules are interpreted as imposing Markovian conditions that further constrain world rankings to reflect the modularity of causal organizations. These constraints are shown to facilitate reasoning about causal projections, explanations, actions and change.

1. Rankings as an order-of-magnitude abstraction of probabilities

The uncertainty encountered in common sense reasoning fluctuates over an extremely wide range. For example, the probability that the new book on my desk is about astrology may be less than one in a million. However, if I open the wrappings and

* Corresponding author. E-mail: judea@cs.ucla.edu.

[1] E-mail: moises@erg.sri.com.

SSDI 0004-3702(95)00090-9

see a Zodiac sign, the probability becomes close to 1, say 0.999. Intelligent agents are expected to reason with such rare eventualities and to produce explanations and actions whenever these occur. Given this wide range of uncertainty fluctuations and the fact that the majority of everyday decisions involve relatively low payoffs, the full precision of probability calculus may not be necessary, and an order-of-magnitude approximation may be sufficient. Thus, instead of measuring probabilities on a scale from zero to one, we can imagine projecting probability measures onto a quantized logarithmic scale and then treating beliefs that map onto two different quanta as being of different orders of magnitude.

This method of approximation gives rise to a semi-qualitative calculus of uncertainty, one in which degrees of (dis)belief are ranked by nonnegative integers (corresponding perhaps to linguistic quantifiers such as "believable", "unlikely", "very rare") still capable of accounting for retraction and restoration of beliefs by Bayesian conditioning. The origin of this ranked-based approximation can be traced back to Adams [1], who developed a logic of conditionals based on infinitesimal probabilities, and to the ordinal conditional functions (OCFs) of Spohn [71]. Potential applications in nonmonotonic reasoning were noted in [50,54] and further developed in [25,31,34,46,47,57].

One way of motivating integer rankings of beliefs is to consider a probability distribution $P(\omega)$ defined over a set Ω of possible worlds and to imagine that an agent wishes to extract an order-of-magnitude approximate of $P(\omega)$. The traditional engineering method of approximation would be to express each numerical parameter (specifying P) in a base-b representation, where b depends on the precision needed, and then omit all but the most significant figure from each expression. [2] All arithmetic operations would then be performed on these approximate, single digit quantities, in lieu of the original parameters. The abstraction we advocate goes one step further. Instead of retaining the numerical value of the most significant figure, we retain only its *position*. The mechanics of this exercise is equivalent to treating the base b as an infinitesimal number ε, thus mapping every quantity to a polynomial in ε. These polynomials are added and multiplied precisely, but at the end we calculate the limit of the final results as ε goes to zero.

Specifically, if we write the probability $P(\omega)$ as a polynomial in ε; for example, $P_\varepsilon(\omega) = 1 - c_1\varepsilon$ or $\varepsilon^2 - c_2\varepsilon^4$, We define a ranking function $\kappa(\omega)$ as the power of the most significant ε-term in $P_\varepsilon(\omega)$, or,

$$\kappa(\omega) = \begin{cases} \min\left\{n \text{ such that } \lim\limits_{\varepsilon\to 0} P(\omega)/\varepsilon^n \neq 0\right\}, & \text{if } P_\varepsilon(\omega) > 0, \\ \infty, & \text{if } P_\varepsilon(\omega) = 0. \end{cases} \qquad (1)$$

Likewise, since the probabilities assigned to any logical formula φ, as well as all conditional probabilities $P_\varepsilon(\psi|\varphi)$, will be rational functions of ε, we define the ranking function $\kappa(\psi|\varphi)$ as the power of the most significant ε-term in the expansion of

[2] Thus, given a basis b, a quantity p will be expressed by the polynomial $p = a_0 * (b)^0 + a_1 * (b)^1 + a_2 * (b)^2 + \cdots$, and will be approximated by the most significant term of this polynomial, namely, the first term where $a_i \neq 0$, thus $p \approx a_i * (b)^i$.

Table 1
Linguistic quantifiers and ε^n

$P(\phi) = \varepsilon^0$	ϕ and $\neg\phi$ are possible	$\kappa(\phi) = 0$
$P(\phi) = \varepsilon^1$	$\neg\phi$ is believed	$\kappa(\phi) = 1$
$P(\phi) = \varepsilon^2$	$\neg\phi$ is strongly believed	$\kappa(\phi) = 2$
$P(\phi) = \varepsilon^3$	$\neg\phi$ is very strongly believed	$\kappa(\phi) = 3$
$\vdots$	$\vdots$	$\vdots$

$P_\varepsilon(\psi|\varphi)$. In other words, $\kappa(\psi|\varphi) = n$ iff $P_\varepsilon(\psi|\varphi)$ has the same order of magnitude as ε^n. [3]

Parameterizing a probability measure by ε and extracting the lowest exponent of ε as the measure of (dis)belief was proposed in [55] as a model of the process by which people abstract qualitative beliefs from numerical probabilities and accept them as tentative truths. For example, we can make the correspondence between linguistic quantifiers and ε^n depicted in Table 1. This abstraction yields an integer-addition calculus which combines the benefits of logic and probabilities. Like logic, it permits us to reason symbolically and form deductively closed beliefs and, like probability, it permits us to retract beliefs in response to changing observations, using the ranking-equivalent of Bayesian conditioning as shown below. The following properties of ranking functions (left-hand side below) reflect, on a logarithmic scale, the usual properties of probability functions (right-hand side), with min replacing addition, and addition replacing multiplication:

$$\kappa(\varphi) = \min_{\omega\models\varphi} \kappa(\omega): \qquad P(\varphi) = \sum_{\omega\models\varphi} P(\omega), \tag{2}$$

$$\kappa(\varphi) = 0 \text{ or } \kappa(\neg\varphi) = 0: \qquad P(\varphi) + P(\neg\varphi) = 1, \tag{3}$$

$$\kappa(\psi|\varphi) = \kappa(\psi \wedge \varphi) - \kappa(\varphi): \quad P(\psi|\varphi) = P(\psi \wedge \varphi)/P(\varphi). \tag{4}$$

This correspondence dictates the following principles on the semantics of rankings and beliefs.

(1) Each world is ranked by a nonnegative integer κ representing the degree of surprise associated with finding such a world.
(2) Each wff is given the rank of the world with the lowest κ (most normal world) that satisfies that wff.
(3) Given a ranking κ and a collection of facts ϕ, we say that σ *is believed given* ϕ if $\kappa(\neg\sigma|\phi) > 0$, or, equivalently, if the σ holds in all the lowest κ (most normal) worlds satisfying ϕ.

Principles (1) and (2) follow immediately from Eq. (2). Principle (3) associates beliefs with extreme conditional probabilities, saying that σ is believed given ϕ iff $P(\sigma|\phi) \geqslant 1 - C\varepsilon$ (for sufficiently small $\varepsilon > 0$), where P is the ε-parameterized probability associated with that particular ranking κ. This abstraction of probabilities

[3] Spohn [71] was the first to study such ranking functions, which he named ordinal conditional function (OCF) for the representation of *plain beliefs*. He also noted their equivalence to non-standard probabilities but considered this coincidence to be of formal rather than conceptual interest. Rankings are also implicit in Adams' consistency test [1].

matches the notion of *plain belief* [71] in that it is deductively closed: If A is believed and B is believed, then $A \wedge B$ is believed as well because $\kappa(\neg(A \wedge B)) > 0$ whenever $\kappa(\neg A) > 0$ and $\kappa(\neg B) > 0$. This deviates from the probability-threshold conception of belief: if both $P(A)$ and $P(B)$ are above a certain threshold, $P(A \wedge B)$ may still be below that threshold. The drawback of this abstraction is that many small probabilities do not accumulate into a strong argument (as in the lottery paradox [44]). However, in mundane reasoning applications, where reasoning chains are relatively shallow, such deviations from numerical probability calculus are usually tolerable—a reasonable price for achieving deductive closure [55].

Reasoning using principles (1)–(3) requires complete specification of the κ function, which is not readily available in practice. We are usually given information in the form of statements such as "birds normally fly" which we may interpret as $P(f|b) \geqslant 1 - \varepsilon$ (see also [33]) or, equivalently, $\kappa(\neg f|b) > 0$, and no information whatsoever about the flying habits of red birds or non-birds. In this case, we still would like to conclude "red birds normally fly", even though the information given is not sufficient for defining a complete ranking function. Drawing plausible conclusions from such fragmentary pieces of information requires additional inferential machinery with two features: it should enrich the specification of the ranking function with the needed information, and it should operate directly on the specification sentences in the knowledge base, rather than on the rankings of worlds (which are too numerous to list). Such machinery is provided by the formalism we propose in this paper, which accepts knowledge in the form of if–then rules (interpreted as constraints on κ) and computes the confidence in (i.e., ranking of) any given query by syntactic manipulation of these rules.

To accomplish these functions, we incorporate two principles in addition to those given above:

(4) Each input rule "if φ then ψ", written $\varphi \to \psi$, is interpreted as a constraint on the ranking κ, forcing every world in $\varphi \wedge \neg\psi$ to rank at least one rank above the most normal world in φ, that is, $\kappa(\neg\psi|\varphi) > 0$.

(5) Out of all rankings satisfying the constraints above, we adopt only those that are *minimal*, in the sense of assigning each world the lowest possible (most normal) rank. Remarkably, unlike most notions based on minimality, this ranking will turn out to be unique, denoted κ^+.

Principle (4) is a straightforward consequence of the probabilistic reading of the rules, $P(\psi|\varphi) \geqslant 1-\varepsilon$. Principle (5) reflects an assumption of reasonable cautiousness; unless compelled otherwise, assume every situation to be as serious a possibility as permitted by the given input information. We remark that although different sets of rules can give rise to the same ranking function κ^+, such sets are not entirely equivalent; while they yield the same answers to queries, and same responses to new observations, they differ in the way the knowledge base absorbs new rules.[4]

An inference system based on principles (1)–(5), called system-Z, is described in [57] and is reviewed in Section 2 (readers familiar with system-Z may wish to skip this section and go directly to Section 3). The distinctive feature of this system is that all inferences are conducted by syntactic processing of the rules in the knowledge base

[4] See Section 6 for a more detailed discussion.

$\Delta = \{\psi_i \rightarrow \phi_i\}$, $1 \leqslant i \leqslant n$, and not on the rankings of worlds (as in [71]) or belief sets (as in [24]). To this end, the knowledge base is first processed so as to assign each rule $r_i \in \Delta$ a priority number, $Z(r_i)$, which summarizes the interactions of r_i with other rules in Δ. Section 2 shows that the Z priorities can be computed in $O(n^2)$ propositional satisfiability tests and then, once Z is compiled, queries can be answered using $O(\log n)$ such tests, where n is the number of rules in Δ, and the satisfiability tests are performed on the *material counterpart*[5] of the rules in Δ. Section 2 also includes a test for the consistency of Δ and examples illustrating the use of system-Z for default reasoning.

The main focus of this paper lies in augmenting system-Z with the capability of handling richer types of input information, including variable-strength rules (Section 3), indirect evidence (Section 5) and causal rules (Section 7).

The realization that some default rules are stated with greater firmness than others has occurred in many contexts. For example, action-response defaults of the type "if Fred is shot with a loaded gun, Fred is dead" are normally stated 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 seconds. In diagnosis applications, likewise, the analyst may feel strongly that failures are more likely to occur in one type of device (e.g., multipliers) than in another (e.g., adders). Although numerical probabilities or degrees of certainty have been suggested for expressing this valuable knowledge, if the full precision provided by numerical calculi is not necessary, an intermediate qualitative language like the one proposed in this paper might be more suitable. For this purpose Section 3 augments principles (3) and (4) as follows:

(3′) Given a ranking κ and a collection of facts ϕ, we say that σ *is believed with strength* δ, given ϕ, if $\kappa(\neg\sigma|\phi) > \delta$, or, equivalently, if the κ-rank of $\phi \wedge \neg\sigma$ is at least $\delta + 1$ degrees above that of ϕ.

(4′) Each input rule "if φ then ψ (with strength δ)", written $\varphi \xrightarrow{\delta} \psi$, is interpreted as a constraint forcing every world in $\varphi \wedge \neg\psi$ to rank *at least* $\delta + 1$ *degrees above* the most normal world in φ, that is, $\kappa(\neg\psi|\varphi) > \delta$.

In probabilistic terms, principle (3′) says that given ϕ, σ is *believed to a degree* δ iff $P(\sigma|\phi) \geqslant 1 - C\varepsilon^{\delta+1}$, where P is the ε-parameterized probability associated with that particular ranking κ. Principle (4′) encodes the probabilistic reading of the rules, $P(\psi|\varphi) \geqslant 1 - \varepsilon^{\delta+1}$. The parameter δ is an optional feature for the rule encoder that augments the expressiveness of the knowledge base. If δ is unspecified, it is assumed to be equal to zero, and rules are interpreted as in principle (4) above (i.e., $P(\psi|\varphi) \geqslant 1 - \varepsilon$).

The inference system devised to accommodate variable-strength rules is called system-Z^+. A knowledge base with all $\delta = 0$ will be called *flat* and simply reduces to the one analyzed in Section 2. Remarkably, the introduction of variable-strength measures does not change the procedure required for consistency checking (Section 3.1), and results in only slightly higher complexity of the inference process, when compared with that of a flat system (Section 3.2). It is shown in Section 3.2 that the complexity of query-answering procedures increases, to account for the δ's, from $O(n^2)$ (for a

[5] The material counterpart of $\varphi \rightarrow \psi$ is the propositional formula $\varphi \supset \psi$.

flat knowledge base) to $O(n^2 \times \log n)$ satisfiability tests. Parallel to system-Z, query answering is facilitated by computing for each rule a priority number $Z^+(r_i)$, which accounts for both specificity and rule interaction under the constraints imposed by the variable-strength rules.

These procedures are polynomial for Horn expressions, network theories, or acyclic databases. Comparisons to related proposals for default reasoning can be found in Section 4.

In Section 5, system-Z^+ is equipped with the capability to reason with *soft* evidence or imprecise observations. Such a capability is important when we wish to assess the plausibility of σ (using principle (3) above) but the context ϕ is not given with absolute certainty; all that can be ascertained is "ϕ is supported to a degree m". We propose two different strategies for computing a new ranking κ' from such soft evidential reports. The first strategy, named *J-conditioning* (Section 5.1), is based on Jeffrey's rule of conditioning [56] which interprets the report as taking "all things" into consideration: the new degree of disbelief for $\neg\phi$ should be $\kappa'(\neg\phi) = m$. The second strategy, named *L-conditioning* (Section 5.2), interprets the report as specifying the desired *shift* in the degree of belief in ϕ, as warranted by that report *alone*. We show that both J- and L-conditioning have roughly the same complexity as ordinary conditioning.

Section 6 relates system-Z^+ to the theory of belief revision in [2] and shows that J-conditioning offers a natural realization of rational belief revision, overcoming several deficiencies in the AGM formulation. Additionally, the section identifies five belief revision operations which cannot be characterized using operators on belief sets alone, but require formulation in terms of conditional rules.

Section 7 deals with default rules that convey causal relationships. Such defaults, especially those specifying the effect of actions, require special treatment in most default formalisms, but present no special difficulties in probabilistic analysis based on Bayesian networks [46]. In Section 7 we borrow from probabilistic analysis the independence conditions that are typical to causal organizations and show that by imposing these conditions as constraints on ranking functions, we endow default rules with the causal character necessary to support reasoning about actions, their indirect consequences and their interaction with observations.

2. Ranking functions and system-Z: a review

2.1. Consistency

Consider the basic language to be a finite set $\mathcal{L}$ of atomic propositions augmented with two propositional constants $\top$ (*true*) and $\perp$ (*false*). Let $\mathcal{L}_P$ be a closed set of propositional well-formed formulas (wffs) generated as usual from the atomic propositions in $\mathcal{L}$ and the connectives $\vee$, $\wedge$, $\supset$ and $\neg$. We define a *world* ω as a truth assignment for the atomic propositions in $\mathcal{L}$. The set of possible worlds is denoted by Ω, and if there are n atomic propositions in $\mathcal{L}$, the size of Ω will be 2^n. The satisfaction of a wff $\varphi \in \mathcal{L}_P$ by a world ω is defined as usual and denoted by $\omega \models \varphi$. If ω satisfies φ, we say that ω is a *model* for φ. Ranking functions are defined as follows:

Definition 1 (*Ranking functions*). A *ranking function* κ is an assignment of non-negative integers to the elements in Ω, such that $\kappa(\omega) = 0$ for at least one $\omega \in \Omega$.

We extend this definition to induce rankings on wffs (in $\mathcal{L}_P$) in accordance with the probabilistic interpretation given by Eq. (2):

$$\kappa(\varphi) = \begin{cases} \min\limits_{\omega \models \varphi} \kappa(\omega), & \text{if } \varphi \text{ is satisfiable,} \\ \infty, & \text{otherwise.} \end{cases} \tag{5}$$

Similarly, following Eq. (4), we define the conditional ranking $\kappa(\psi|\varphi)$ for a pair of wffs φ and ψ (from $\mathcal{L}_P$) as

$$\kappa(\psi|\varphi) = \begin{cases} \kappa(\psi \wedge \varphi) - \kappa(\varphi), & \text{if } \kappa(\varphi) \neq \infty, \\ \infty, & \text{otherwise.} \end{cases} \tag{6}$$

Intuitively, $\kappa(\psi|\varphi)$ stands for the degree of incremental *surprise* or *abnormality* associated with finding ψ to be true, given that we already know φ. The inequality $\kappa(\neg\psi|\varphi) > 0$ means that given φ it would be surprising (i.e., abnormal) by at least one additional rank to find $\neg\psi$, which is precisely the interpretation we attribute to the conditional sentence (rule) "if φ then ψ".

A rule (or *default*) is the formula $\varphi \rightarrow \psi$, where φ and ψ are wffs in $\mathcal{L}_P$ and $\rightarrow$ is a new binary connective, conveying generic domain knowledge. Such rules express what is normally the case in the domain without excluding the possibility of exceptions. [6] The generic background information carried by a rule "if φ then ψ" is to be distinguished from the contingent information conveyed by ordinary wffs or the introspective information carried by the conditional sentence "if I presume ψ then φ would be believed". The distinction amounts to viewing the inequality $\kappa(\neg\psi|\varphi) > 0$ as a permanent requirement, to be upheld regardless of other rules in the knowledge base, rather than a feature specific to a given collection of rules. In other words, the inequality $\kappa(\neg\psi|\varphi) > 0$ is to be treated as a specification constraint for forming ranking functions, rather then a feature of an existing ranking function.

Given a knowledge base $\Delta = \{\varphi_i \rightarrow \psi_i\}$, $1 \leqslant i \leqslant n$, consistency can be defined as follows

Definition 2 (*Consistency*). A ranking κ is said to be *admissible* relative to a given Δ iff

$$\kappa(\varphi_i \wedge \psi_i) < \kappa(\varphi_i \wedge \neg\psi_i) \tag{7}$$

(equivalently $\kappa(\neg\psi_i|\varphi_i) > 0$) for every rule $\varphi_i \rightarrow \psi_i \in \Delta$. A knowledge base Δ is *consistent* iff there exists an admissible ranking κ relative to Δ.

[6] The use of conditional sentences to represent and reason with default information has been proposed by a number of researchers; see for example [8,20,25,26,28,34,43].

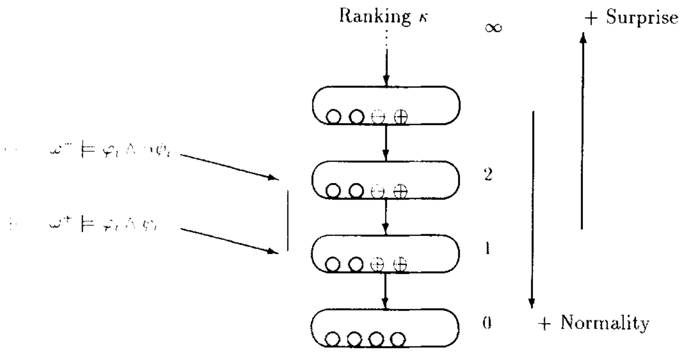

Fig. 1. Admissibility constraint imposed by $\varphi_i \to \psi_i$.

Eq. (7) echoes the preferential interpretation of *default* rules [68], according to which ψ holds in all *minimal* models for φ. In our case, minimality is reflected in having the lowest (i.e., most normal) rank. Let us say that a world ω *verifies* a rule $\varphi \to \psi$ if $\omega \models \varphi \wedge \psi$. A world ω *falsifies* $\varphi \to \psi$ if $\omega \models \varphi \wedge \neg\psi$. Admissibility requires that each time we find a world ω^- falsifying $\varphi_i \to \psi_i$, there exists a world ω^+ verifying $\varphi_i \to \psi_i$ such that $\kappa(\omega^+)$ is lower than $\kappa(\omega^-)$ (see Fig. 1). In probabilistic terms, consistency guarantees that for every $\varepsilon > 0$, there exists a probability distribution P such that if $\varphi_i \to \psi_i \in \Delta$, then $P(\psi_i|\varphi_i) \geqslant 1 - \varepsilon$. A more elaborate definition of consistency applies to knowledge bases containing a mixture of defeasible and non-defeasible (strict) rules [33]. For simplicity we skip the treatment of strict rules in this paper. [7]

The notion of *tolerance* below is a central component in establishing consistency (and priorities among rules), and identifies a satisfiability test crucial for most of the procedures in this paper.

Definition 3 (*Tolerance*). A rule $\alpha \to \beta$ is tolerated by a knowledge base $\Delta = \{\varphi_i \to \psi_i\}$, $1 \leqslant i \leqslant n$, iff there exists a ω such that

$$\omega \models \alpha \wedge \beta \bigwedge_{i=1}^{n} \varphi_i \supset \psi_i. \tag{8}$$

In other words, $\alpha \to \beta$ is tolerated by Δ iff there exists a ω, such that ω verifies $\alpha \to \beta$ and ω does not falsify any rules in Δ.

Theorem 4 (see [1]). *Δ is consistent iff in every non-empty subset $\Delta' \subseteq \Delta$ there exists a rule tolerated by Δ'.*

[7] To include strict rules Definition 2 should interpret a strict rule $\phi \Rightarrow \sigma$ as imposing the constraint: $\kappa(\neg\sigma \wedge \phi) = \infty$ and $\kappa(\phi) < \infty$. The procedure in Fig. 2 must also be modified, and its complexity requires an extra $O(s)$ satisfiability tests, where s is the number of strict rules in the knowledge base (see [28,33] for details).

Procedure Consistency-Test
Input: A knowledge base $\Delta = \{\varphi_i \rightarrow \psi_i \mid 1 \leqslant i \leqslant n\}$.
Output: An ordered partition of $\Delta = (\Delta_0, \Delta_1, \ldots, \Delta_k)$ iff Δ is consistent.
1. Let $i = 0$.
2. While Δ is non-empty
   (a) Find the set of rules $\Delta_i = \{\varphi_j \rightarrow \psi_j\}$ from Δ, such that each $\varphi_j \rightarrow \psi_j$ is tolerated by Δ
   (b) If none can be found then ABORT: Δ is inconsistent
   (c) Else remove Δ_i from Δ and set $i = i + 1$.
3. Return $\Delta = (\Delta_0, \Delta_1, \ldots, \Delta_k)$

End Procedure

Fig. 2. Procedure for testing consistency.

It turns out that not every subset needs to be checked for toleration:

Theorem 5 (see [57]). *Δ is consistent iff an ordered partition of $\Delta = (\Delta_0, \Delta_1, \ldots, \Delta_k)$ can be built such that every rule in Δ_i is tolerated by the set $\bigcup_{j=i}^{k} \Delta_j$.*

A procedure (Consistency-Test) for building this partition and testing consistency is given in Fig. 2. The complexity of Consistency-Test is essentially determined by step 2(a) which identifies the rules in the set Δ_i of the partition. Note that in the worst case (when no such rule is found) this step requires at most n propositionally satisfiability tests like the one described in Eq. (8), where n is the number of rules in Δ. Also, in the worst case, step 2(a) is executed n times. Defining $\varphi_i \supset \psi_i$ to be the material counterpart of $\varphi_i \rightarrow \psi_i$, we have the following corollary:

Corollary 6. *The consistency of a knowledge base Δ containing n conditional rules can be tested in $O(n^2)$ propositional satisfiability tests on the material counterpart of Δ.*

The next subsection shows that the partition $\Delta = (\Delta_0, \Delta_1, \ldots, \Delta_k)$ defines a natural ordering of the rules in Δ which in turn defines the minimal ranking among possible worlds and a notion of entailment.

2.2. Entailment: drawing plausible conclusions

Given a generic knowledge base in the form of a consistent set Δ of rules, together with a wff ϕ describing the collection of specific facts known about the domain, we wish to characterize the set of conclusions that can plausibly be entailed from ϕ in the context of Δ. For any specific ranking function the conclusions entailed by ϕ are dictated by principle (3) (Section 1) and form a *consequence relation* defined as follows.

Definition 7 (*Consequence relation*). A ranking κ induces a consequence relation $\mid\!\sim_\kappa$ among wffs, where

$$\phi \mid\!\sim_\kappa \sigma \quad \text{iff} \quad \kappa(\sigma \wedge \phi) < \kappa(\neg\sigma \wedge \phi). \tag{9}$$

Since Δ permits not one but several ranking functions, a straightforward way to define entailment would be to take into consideration all the consequence relations induced by the set of admissible rankings κ with respect to Δ.

Definition 8 (*p-entailment*). Given a consistent Δ, σ is *p-entailed* by ϕ, in the context of Δ, written $\phi \mid\!\sim_p \sigma$, if $\phi \mid\!\sim_\kappa \sigma$ is in every consequence relation $\mid\!\sim_\kappa$ induced by a ranking κ admissible with Δ.

p-entailment is named after the relation proposed by Adams [1], and the equivalent to ε-entailment by Pearl [54] and r-entailment by Lehmann and Magidor [47]. Probabilistically, p-entailment guarantees that conclusions will receive arbitrarily high probabilities whenever the premises receive sufficiently high probabilities (i.e., for every $\varepsilon > 0$ there exists an $\varepsilon' > 0$ such that if $P(\psi_i|\varphi_i) \geqslant 1 - \varepsilon'$ for every $\varphi_i \rightarrow \psi_i \in \Delta$ then $P(\sigma|\phi) \geqslant 1 - \varepsilon$, see [1,33] for details).

p-entailment can be characterized syntactically in terms of the rules of inference provided in [25] or in [47]. p-entailment can also be characterized by the notion of consistency as indicated in Theorem 9.

Theorem 9. *Given a consistent Δ, $\phi \mid\!\sim_p \sigma$ iff $\Delta \cup \{\phi \rightarrow \neg\sigma\}$ is inconsistent.* [8]

It follows immediately from Theorem 9 that entailment can be decided using procedure Consistency-Test (Fig. 2), and that the complexity is the same as testing consistency.

Corollary 10. *Given a consistent $\Delta = \{\varphi_i \rightarrow \psi_i\}$, $1 \leqslant i \leqslant n$, the question of whether $\phi \mid\!\sim_p \sigma$ holds can be determined in $O(n^2)$ propositional satisfiability tests on the material counterpart of Δ.* [9]

Another property of p-entailment is *semi-monotonicity*, that is, monotonicity relative to the addition of rules as distinct from the addition of factual information.

Corollary 11. *If $\Delta' \subseteq \Delta$, and $\phi \mid\!\sim_p \sigma$ given Δ', then $\phi \mid\!\sim_p \sigma$ given Δ.*

The proof is immediate: if $\Delta' \cup \{\phi \rightarrow \neg\sigma\}$ is inconsistent, then $\Delta \cup \{\phi \rightarrow \neg\sigma\}$ must be inconsistent as well. Semi-monotonicity reflects a strategy of extreme caution; no conclusion will ever be issued if it is possible to add rules to Δ (consistently) in such a way as to render the conclusion no longer valid. Thus, p-entailment generates the maximal set of "safe" conclusions that can be drawn from Δ, and hence, was proposed by Pearl [55] as a *conservative core* that ought to be common to all nonmonotonic formalisms.

Like other systems based on conditional logic [52], p-entailment does not properly handle irrelevant features, e.g., from $a \rightarrow c$ we cannot conclude $a \wedge b \mid\!\sim_p c$ even in cases

[8] Theorem 9 was first proven by Adams [1] for defeasible rules and was extended to strict (non-defeasible) rules in [33].

[9] A similar complexity result is given in [47].

where Δ makes no mention of b. To sanction bolder inferences we now define plausibility with respect to a distinguished admissible ranking, κ^z, assigning to each world the lowest possible κ permitted by the constraints imposed by Δ. We will first introduce a syntactic definition of κ^z, define the notion of z-entailment, study its computational properties, and then show that κ^z satisfies the desired minimality condition.

Definition 12 (*The ranking κ^z*). Let $\Delta = \{r_i : \varphi_i \rightarrow \psi_i\}$ be a consistent set of conditional rules. Let $\Delta = (\Delta_0, \Delta_1, \ldots, \Delta_k)$ be the partition that results from applying procedure Consistency-Test on Δ. We define Z to be an ordering on Δ, such that $Z(r_i) = j$ iff r_i belongs to the set Δ_j of the partition. κ^z is defined as follows:

$$\kappa^z(\omega) = \begin{cases} 0, & \text{if } \omega \text{ does not falsify any rule in } \Delta, \\ \max\limits_{i \leqslant n}\{Z(r_i) \mid \omega \models \varphi_i \wedge \neg\psi_i\} + 1, \\ \quad \text{otherwise.} \end{cases} \tag{10}$$

Definition 13 (*z-entailment*). Given a consistent Δ, and the ranking κ^z, we say that ϕ z-entails σ in the context of Δ iff $\phi \mid\!\sim_z \sigma$ is in the consequence relation induced by κ^z, that is iff $\kappa^z(\neg\sigma|\phi) > 0$.

Theorem 14. *Given a consistent $\Delta = \{r_i : \varphi_i \rightarrow \psi_i\}$, $1 \leqslant i \leqslant n$, then:*
(1) *the function Z can be computed in $O(n^2)$ satisfiability tests on the material counterpart of Δ;*
(2) *given Z and a wff ϕ, $\kappa^z(\phi)$ can be computed in $O(\log n)$ satisfiability tests on the material counterpart of Δ.*

Clearly, if the rules in Δ are of Horn form, computing the priority ranking Z and deciding consequences of queries ($\phi \mid\!\sim_z \sigma$) can be done in polynomial time [21]. For the first part of Theorem 14, recall that Z can be computed using procedure Consistency-Test in Fig. 2 as follows: first, identify all rules $r_i : \varphi_i \rightarrow \psi_i$ in Δ for which the formula

$$\varphi_i \wedge \psi_i \bigwedge_{j \neq i, r_j \in \Delta} \varphi_j \supset \psi_j \tag{11}$$

is satisfiable (this corresponds to step 2(a) in Fig. 2). Next, assign to these defaults priority $Z = 0$, remove them from Δ, and repeat the process, assigning to the next set of defaults the priority $Z = 1$, then $Z = 2$, and so on. For the second part, note that once Z is known, the rank κ^z of any wff ϕ is given by

$$\kappa^z(\phi) = \min\left\{ i \mid \phi \bigwedge_{j: Z(r_j) \geqslant i} \varphi_j \supset \psi_j \text{ is satisfiable} \right\}. \tag{12}$$

Thus, $\kappa^z(\phi)$ can be computed by running a binary search on Δ looking to find the lowest $Z(r)$ such that there is a model for ϕ that does not violate any rule r' with priority $Z(r') \geqslant Z(r)$. This is done by dividing Δ into two roughly equal sections: top-half (r_{mid} to r_{high}) and bottom-half (r_{low} to r_{mid}). A satisfiability test on the wff

$\alpha = \phi \bigwedge_{j=\mathrm{mid}}^{n} \varphi_j \supset \psi_j$ decides whether the search should continue (in a recursive fashion) on the bottom-half or top-half.

Eqs. (11) and (12) give a method of constructing a propositional theory $Th(\phi)$ that characterizes precisely the set of conclusions σ that plausibly follow from ϕ (in the context of Δ), that is, $\phi \mathrel{|\!\!\sim}_z \sigma$ iff $Th(\phi) \supset \sigma$. Such a theory is given by the formula

$$Th(\phi) = \phi \bigwedge_{i:\ Z(r_i) \geqslant \kappa^z(\phi)} \varphi_i \supset \psi_i. \tag{13}$$

This is somewhat reminiscent of Brewka's [12] and Poole's [63] idea of constructing preferred subtheories that are maximally consistent with the context ϕ. Here, the construction is more cautious; it stops as soon as all rules of priority $Z \geqslant \kappa^z(\phi)$ are included in the theory.[10] Note, however, that in contrast to Brewka's and Poole's proposals, our priorities are computed automatically from the knowledge base.

Before discussing an illustrative example, note (Theorem 16) that Eq. (10) in Definition 12 defines a unique admissible ranking function κ^z that is minimal in the following sense:

Definition 15 (*Minimal ranking*). A ranking function κ is said to be *minimal* if every other admissible ranking κ' satisfies $\kappa'(\omega) > \kappa(\omega)$ for at least one possible world ω.

Theorem 16 (see [57]). *Every consistent Δ has a unique minimal ranking given by κ^z.*

2.3. Examples

Example 17. Consider the following collection of rules Δ_{pb}:

r_1: "Birds fly" $b \to f$.
r_2: "Penguins are birds" $p \to b$.
r_3: "Penguins do not fly" $p \to \neg f$.
r_4: "Birds have wings" $b \to w$.
r_5: "Animals that fly are airborne" $f \to a$.

It can be readily verified that r_1, r_4, and r_5 are each tolerated by all five rules in Δ_{pb}. For example, the truth assignment

$$\omega' \models \neg p \wedge f \wedge b \wedge w \wedge a$$

satisfies both

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

and

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

[10] Different ways of completing the construction were proposed by Boutilier [9] (see discussion in Section 8).

Table 2
Plausible conclusions for Δ_{pb}, with respect to the notions of p-entailment and z-entailment. A query such as (b, p) means a test for whether $b \mid\!\sim p$ holds in the consequence relations defined by p-entailment and z-entailment (respectively)

Queries	p-entailment ($\mid\!\sim_p$)	z-entailment ($\mid\!\sim_z$)
$(p \wedge b, f)$—"Do penguin-birds fly?"	NO	NO
(b, p)—"Are birds typically penguins?"	NO	NO
$(r \wedge b, f)$—"Do red birds fly?"	undecided	YES
(b, a)—"Are birds airborne?"	undecided	YES
(p, w)—"Are penguins winged animals?"	undecided	undecided

Thus, r_5, r_4 and r_1 are each assigned a Z-label 0 indicating that these rules pertain to the most normal state of affairs. No other rule can be labeled 0 because, once we assign p the truth value $\top$, we must assign $\top$ to b and $\perp$ to f, which is inconsistent with $b \supset f$. The remaining two rules can now be Z-labeled 1, because each one of the two is tolerated by the other.

Examples of plausible consequences one would expect to draw from Δ_{pb} are depicted in Table 2. The first column contains the queries, the second contains p-entailed conclusions, and the last contains z-entailed conclusions. The pair (ϕ, σ) indicates the query: "is σ-entailed given ϕ?" Where a "NO" indicates $\phi \mid\!\sim_p \neg\sigma$ $(\phi \mid\!\sim_z \neg\sigma)$, a "YES" indicates $\phi \mid\!\sim_p \sigma$ $(\phi \mid\!\sim_z \sigma)$, and an "undecided" indicates neither.

We see that z-entailment sanctions plausible inference patterns that are not p-entailed, among them rule chaining, contraposition and the discounting of irrelevant features. For example, we cannot conclude by p-entailment that birds are airborne, $b \mid\!\sim_p a$, because neither $b \rightarrow a$ nor $b \rightarrow \neg a$ would render Δ_{pb} inconsistent. However, a is z-entailed by b, because the rule $b \rightarrow a$ is tolerated by all rules in Δ_{pb} while $b \rightarrow \neg a$ is tolerated by only those Z-labeled 1. Thus, $\kappa^z(b \wedge a) < \kappa^z(b \wedge \neg a)$. Similarly, if r is an irrelevant feature (i.e., not appearing in Δ_{pb}), we obtain $b \wedge r \mid\!\sim_z f$ but not $b \wedge r \mid\!\sim_p f$.

The main weakness of z-entailment is its inability to sanction property inheritance from classes to exceptional subclasses. For example, it will not conclude that penguins have wings $(p \rightarrow w)$ by virtue of being birds (albeit exceptional birds). The reason is that the Z-label 1 assigned to all rules emanating from p amounts to proclaiming penguins an exceptional type of birds in *all* respects, barred from inheriting *any* bird-like properties (e.g., laying eggs, having beaks, etc.). This is a drawback that cannot be remedied by methods based solely on the Z-ordering of defaults; a more refined ordering is required which also takes into account the *number* of rules tolerating a formula, not merely their rank orders. One such refinement is provided by the maximum entropy approach [31] where each model is ranked by the sum of weights on the rules falsified by that model. Another refinement is provided by Geffner's conditional entailment [25], where the rules are partially ordered. These two refinements and other alternatives will be discussed in Section 4.

We now augment the capabilities of system-Z to handle variable-strength rules, thus permitting some defaults to be stated with greater firmness than others.

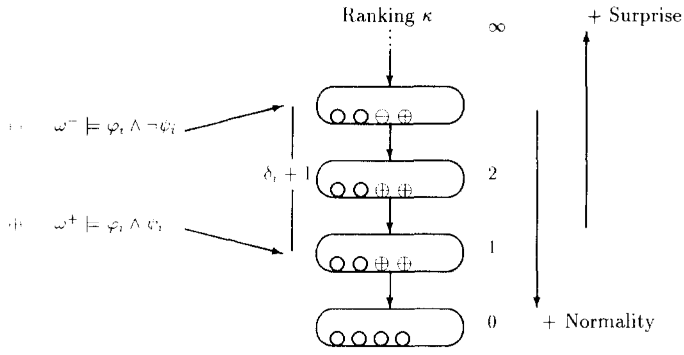

Fig. 3. δ-admissibility constraint imposed by *examplei*.

3. Variable-strength conditionals: system-Z^+

We extend the specification of the rules $\varphi_i \to \psi_i$ with a parameter δ_i representing the *degree of strength* or *firmness* of the rule. Following the previous section, we first study a notion of consistency, and then propose a notion of entailment. Remarkably, the main properties of flat systems are retained after the δ_i are added. Consistency can be tested using the same procedure (Fig. 2) and consequences are generated by the same general formula (Eq. (13)). The computation of the rule priorities, however, requires an additional factor of $O(\log n)$ in complexity.

3.1. Consistency revisited: δ-consistency

Consider a set $\Delta^+ = \{r_i \mid r_i = \varphi_i \xrightarrow{\delta_i} \psi_i,\ 1 \leqslant i \leqslant n\}$, where φ_i and ψ_i are propositional formulas, "$\to$" denotes a default connective as before, and δ_i is a nonnegative integer.

Definition 18 (*δ-consistency*). A ranking κ is said to be *admissible* relative to a given Δ^+, iff

$$\kappa(\varphi_i \wedge \psi_i) < \delta_i + \kappa(\varphi_i \wedge \neg\psi_i) \tag{14}$$

(equivalently $\kappa(\neg\psi_i|\varphi_i) > \delta_i$) for every rule $\varphi_i \xrightarrow{\delta_i} \psi_i \in \Delta^+$. A knowledge base Δ^+ is *consistent* iff there exists an admissible ranking κ relative to Δ^+.

As depicted in Fig. 3, δ-admissibility requires that for each world ω^- satisfying $\varphi_i \wedge \neg\psi_i$ there must be a world ω^+ satisfying $\varphi_i \wedge \psi_i$ such that $\kappa(\omega^+)$ is at least $\delta_i + 1$ ranks less surprising than $\kappa(\omega^-)$.

Let Δ be the flat version of Δ^+; that is, if $\Delta^+ = \{\varphi_i \xrightarrow{\delta_i} \psi_i\}$, then $\Delta = \{\varphi_i \to \psi_i\}$. The next theorem establishes that the δ-consistency of Δ^+ can be decided by applying procedure Consistency-Test on its flat version Δ.

Theorem 19. *Δ^+ is δ-consistent iff Δ is consistent (in the sense of Definition 2).*

It is reassuring to know that once a knowledge base is consistent with respect to one set of assignments to the δ_i, it will be consistent with respect to any such assignment. This means that the rule author has the freedom to modify the firmness of the rules without fear of introducing an inconsistency. We will therefore use the term "consistency" when referring to "δ-consistency".

3.2. Entailment revisited: δ-plausible conclusions

As done in Section 2.2 we will define a consequence relation relative to a unique minimal ranking k^+, which assigns to each world the lowest possible rank permitted by the admissibility constraints of Eq. (14). We will first introduce a syntactic definition of κ^+ (an extension of Definition 12) and then show that it satisfies the desired minimality condition in Theorem 21, which parallels Theorem 16. [11]

Definition 20 (*The ranking κ^+*). Let $\Delta^+ = \{r_i \mid r_i = \varphi_i \xrightarrow{\delta_i} \psi_i\}$ be consistent. κ^+ is defined as follows:

$$\kappa^+(\omega) = \begin{cases} 0, & \text{if } \omega \text{ does not falsify any rule in } \Delta^+, \\ \max\{Z^+(r_i) \mid \omega \models \varphi_i \wedge \neg\psi_i\} + 1, \\ & \text{otherwise,} \end{cases} \tag{15}$$

where $Z^+(r_i)$ is a *priority* ordering on rules, defined by

$$Z^+(r_i) = \min\{\kappa^+(\omega) \mid \omega \models \varphi_i \wedge \psi_i\} + \delta_i. \tag{16}$$

Eqs. (15) and (16) can be viewed as two coupled equations; one defines κ^+ in terms of Z^+, the second defines Z^+ in terms of κ^+. Fig. 4 presents an effective procedure, called Z^+_order, for computing Z^+ from Δ^+. The significance of Eq. (15) is that the priority function Z^+ constitutes an economical encoding, linear in the size of Δ^+, from which the κ^+ of any world ω can be computed in $O(\log|\Delta^+|)$ of satisfiability tests by searching for the highest Z^+ rule violated by ω. The resulting consequence relation $\mid\!\sim_+$ and its associated reasoning procedures are called system-Z^+. Note that if all δ_i are equal to zero, the ranking κ^+ reduces to κ^z (Definition 12).

Theorem 21 establishes the uniqueness and minimality of κ^+, while Theorems 22 and 23 establish the correctness of procedure Z^+_order and its complexity.

Theorem 21. *Every consistent Δ^+ has a unique minimal ranking given by κ^+.*

Theorem 22. *The function Z computed by procedure* Z^+_order *satisfies Definition* 20. [12]

[11] We remark that the notions or falsification and verification of a rule, as well as the notion of tolerance (Definition 3) remain the same; they are not modified by the introduction of the δ-parameters.

[12] Note that Eqs. (17) and (18) correspond to Eqs. (16) and (15) in Definition 20.

Procedure Z^+_order
Input: A consistent knowledge base Δ^+.
Output: Z^+-ordering 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 $r_i = \varphi_i \xrightarrow{\delta_i} \psi_i \in \Delta_0$, 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 Δ^* be the set of rules in $\Delta' = \Delta^+ - \mathcal{R}Z^+$ tolerated by Δ'.
   (b) For each $r : \phi \xrightarrow{\delta} \sigma \in \Delta^*$, let Ω_r denote the set of models for $\phi \wedge \sigma$ that do not violate any rule in Δ'; compute

$$Z(r) = \min_{\omega_r \in \Omega_r} [\kappa(\omega_r)] + \delta \tag{17}$$

where

$$\kappa(\omega_r) = \max_{r_j \in \mathcal{R}Z^+} \{Z(r_j) \mid \omega_r \models \varphi_j \wedge \neg\psi_j\} + 1 \tag{18}$$

and $r_j : \varphi_j \xrightarrow{\delta_j} \psi_j \in \mathcal{R}Z^+$.
   (c) Let r^* be a rule in Δ^* having the lowest Z; set $\mathcal{R}Z^+ = \mathcal{R}Z^+ \cup \{r^*\}$.

End Procedure

Fig. 4. Procedure for computing the Z^+-ordering on rules.

Theorem 23. *Given a consistent* $\Delta^+ = \{r_i \mid r_i = \varphi_i \xrightarrow{\delta_i} \psi_i\}$, $1 \leqslant i \leqslant n$, *the computation of the ranking* Z^+ *requires* $O(n^2 \times \log n)$ *satisfiability tests.*

Two remarks are in order. First, the complexity of the procedure may seem surprising given that Eqs. (17) and (18) are manipulating worlds the number of which grow exponentially. These equations are written this way in the algorithm to show the connection with the equations in Definition 20. Yet, as shown in Lemmas A.5 and A.6, the values needed in Eqs. (17) and (18) are computed by manipulating the *rules* in $\mathcal{R}Z^+$, and not *worlds*.

The second remark concerns step 3(c) in the procedure. It seems that instead of including only one rule in $\mathcal{R}Z^+$ all of the rules in Δ^* should be added. The following example illustrates why doing this may result in a ranking not satisfying the compactness properties of κ^+. Consider the knowledge base

$$\{r_1 : a \xrightarrow{35} b,\ r_2 : c \rightarrow d,\ r_3 : (c \wedge f) \rightarrow \neg d,$$
$$r_4 : a \rightarrow ((\neg b) \vee (c \wedge f \wedge g \wedge \neg d)),\ r_5 : (c \wedge f \wedge g) \rightarrow d\}.$$

If we use the suggestion above, then the sequence by which rules are added to $\mathcal{R}Z^+$ is $[r_1$ and r_2, r_3 and $r_4, r_5]$, and the priorities Z on rules will be: $Z(r_1) = 35$, $Z(r_2) = 0$, $Z(r_3) = 1$, $Z(r_4) = 36$ and $Z^+(r_5) = 2$. On the other hand, if procedure Z^+_order is used, then the sequence by which rules are added to $\mathcal{R}Z^+$ is $[r_1$ and $r_2, r_3, r_5, r_4]$, and the Z^+ are: $Z^+(r_1) = 35$, $Z^+(r_2) = 0$, $Z^+(r_3) = 1$, $Z^+(r_5) = 2$ and $Z^+(r_4) = 3$. By virtue of the difference in the priority of rule r_4, the ranking induced by Z^+ will be more compact than the one induced by Z.

Once Z^+ is known, determining the strength δ with which an arbitrary query σ is confirmed, given the information ϕ, written $\phi \mid\!\sim^{\delta}_{+} \sigma$ requires $O(\log n)$ satisfiability tests: first $\kappa^+(\phi \wedge \sigma)$ and $\kappa^+(\phi \wedge \neg\sigma)$ are computed, using a binary search as in Lemma A.5. Then, these two values are compared and the difference is equated with the strength δ. Clearly, as in the case of a flat knowledge base, if the rules in Δ^+ are of Horn form, computing the priority ranking Z^+ and deciding the plausibility of queries $(\phi \mid\!\sim^{\delta}_{+} \sigma)$ can be done in polynomial time [21], and moreover, Eq. (13) holds.

3.3. Examples

Example 24. Consider again the collection of rules of Section 2.3, augmented with δ's, and denoted here by Δ^+_{pb}:

$$r_1 : b \xrightarrow{\delta_1} f.$$
$$r_2 : p \xrightarrow{\delta_2} b.$$
$$r_3 : p \xrightarrow{\delta_3} \neg f.$$
$$r_4 : b \xrightarrow{\delta_4} w.$$
$$r_5 : f \xrightarrow{\delta_5} a.$$

The Z^+-ordering is computed as follows: Since both r_1, r_4, and r_5 are tolerated by all the rules in the knowledge base, $Z^+(r_1) = \delta_1$, $Z^+(r_4) = \delta_4$, and $Z^+(r_5) = \delta_5$. Any κ^+-minimal world verifying r_2 and r_3 must violate r_1; therefore, following procedure Z^+_order, $Z^+(r_2) = \delta_1 + \delta_2 + 1$ and $Z^+(r_3) = \delta_1 + \delta_3 + 1$.

All the plausible conclusions shown in Table 2 are also in the consequence relation $\mid\!\sim_{+}$ induced by Δ^+_{pb}. As an illustration, consider the conclusion $p \wedge b \mid\!\sim_{+} \neg f$ ("penguin-birds don't fly"), which amounts to $\kappa^+(p \wedge b \wedge \neg f) < \kappa(p \wedge b \wedge f)$. Note that any minimally ranked world ω_1 satisfying $p \wedge b \wedge \neg f$ must violate $r_1 : b \xrightarrow{\delta_2} f$, and thus $\kappa^+(\omega_1) = Z^+(r_1) = \delta_1$. Similarly, any minimally ranked world ω_2 satisfying $p \wedge b \wedge f$ must violate $r_3 : p \xrightarrow{\delta_3} \neg f$, and thus $\kappa^+(\omega_2) = Z^+(r_3) = \delta_3 + \delta_1 + 1$. The preference for r_3 over r_1 is established independently of the initial δ's assigned to these rules. In the knowledge base above, the priority of r_3 ("typically penguins do not fly") was adjusted to $\delta_1 + \delta_3 + 1$, so as to supersede δ_1, the priority of the conflicting rule "typically birds fly". As a result of such adjustments, the relative importance of rules is maintained throughout the system, and compliance with specificity type constraints is automatically preserved. This should come as no surprise since the δ and the Z^+ reflect different considerations. The δ_i in $\varphi_i \xrightarrow{\delta_i} \psi_i$ establishes the relative strength with which I am committed to accept ψ_i in the context of φ_i. The $Z^+(r_i)$ priority, on the other hand, refers to the degree of surprise of finding a world that violates r_i which includes the surprise associated with φ_i: $Z^+(r_i) \geqslant \delta_i + \kappa^+(\varphi_i)$. Thus, while the δ represent individual properties of externally imposed constraints, the Z^+ also represent the interactions among these constraints. The independence between Z^+ and δ in relation to specificity considerations is formalized in the following proposition.

Theorem 25. *Let $r_1 : \varphi \xrightarrow{\delta_1} \psi$ and $r_2 : \phi \xrightarrow{\delta_2} \sigma$ be two rules in a consistent Δ such that*
(1) $\varphi \mathrel{|\!\sim}_p \phi$ *(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 $Z^+(r_1) > Z^+(r_2)$ independent of the values of δ_1 and δ_2.

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

In Section 2.3 was pointed out that $p \mathrel{|\!\sim}_+ w$ ("penguins have wings") is not in the consequence relation induced by κ^z. In the case of Δ^+_{pb} this conclusion is sanctioned whenever δ_4 is set to be bigger than δ_1, reflecting perhaps the intuition that anatomic features (e.g., wings) are more typical than performance characteristics (e.g., flying). This solution to property inheritance however, is not entirely satisfactory. If we add to this new set of rules a class of "birds" which are "wingless", system-Z^+ will conclude that either "penguins have wings" or "wingless birds fly" but not both. The fact that "penguins" are only exceptional with respect to "flying" (and not necessarily with respect to "having wings") is automatically encoded in the Z^+ ranking by forcing $Z^+(r_3)$ to exceed $Z^+(r_1) + \delta_3$ independently of δ_4 (and $Z^+(r_4)$). These independencies among the Z^+ assigned to the rules may be exploited in future proposals as a basis for the formulation of more complex rule interactions, similar to the partial orders among priorities proposed in [25] (see Section 4).

The next example illustrates the use of δ's to establish preferences among defaults when there are no *specificity considerations* available.

Example 26. Consider the set Δ^+_{pq}:

r_1: "quakers are pacifists (with strength δ_1)" $q \xrightarrow{\delta_1} p$.
r_2: "republicans are non-pacifists (with strength δ_2)" $r \xrightarrow{\delta_2} \neg p$.

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 larger than, less than, or equal to δ_2. This is because any model ω_{rqp} for quakers, republicans, and pacifists must violate r_2, and consequently $\kappa^+(\omega_{rqp}) = \delta_2$, while any model $\omega_{rq\neg p}$ for quakers, republicans, and non-pacifists must violate r_1, that is, $\kappa^+(\omega_{rq\neg p}) = \delta_1$. In this case the decision to prefer one world over the other does not depend on specificity considerations but rather on whether the rule encoder believes that religious convictions carry more weight than political affiliations.

4. Related approaches

Lehmann [46] introduced a consequence relation called *rational closure* which extends the inferential power of p-entailment.[13] Rational closure is defined in terms

[13] As mentioned in Section 2.2 and shown in [46], r-entailment and p-entailment are equivalent notions.

of a relation called *more exceptional*, where a wff α is said to be more exceptional than β if

$$\alpha \vee \beta \mathrel{|\!\!\sim_p} \neg\alpha. \tag{19}$$

Based on this relation, Lehmann then used an inductive definition to assign a *degree* to each wff α in the language: $degree(\alpha) = i$ if $degree(\alpha)$ is not less than i and every β that is less exceptional than α has $degree(\beta) < i$. Finally, $\alpha \mathrel{|\!\!\sim} \beta$ was defined to be in the rational closure of Δ iff

$$degree(\alpha) < degree(\alpha \wedge \neg\beta). \tag{20}$$

Goldszmidt and Pearl [32] showed that $degree(\alpha)$ is identical to $\kappa^z(\alpha)$ and, hence, rational closure is semantically equivalent to z-entailment. [14] The difference between the two systems is both conceptual and computational, concerning the role of the Z-ordering. System-Z begins by computing a priority ordering on rules, from which ranking on worlds and formulas can be computed, if needed. Rational closure begins with an inductive definition of degrees of formulas, from which priorities on rule can be computed as a special case, setting $degree(\phi \to \psi) = degree(\phi)$. As we remarked earlier, rule ordering constitutes remarkably efficient encoding of both worlds and formula rankings, the computation of which can be amortized over many queries. This ordering also facilitates a concise characterization of the consequence relation as shown in Eq. (13). It is natural, therefore, to view the Z-ordering as an intrinsic compilation of the knowledge base, and the basis from which inferences commence. [15]

Lehmann and Magidor [47] provide an axiomatic characterization of the rational closure and showed that the rational closure can be obtained by closing the relation of p-entailment under a rule (suggested by David Makinson) called *rational monotony*. Rational monotony permits us to conclude $\alpha \wedge \beta \mathrel{|\!\!\sim} \gamma$ from $\alpha \mathrel{|\!\!\sim} \gamma$ as long as the consequence relation does not contain $\alpha \mathrel{|\!\!\sim} \neg\beta$. Since rational monotony is induced by any admissible ranking, not necessarily the minimal one defined by system-Z, z-entailment can be thought of as an enrichment of p-entailment with properties that are sound in any individual (admissible) ranking function.

A related system was developed by Delgrande [20], albeit from a different perspective, and using a looser definition of ranking. Delgrande based his system on first-order conditional logic [52] and augmented its inferential power using a fixed-point construction to obtain an extension of Δ. A conditional is added to the extension if it is *supported* by what is already in the extension. The final extension constitutes a consequence relation that is similar to that of rational closure.

From the perspective of defeasible reasoning, system-Z^+ extends system-Z and rational closure by providing the user with the power to explicitly set priorities among

[14] Another equivalent system was independently developed by Rott [66], who intended to capture the behavior of counterfactual conditionals.

[15] The origin of the Z-ordering can be traced back to Adams [1] where it is used to build "nested sequences" of confirmable subsets of Δ. A similar construction was also used in [46,47] to prove the co-NP-completeness of the rational closure.

default rules, while simultaneously maintaining a proper account for specificity relations; and, as will be seen in Section 5, system-Z^+ also permits reasoning with imprecise observations. However, it inherits the main deficiency of these two formalisms, namely, the inability to sanction inheritance across exceptional subclasses (see Section 3.3). This difficulty can be overcome by reshaping the ranking function so as to conform to the following principle: A world ω is preferred to ω' if it falsifies a proper subset of the rules falsified by ω'. Three formalisms which embody this principle will be briefly reviewed next, including the maximum entropy approach in [31], Geffner's conditional entailment [25,27], and the proposal by Boutilier in [9].

The maximum entropy (ME) approach initially proposed in [54, Chapter 10] and further developed in [4,31] is motivated by the convention that, unless mentioned explicitly, properties are presumed to be independent of one another; such presumptions are normally embedded in probability distributions that attain the maximum entropy subject to a set of constraints. Given a set Δ of rules and a family of probability distributions that are admissible relative the constraints conveyed by Δ (i.e., $P(\varphi_i \rightarrow \psi_i) \geqslant 1-\varepsilon$, for all $r_i \in \Delta$), we can single out a distinguished distribution $P^*_{\varepsilon,\Delta}$ having the greatest entropy $\sum_{\omega\in\Omega} P(\omega)\log P(\omega)$, and define entailment relative to this distribution by $\phi \mid\!\sim_* \sigma$ iff $\lim_{\varepsilon\to 0} P^*_{\varepsilon,\Delta}(\sigma|\phi) = 1$.

An infinitesimal analysis of the ME approach also yields a ranking function κ^* on worlds, where κ^* can be represented as a set of recursive equations similar to κ^z and κ^+ (Definitions 12 and 20):

$$\kappa^*(\omega) = \begin{cases} 0, & \text{if } \omega \text{ does not falsify any rule in } \Delta, \\ \displaystyle\sum_{\omega\models\varphi_i\wedge\neg\psi_i} [Z^*(r_i)] + 1, & \text{otherwise.} \end{cases} \tag{21}$$

While $\kappa^+(\omega)$ in Section 3 is defined by the maximum priority rule violated in ω, $\kappa^*(\omega)$ depends on the summation of these priorities. This difference has implications for both the computational complexity and the quality of conclusions that these two proposals sanction. Although the procedure for computing the Z^* priorities in the ME approach is very similar to the one presented in Fig. 4 the computation of the Z^* priorities (and the query-answering procedures) has been proven to be NP-hard even for Horn clauses (see [7]). On the other hand, the ME approach allows the sanctioning of inheritance among exceptional subclasses (see [28,31] for further discussion on the advantages and disadvantages of this formalism).

In Geffner's conditional entailment, rather than letting rule priorities dictate a ranking function on models, a partial order on interpretations is induced instead. To determine the preference between ω and ω', we examine the highest priority rules that distinguish between the two, i.e., that are falsified by one and not by the other. If all such rules remain unfalsified in one of the two possible worlds, then this model is the preferred one. Formally, if $\mathcal{F}[\omega]$ and $\mathcal{F}[\omega']$ stand for the set of rules falsified by ω and ω' respectively, then ω is preferred to ω' iff $\mathcal{F}[\omega] \neq \mathcal{F}[\omega']$ and for every rule in $\mathcal{F}[\omega] - \mathcal{F}[\omega']$ there exists a rule r' in $\mathcal{F}[\omega'] - \mathcal{F}[\omega]$ such that r' has a higher priority than r. Thus, a model ω will always be preferred to ω' if it falsifies a proper subset of the rules falsified by ω'.

Another difference in Geffner's proposal is that the rule priority relation is a partial order as well. This partial order is determined by the following interpretation of the rule $\varphi \rightarrow \psi$: if φ is all we know, then, regardless of other rules that Δ may contain, we are authorized to assert ψ. This means that $r : \varphi \rightarrow \psi$ should get a higher priority than any argument (a chain of rules) leading from φ to $\neg\psi$ and, more generally, if a set $\Delta' \subset \Delta$ does not tolerate r, then at least one rule in Δ' ought to have a lower priority than r. In general, we say that a proposition σ is conditionally entailed by ϕ (in the context of a set Δ) if σ holds in all the preferred models for ϕ induced by every priority ordering *admissible* for Δ. Conditional entailment rectifies many of the shortcomings of system-Z, as well as some weaknesses of the entailment relation induced by maximum entropy. 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 [25].

Boutilier [9] proposed a system which combines the priority ordering of the flat version of system-Z^+ (i.e., system-Z) with Brewka's [12] notion of preferred subtheories. Thus, whereas system-Z^+ assigns equal rank to any two worlds that violate a rule r with $Z^+(r) = z$ and no rule of higher Z^+, the proposal in [9] will make further comparisons in terms of rules of lower priority violated in these worlds. In the case of the example discussed in Section 3.3, since any *minimal* world satisfying $p \wedge w$ must violate a proper subset of the rules violated by any *minimal* model for $p \wedge \neg w$, the desired conclusion is certified. These notions are formalized in terms of the modal logic CO^* which is semantically related to the probabilistic interpretation proposed in this paper [8]. Nevertheless, counterintuitive examples to this notion of entailment can still be found in [25,32]. While Boutilier's proposal appears to be simpler than conditional entailment (as it does not require partial orders), its computational effectiveness is yet to be analyzed.

In the next section system-Z^+ is extended to permit for inferences from imprecise observations.

5. Indirect evidence, and imprecise observations

So far, a query was defined as a pair of boolean formulas (ϕ, σ), where ϕ (the *context*) stands for the set of observations at hand and σ (the *target*) stands for the conclusion whose belief we wish to confirm, deny, or assess. A query (ϕ, σ) would be answered in the affirmative if σ was found to hold in all minimally ranked models of ϕ, and the degree of belief in σ would be given by $\kappa(\neg\sigma \wedge \phi) - \kappa(\sigma \wedge \phi)$.

In many cases, however, the queries we wish to answer cannot be cast in this format, because our set of observations is not precise enough to be articulated as a crisp boolean formula ϕ in the language chosen for analysis. Instead, the observations at hand provide merely indirect (or "soft") evidence in favor of ϕ.

Example 27. Assume that we are throwing a formal party and our friends Mary and Bill are invited. However, judging from their previous behavior, we believe "if Mary goes to the party, Bill will stay home (with strength δ)", written $M \xrightarrow{\delta} \neg B$. Now assume that we have some indirect evidence (of strength K) that Mary will go to the party (perhaps because she is extremely well dressed and is not consulting the movie section in the *Times*) and we wish to inquire whether Bill will stay home.

It would be inappropriate to query the system with the pair $(M, \neg B)$, because the context M has not been established beyond doubt. The difference could be critical if we also have some indirect evidence against "Bill staying home", (e.g., he was seen renting a tuxedo). A flexible system should allow the user to assign a separate degree of belief to each component of ϕ and proceed with analyzing its rational consequences. Thus, a query should consist of a tuple like $(\phi_1, K_1; \phi_2, K_2; \ldots; \phi_m, K_m : \sigma)$, where each K_i measures the degree to which the contextual proposition ϕ_i is supported by evidence. [16]

At first glance it might seem that system-Z^+ would automatically provide such a facility through the use of variable-strength rules. For example, to express the fact that Mary seems to be going to the party, we can perhaps add to Δ a *dummy* rule $Obs_1 \xrightarrow{K} M$ (stating that if Mary meets the set of observations Obs_1, then Mary is believed to be going to the party) and then add the proposition Obs_1 to the context part ϕ of the query, to indicate that Obs_1 has taken place.

This proposal has several shortcomings. First, in many systems it is convenient to treat if–then rules as a stable part of our knowledge, unperturbed by observations made about a particular individual or in any specific set of circumstances. This permits us to compile rules into a structure that allows efficient processing over a long stream of queries. Adding query-induced rules to the knowledge base will neutralize this facility.

Second, rules and observations combine differently: The latter should accumulate, the former do not. For example, if we have two rules $a \xrightarrow{\delta_1} c$ and $b \xrightarrow{\delta_2} c$ and we observe a and b, system-Z^+ would believe c to a degree $\max(\delta_1, \delta_2)$. However, if a and b provide two independent reasons for believing c, the two observations together should endow c with a belief that is stronger than any one component in isolation. To incorporate such cumulative pooling of evidence, we must encode the assumption that a and b are conditionally independent (given c), which is not automatically embodied in system-Z^+. [17]

To avoid these complications, the method we propose treats imprecise observations by invoking specialized conditioning operators, unconstrained by a rule's semantics. We distinguish between two types of evidential reports:

(1) Type-J: "All things considered", our current belief in ϕ should *become J*.
(2) Type-L: "Nothing else considered", our current belief in ϕ should *shift by L*.

Example 28. We can illustrate the distinction between the two evidential reports through the party example consisting of the single rule $r_m : M \xrightarrow{2} \neg B$ ("if Mary goes to the party, then Bill will not go"). The resulting ranking is depicted in on the left-hand side of Fig. 5, as can be seen by a trivial application of procedure Z^+_order, yielding

[16] We remark that evidence in this paper is regarded as setting the context of a query and not as a modifier of the knowledge in Δ. Statistical methods for accomplishing the latter task are explored by Bacchus [3].

[17] The assumptions of conditional independence among converging rules is embodied in the formalism of maximum entropy [31], as well as in the causal interpretation of rules introduced in Section 7.

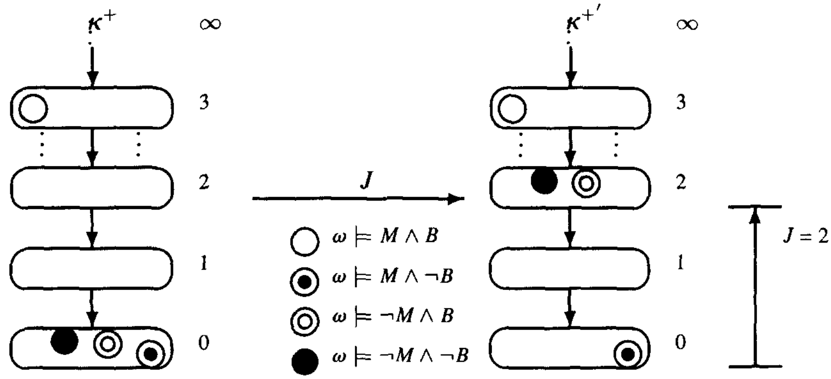

Fig. 5. J-conditioning on the Bill and Mary party example, showing the effect of a strong ($J = 2$) evidence in favor of M.

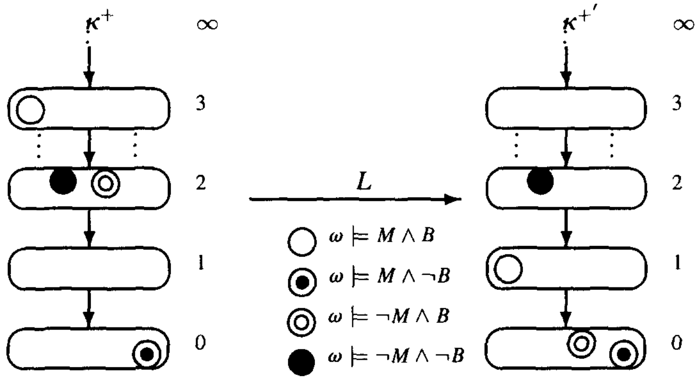

Fig. 6. L-conditioning on the Bill and Mary party example, showing the effect of a strong ($L = 2$) evidence in favor of B.

$Z^+(r_m) = 4$, and Eqs. (5) and (15). We find $\kappa(\omega) = 0$ for every world ω, except for worlds $\omega \models B \wedge M$, for which $\kappa^+(\omega)$ is at least 3. This means that we have no reason to believe that either Mary or Bill will go to the party, but we are pretty sure that both of them will not show up.

Now suppose we see that Mary is very well dressed, and this observation makes our belief in M increase to 2, that is, $\kappa^{+'}(\neg M) = 2$.[18] To conform to this observation, we shift all the $(\neg M)$-worlds upward, relative to those of the M-worlds, by as many increments as required to satisfy the condition $\kappa^{+'}(\neg M) = 2$, and obtain the ranking depicted on the right-hand side of Fig. 5. As a consequence, our belief in Bill staying home also increases to 2 since $\kappa^{+'}(B) = 2$.

[18] Where $\kappa^{+'}$ denotes the revised ranking.

Next, suppose that someone tells us that he has a strong hunch that Bill plans to show up for the party, but fails to tell us why. There are two ways in which this report can influence our beliefs. The natural way (Type-L) would be to assume that our informer has not seen Mary's dress and even might not be aware of Bill and Mary's relationship—hence we assess the impact of his report in isolation and say that whatever the value of our current belief in Bill going, it should increase by two increments, or $L = 2$. The ranking function $\kappa^{+''}$ resulting from this shift is depicted on the right-hand side of Fig. 6, showing $\kappa^{+''}(B) = 0$ and $\kappa^{+''}(\neg M) = 0$, and we are back to the initial ranking of B and of M, except that our disbelief in both Mary and Bill being at the party has decreased to $\kappa^{+''}(M \wedge B) = 1$. A second way would be to assume that our informer is omniscient and having taken into consideration all we know about Bill and Mary, he now instructs us to revise our rankings so that the final belief in "Bill going" will be of strength 2. With this Type-J interpretation, we shift the $\neg B$-worlds upward as many increments as required to establish $\kappa^{+''}(\neg B) = 2$ and obtain (after four increments) $\kappa^{+''}(M) = 2$, thus concluding that Mary will not show up to the party after all.

The transformations dictated by these two interpretations of evidential reports parallel the probabilistic notions of Jeffrey's conditioning (for Type-J) [56] and *virtual conditioning* (for Type-L) [54, p. 44]. We now give a formal definition of these transformations and assess their computational complexity.

5.1. Type-J: all things considered

Let $\kappa'(\omega)$ be a revision of $\kappa(\omega)$ representing evidence, of total strength J, directly supporting a wff ϕ. The Type-J reading of such evidence translates to the condition $\kappa'(\neg\phi) = J > 0$ (and, consequently, $\kappa'(\phi) = 0$). In order to compute $\kappa'(\psi)$ for every wff ψ, we make the assumption that when an agent reports of an observation bearing *directly* on ϕ, such observation does not normally change the *conditional degree of belief* in any propositions conditional on the evidence ϕ or on the evidence $\neg\phi$ [40, 56]. Thus, letting P and P' denote the agent's probability distribution before and after the observation respectively, we have [19]

$$P'(\psi|\phi) = P(\psi|\phi) \quad \text{and} \quad P'(\psi|\neg\phi) = P(\psi|\neg\phi), \tag{22}$$

which leads to Jeffrey's rule [40],

$$P'(\psi) = P(\psi|\phi)P'(\phi) + P(\psi|\neg\phi)P'(\neg\phi). \tag{23}$$

Translated into the language of rankings (using Eqs. (2)–(4)), Eq. (23) yields

$$\kappa'(\psi) = \min[\kappa(\psi|\phi) + \kappa'(\phi); \kappa(\psi|\neg\phi) + \kappa'(\neg\phi)], \tag{24}$$

which offers a convenient way of computing $\kappa'(\psi)$ once we specify $\kappa'(\phi) = 0$ and $\kappa'(\neg\phi) = J$. Eq. (24) assumes an especially attractive form when computing the κ' of a world ω:

[19] Eq. (22) is known as the *J-condition* [56].

$$\kappa'(\omega) = \begin{cases} \kappa(\omega|\phi) + \kappa'(\phi), & \text{if } \omega \models \phi, \\ \kappa(\omega|\neg\phi) + \kappa'(\neg\phi), & \text{if } \omega \models \neg\phi. \end{cases} \tag{25}$$

Eq. (25) corresponds exactly to the α-conditioning proposed in Spohn [71, Definition 6, p. 117], with $\alpha = J$. If $\kappa'(\neg\phi) = \infty$, this process is equivalent to ordinary Bayesian conditioning, since $\kappa'(\omega) = \kappa(\omega|\phi)$ if $\omega \models \phi$ and $\kappa'(\omega) = \infty$ otherwise. Note, however, that in general this conditioning is not commutative; if ϕ_1 and ϕ_2 are mutually dependent (i.e., $\kappa(\phi_2|\phi_1) \neq \kappa(\phi_2)$), the order in which we establish $\kappa(\neg\phi_1) = J_1$ and $\kappa(\neg\phi_2) = J_2$ might make a difference in our final belief state. [20] This is not surprising since in the "all things considered" interpretation the last report is presumed to summarize *all* previous observations.

5.2. Type-L reports: nothing else considered

L-conditioning is appropriate for evidential reports of the type "new evidence was obtained which, by its own merit, would *support* ϕ to degree L". Unlike J-conditioning, the degree L now specifies *change* in the belief of ϕ, not the absolute value of the final belief in ϕ. As in the case of Type-J reports, we assume that in naming ϕ as the direct beneficiary of the evidence, the intent is to convey the assumption of conditional independence, as formulated in Eq. (23). Next, following the method of *virtual evidence* [54, Chapter 2], we assume that the degree of support L quantifies the likelihood ratio $\lambda(\phi)$ associated with some undisclosed observation *Obs*:

$$\lambda(\phi) = \frac{P(Obs|\phi)}{P(Obs|\neg\phi)}, \tag{26}$$

which governs the updates via the product rule

$$\frac{P'(\phi)}{P'(\neg\phi)} = \frac{\lambda(\phi)P(\phi)}{P(\neg\phi)}. \tag{27}$$

Translated into the language of rankings, this assumption yields

$$\kappa'(\phi) - \kappa'(\neg\phi) = \kappa(\phi) - \kappa(\neg\phi) - L \tag{28}$$

and, since either $\kappa'(\phi)$ or $\kappa'(\neg\phi)$ must be zero, we obtain

$$\kappa'(\phi) = \max[0; \kappa(\phi) - \kappa(\neg\phi) - L], \tag{29}$$

$$\kappa'(\neg\phi) = \max[0; \kappa(\neg\phi) - \kappa(\phi) + L]. \tag{30}$$

We see that the effect of L-conditioning is to shift the difference between the ranks of ϕ and $\neg\phi$ by the specified amount L. Once $\kappa'(\phi)$ is known, Jeffrey's rule (Eq. (24)) can be used to compute the $\kappa'(\sigma)$ for an arbitrary wff σ yielding

[20] Spohn [71, p. 118] has acknowledged the desirability of commutativity in evidence pooling but has not stressed that α-conditioning commutes only in a very narrow set of circumstances (partially specified by his Theorem 4). These circumstances require that successive pieces of evidence support only propositions that are relatively independent—the truth of one proposition should not imply a belief in another. Shenoy [67] has corrected this deficiency by devising a commutative combination rule that behaves like L-conditioning.

$$\kappa'(\sigma) = \begin{cases} \min[\kappa(\sigma|\phi) + \kappa(\phi) - \kappa(\neg\phi) - L; \kappa(\sigma|\neg\phi)], \\ \qquad\qquad \text{if } \kappa(\neg\phi) + \kappa(\phi) < L, \\ \min[\kappa(\sigma|\phi); \kappa(\sigma|\neg\phi) + \kappa(\neg\phi) + L - \kappa(\phi)], \\ \qquad\qquad \text{if } \kappa(\neg\phi) + \kappa(\phi) > L, \\ \min[\kappa(\sigma|\phi); \kappa(\sigma|\neg\phi)], \quad \text{if } \kappa(\neg\phi) + \kappa(\phi) = L. \end{cases} \tag{31}$$

This expression takes the following form for $\kappa'(\omega)$:

$$\kappa'(\omega) = \begin{cases} \kappa(\omega|\phi) + \max[0; \kappa(\phi) - \kappa(\neg\phi) - L], & \text{if } \omega \models \phi, \\ \kappa(\omega|\neg\phi) + \max[0; \kappa(\neg\phi) - \kappa(\phi) + L], & \text{if } \omega \models \neg\phi. \end{cases} \tag{32}$$

As in J-conditioning, if $L = \infty$ then $\kappa'(\omega) = \kappa(\omega|\phi)$. For the general case, we can see that the effect of L-conditioning is to shift *downward* the κ of all worlds that are models of the supported proposition ϕ relative to the κ of all worlds that are not models for ϕ. However, unlike J-conditioning, the net relative shift is constant and is equal to L, independent of the initial value of $\kappa(\phi)$. It is easy to verify that L-conditioning is commutative (as is its probabilistic counterpart, see Eq. (27)), and hence it permits iterated belief revision in the case of multiple evidence. Note also that J-conditioning respects evidence independence; if two pieces of evidence support a given proposition ϕ, with strengths L_1 and L_2, then the combined effect is equivalent to shifting the rank of ϕ (relative to $\neg\phi$) by $L_1 + L_2$, as is expected from two independent bodies of evidence.

5.3. *Complexity analysis*

From Eq. (24) we see that $\kappa'(\psi)$ can be computed from $\kappa(\psi|\phi)$ and $\kappa(\psi|\neg\phi)$, which, assuming we have Z^+, requires a logarithmic number of propositional satisfiability tests (see Section 3.2). L-conditioning can follow a similar route (see Eq. (31)).

Special precautions must be taken when simultaneous, multiple pieces of evidence become available. First, J-conditioning is not commutative, hence we cannot simply compute κ' by J-conditioning on ϕ_1 and then J-conditioning κ' on ϕ_2 to get κ''. We must J-condition simultaneously on ϕ_1 and ϕ_2 with their respective J-levels, say J_1 and J_2. Worse yet, an auxiliary effort must be expended to compute the J-level of each combination of ϕ's, in our case $\phi_1 \wedge \phi_2$, $\phi_1 \wedge \neg\phi_2$, etc. This is no doubt a hopeless computation when the number of observations is large.

L-conditioning, by virtue of its commutativity, can process multiple observation by recursive computations. Assume we wish to assess the impact on a sentence ψ of two (undisclosed) pieces of evidence, one supporting ϕ_1 (with strength L_1) and the other supporting ϕ_2 (with strength L_2). We first L-condition κ on ϕ_1 and calculate $\kappa'(\phi_1)$ and $\kappa'(\phi_2)$ using Eq. (30) and (31), respectively. Applying Eq. (31) this time to $\kappa'(\psi \wedge \phi_2)$, we calculate $\kappa'(\psi|\phi_2)$. Second, we L-condition κ' on ϕ_2, compute $\kappa''(\phi_2)$ using Eq. (30), and finally, using $\kappa'(\psi|\phi_2)$ and $\kappa''(\phi_2)$ in Eq. (31) obtain $\kappa''(\psi)$.[21] Note that, although each of these calculations requires only $O(\log|\Delta|)$ satisfiability tests,

[21] The generalization to more than two pieces of evidence is straightforward.

this computation is effective only when we have a well-designated target hypothesis ψ to estimate. The computation must be repeated each time we change the target hypothesis, even when the context remains unaltered. This is because we no longer have a facility for economically encoding a complete description of κ', as we had for κ (using the Z^+-function). Thus, the encoding for κ' may not be as *economical* as that for κ (the number of worlds is astronomical), unless we manage to find dummy rules that emulate the constraints imposed on ϕ_1 by the (undisclosed) observation. Such dummy rules must enforce the conditional independence constraints embedded in Eq. (23), without violating the admissibility constraints (Eq. (7)) in Δ. These dummy rules can be encoded using the stratification mechanism proposed in [35,36] using the causal rule $\neg\phi \xrightarrow{L_1} \neg Obs$.

6. Belief revision and epistemic states

Alchourrón, Gärdenfors and Makinson (AGM) have advanced a set of postulates that have become a standard against which proposals for belief revision are tested [2]. The AGM postulates model beliefs as a deductively closed set K of sentences and constrain how a rational agent should change its belief set K when new information ϕ arrives. The guiding principle behind the AGM postulates is that of minimal change, that is, the new belief set, $K' = K * \phi$, should not differ from K by more than the evidence ϕ requires. The central result of the AGM theory is that the postulates are equivalent to the existence of a complete preordering of all propositions according to their degree of *epistemic entrenchment* such that belief revisions always retain more entrenched propositions in preference to less entrenched ones.

From a computational viewpoint, the existence of such an ordering immediately suggests the existence of an *epistemic state*, supplementing the belief set K, in which the entrenchment ordering is encoded and processed, much like the $\kappa(\phi)$-ordering defined in Sections 3 and 5. However, from an abstract epistemological viewpoint, one can argue that for the purpose of merely characterizing the process of belief revision, we need not concern ourselves with properties of epistemic states outside K; any revision process satisfying the AGM postulates is guaranteed to behave as though propositions were ordered on some scale of entrenchment, thus preserving the principle of minimal change. In this section we will show that implementation as well as characterization of belief revision requires specific reference to a concrete representation of an epistemic state and, moreover, that a system of conditionals (i.e., Δ) constitutes an adequate representation for both functions. We also explicate those aspects of belief revision that would not receive adequate characterization by any AGM-type revision operator.

6.1. Implementational issues

The representational requirements of belief revision are fairly clear: we must find some efficient code for deciding when a given proposition is believed and which beliefs should be given up when new information arrives. In view of the fact that the number of propositions in K is unbounded, the relative entrenchment of propositions must be computed from some finite syntactic code, for example, a ranking function. Indeed, once we specify a single ranking function $\kappa(\omega)$, each proposition is assigned the rank of its lowest ranked world, and we can associate the set of beliefs K with those propositions β for which $\kappa(\neg\beta) > 0$ (see principle (3), Section 1). It follows, then, that K can be represented by any theory ψ whose models have rank 0, that is, $Mods(\psi) = \{\omega \mid \kappa(\omega) = 0\}$. To revise K by a new belief ϕ, we can raise the κ of all models of $\neg\phi$ relative to those of ϕ until $\kappa(\neg\phi)$ becomes $\alpha \geqslant 1$, at which point the newly shifted ranking defines a new set of beliefs $K' = K * \phi$.

This process of belief revision, which Spohn [71, p. 133] called α-conditioning, was presented by Gärdenfors [24] as an example of how belief revision complying with the AGM postulates can be realized. [22] Gärdenfors also suggested that $\kappa(\omega)$ be regarded as the epistemic state supporting K. However, since the number of worlds is still astronomical, identification of an epistemic state with a ranking function, although theoretically feasible, could not serve as a realistic model for human belief revision, nor can it be used in practical reasoning systems. Any realistic representation of "epistemic state" should contain an economical code from which $\kappa(\omega)$ can be computed and, indeed, the analysis of Sections 3 and 5 identifies such a code.

In Sections 3 and 5 we saw that Spohn's process of belief revision can be performed by syntactic operations on the rules in Δ, with Bayes conditioning (hard evidence) corresponding to $\alpha = \infty$, and J-conditioning (soft evidence) corresponding to $\alpha = J$. This suggests that the conditionals residing in Δ provide a sufficient characterization of an epistemic state; no additional information such as ranking function or entrenchment ordering is necessary. In particular, for $\alpha = \infty$, the belief set K coincides with the set of propositions z-entailed by Δ, given all the available evidence ϕ, or, using Eq. (13),

$$\beta \in K \quad \text{iff} \quad Th(\phi) \models \beta \tag{33}$$

where

$$Th(\phi) = \phi \bigwedge_{i:Z(r_i) \geqslant \kappa^z(\phi)} \varphi_i \supset \psi_i, \quad r_i : \varphi_i \to \psi_i \in \Delta. \tag{34}$$

In general, if ϕ_0 stands for some initial evidence supporting K, and ϕ is the new evidence to be incorporated, then the revision of K by ϕ, $K * \phi$, is defined by

$$\beta \in K * \phi \quad \text{iff} \quad Th(\phi_0 \wedge \phi) \models \beta. \tag{35}$$

This syntactic characterization is equivalent to the ranked-based definition of beliefs

$$\beta \in K \quad \text{iff} \quad \kappa_{\phi_0 \wedge \phi}(\neg\beta) > 0 \tag{36}$$

where $\kappa_{\phi_0 \wedge \phi}$ is the *shifted* (or conditional) ranking

$$\kappa_{\phi_0 \wedge \phi}(\omega) = \begin{cases} \kappa_{\phi_0}(\omega) - \kappa_{\phi_0}(\phi), & \text{if } \omega \models \phi, \\ \infty, & \text{otherwise.} \end{cases} \tag{37}$$

[22] In Section 6.2 (footnote 25), we will see that α-conditioning does not, in fact, comply with the AGM postulates, nor with any other postulates of revision operators.

Table 3
Initial ranking for the working student example

κ^+	Possible worlds
0	$(\neg s, a, w)$, $(\neg s, \neg a, w)$, $(\neg s, \neg a, \neg w)$
1	$(\neg s, a, \neg w)$, $(s, a, \neg w)$
2	(s, a, w), $(s, \neg a, \neg w)$, $(s, \neg a, w)$

Table 4
Revised ranking after observing a

κ_a^+	Possible worlds
0	$(\neg s, a, w)$
1	$(\neg s, a, \neg w)$, $(s, a, \neg w)$
2	(s, a, w)

Table 5
Revised ranking after observing s

κ_s^+	Possible worlds
0	$(s, a, \neg w)$
1	(s, a, w), $(s, \neg a, \neg w)$, $(s, \neg a, w)$

The equivalence of these computations is illustrated in the following example.

Example 29. Consider the following collection of rules Δ_s: [23]

r_1: "Typically students don't work" $s \rightarrow \neg w$.
r_2: "Typically students are adults" $s \rightarrow a$.
r_3: "Typically adults work" $a \rightarrow w$.

The Z^+-ordering on the rules (computed according to Eq. (11)) are: $Z^+(a \rightarrow w) = 0$ and $Z^+(s \rightarrow \neg w) = Z^+(s \rightarrow a) = 1$, from which the initial κ^+ ranking can be computed (Eq. (15)), as depicted in Table 3. The rankings in Tables 4 and 23 show the revised rankings after observing $\phi = a$ (e.g., "Joe is an adult") and $\phi = s$ (e.g., "Joe is a student"), respectively. The belief sets associated with these rankings can be computed from the worlds residing in $\kappa^+ = 0$. Thus, the belief set associated with κ_a^+ is the logical closure of the proposition "Joe is a working adult and is not a student", while that associated with κ_s^+ is the logical closure of "Joe is an adult student and does not work".

These belief sets can also be computed using the syntactic characterization of Eq. (13). For example, the theories corresponding to $\phi_0 = true$, $\phi_1 = a$, $\phi_2 = s$, respectively, are given by

$$\begin{aligned} Th(true) &= \{a \supset w, s \supset \neg w, s \supset a\}, \\ Th(a) &= \{a, a \supset w, s \supset \neg w, s \supset a\} \equiv \{\neg s, a, w\}, \\ Th(s) &= \{s, s \supset a, s \supset \neg w\} \equiv \{s, a, \neg w\}. \end{aligned} \tag{38}$$

The two implications in $Th(s)$ mirror the rules $s \rightarrow \neg w$ and $s \rightarrow a$, which are the unique set of rules that are maximally consistent with s.

[23] Note that all δ_i are 0 for this example.

The computational necessity of basing the revision process on a finite set of conditional rules, rather than on the belief entrenchment or ranking functions, has been recognized by several researchers. For example, Nebel [51] adapted the AGM theory so that finite sets of *base* propositions mediate revisions. The basic idea in this syntax-based system is to define a (total) priority order on the set of base propositions and select revisions to be maximally consistent relative to that order, as exemplified in the nonmonotonic systems of Brewka [12] and Poole [64] (and in the example above). Nebel has shown that such a strategy can satisfy almost all the AGM postulates. Boutilier [8] has further shown that the priority function Z^+ does indeed naturally correspond to the epistemic entrenchment ordering of the AGM theory. [24]

However, to fully formalize the practice of belief revision, we also need to specify how the priority order on the base propositions is to be determined. Although one can imagine, in principle, that the knowledge encoder specified this priority order in advance, such specification would be impractical because the order might (and, as we have seen, should) change whenever new rules are added to the knowledge base. By contrast, system-Z^+ extracts both beliefs and rankings of beliefs automatically from the content of Δ; no outside specification of belief orderings is required.

6.2. *Characterization issues*

We return now to the issue of characterization, that is, to whether the process of belief revision can be characterized adequately without specific reference to properties of epistemic states, such as rankings and rules, that are not part of the belief set K. We will show that such epistemic properties must be made explicit in any characterization of belief revision that aims to capture the following cognitive functions:

(1) accommodating variable-weight evidence,
(2) responding to surprising observations (ϕ conflicts with K),
(3) performing iterated belief revisions,
(4) acquiring new conditionals,
(5) dealing with actions and change.

The first three tasks require explicit reference to ranking functions while the last two require, in addition, explicit reference to rules.

The first item has been demonstrated in Section 5. Clearly, any method of handling variable-weight evidence requires comparisons between the weight of evidence at hand and the degree of entrenchment of beliefs to be given up. Since this information is not part of belief sets, it must be supplied by other sources, namely, by ranking functions or some encoding thereof.

Item (2) reflects a major weakness in any operator-based approach to belief revision. Since a revision operator $*\phi$ defines a function from belief sets to belief sets, the resulting belief set $K*\phi$ cannot depend on degrees of disbelief attached to propositions

[24] The proof in [8] considers the priorities Z^+ resulting from a *flat* set of rules as in system-Z [57]. Boutilier [9] also shows that an entrenchment ordering obeying the AGM framework obtains from the Z priorities of the negation of the material counterparts of rules.

outside K.[25] Therefore, two different rankings, say κ_1 and κ_2, having the same belief sets, $K_1 = K_2$ (i.e., $\kappa_1(\omega) = 0$ iff $\kappa_2(\omega) = 0$), must produce the same revised belief set for every ϕ, namely,

$$K_1 * \phi = K_2 * \phi.$$

This carries a disturbing consequence: all background information must be ignored as soon as some surprising evidence arrives, that is, evidence that contradicts the current set of beliefs.[26]

Item (3), on iterated revisions, further emphasizes the difficulties associated with the restriction that $*\phi$ be a function of K only. In the course of responding to a sequence of observations, $\phi_1, \phi_2, \ldots$, it is quite likely that we obtain evidence ϕ_n that discredits one or more of the previous observations. Instead of overriding just the discredited items, as done in J-conditioning, the operator-based characterization would force us to forget all we have seen and start from scratch. The inadequacy of the AGM postulates relative to iterated belief revision is further elaborated in [10, 17], where the difficulties are rectified through the introduction of additional postulates. One of these new postulates [17]:

$$(C_2) \qquad \text{if } \phi_2 \models \neg\phi_1, \text{ then } (K*\phi_1)*\phi_2 \equiv K*\phi_2$$

ensures that later evidence ϕ_2 overrides previous evidence ϕ_1 if the two are logically incompatible. In this formulation, $*\phi$ is not a function but rather a partial function.

We now address item (4), demonstrating that the encoding of epistemic states as sets of conditionals (rather then ranking functions) is necessary for responding not merely to empirical observations but also to linguistically transmitted information in the form of conditional sentences (i.e., if–then rules). For example, suppose someone tells us that, in addition to the story of Example 28, "typically, if a person works, that person is compensated" ($w \rightarrow c$). If our background knowledge is organized as a collection of conditionals, we simply add this new rule to our knowledge base (verifying first that the addition is admissible), recompute Z^+, and are prepared to respond to new observations or hearsay. In a rank-based system, where revisions begin with a given ranking function κ, one cannot properly revise beliefs in response to new conditional sentences, because, to maintain consistency and coherence, such revision must depend not only on the initial ranking but also on the conditional rules that brought about that initial ranking. Two knowledge bases Δ_1 and Δ_2 might give rise to the same ranking function κ^+, yet the new conditional may be consistent with Δ_1 and inconsistent with Δ_2. As an example, consider the sets $\Delta_1 = \{a \rightarrow b\}$ and $\Delta_2 = \{\neg b \rightarrow \neg a\}$. The ranking κ^+ for these knowledge bases is the same (see Table 6). However, the knowledge base $\Delta_2' = \Delta_2 \cup \{a \rightarrow \neg b\}$ is consistent, as shown on the right-hand side of Table 6, while

[25] This problem was brought to our attention by Isaac Levi (see [48]), and was also addressed by Rott [66] and by Boutilier [10].

[26] Although Spohn [71, p. 133] has shown that belief revision conforming to the AGM postulates can be embodied in the context of ranking functions, this requires a major alteration of J-conditioning (Eq. (25)), one that would artificially force K to remain unchanged whenever $\kappa(\neg\phi) = 0$.

Table 6
Ranking κ^+ for $\Delta_1 = \{a \rightarrow b\}$, $\Delta_2 = \{\neg b \rightarrow \neg a\}$, and Δ_2'

κ^+	Δ_1, Δ_2	$\Delta_2' = \Delta_2 \cup \{a \rightarrow \neg b\}$
0	(a,b), $(\neg a,b)$, $(\neg a,\neg b)$	$(\neg a,\neg b)$, $(\neg a,b)$
1	$(a,\neg b)$	$(a,\neg b)$
2	empty	(a,b)

the knowledge base $\Delta_1' = \Delta_1 \cup \{a \rightarrow \neg b\}$ is inconsistent. Clearly, these two situations require different procedures for absorbing the new conditional (see [11]).

The AGM postulates, likewise, are inadequate for characterizing the process of incorporating new conditionals, because they are formulated as transformations on belief sets and are thus oblivious to the set of conditionals that shaped those belief sets, and into which the new conditional is about to join. [27]

The ability to acquire new conditionals (as rules) also provides a simple semantics for interpreting a class of nested conditionals (e.g., "If you wear a helmet whenever you ride a motorcycle, then you won't get hurt badly if you fall off" [28]). Nested conditionals cease to be a mystery when viewed as operations of conditional knowledge bases default rules. The sentence "if $(a \rightarrow b)$ then $(c \rightarrow d)$" is interpreted as

> If I add the default $a \rightarrow b$ to my current Δ, then the conditional $c \rightarrow d$ will be satisfied by the consequence relation $\mid\!\sim_+$ of the resulting knowledge base $\Delta' = \Delta \cup \{a \rightarrow b\}$.

Clearly, such assertions can be given unambiguous tests in the language of default-based ranking systems. Note the essential distinction between having a conditional sentence $a \rightarrow b$ explicitly in Δ and having a conditional sentence $a \rightarrow b$ *satisfied* by the consequence relation $\mid\!\sim_+$ of Δ. In both cases the conditional $a \rightarrow b$ would meet the Ramsey test, but only the former case would resist the adoption of the conditional $a \rightarrow \neg b$, and would trigger a more drastic restructuring of the knowledge base in the spirit of [11]. This distinction gets lost in systems that do not acknowledge defaults as the basis for ranking and beliefs. [29]

The last item on the list, concerning actions and changes, will be discussed more fully in Section 7 where, again, it will be shown that an explicit reference to the rules is required. In this context, the function of each rule is to identify a group of variables that are tied together by a stable mechanism and remain invariant to actions operating on neighboring mechanisms. It is only through such grouping that we can localize the effect of actions and predict their causal ramifications; the ranking function in itself provides no information to support such predictions.

In summary, belief revision by either indirect evidence, surprising observations, or sequences of observations cannot be characterized in terms of a transformation on

[27] Gärdenfors [24, pp. 156–160] has shown that any attempt to characterize the acquisition of conditionals using a revision operator would fail by rendering the revision trivial.

[28] Judea Pearl attributes this example to Philip Calabrese (personal communication).

[29] Belief revision systems proposed in the database literature [13,23] suffer from the same shortcoming. In that context, defaults represent integrity constraints with exceptions.

belief sets; these operations require knowledge of the ranking function. In addition, characterizing the acquisition of new conditionals and the effects of actions requires further knowledge of the rules that have shaped the ranking function. This suggests that conditional sentences form the basic building blocks of one's epistemic state, and that beliefs and degrees of beliefs in propositional sentences follow as natural byproduct of those building blocks. This is to be expected, since conditionals carry generic (hence stable) domain knowledge, while unconditional propositions carry transitory factual information [25,63].

7. Causal relations and actions

Problems associated with representing causal relationships plague almost every proposal for default reasoning [37]. For example, approaches such as circumscription [50] and default logic [65], which are based on extensions to classical logic, fail to block the chaining of the following default expressions [53]:

r_1: "If the grass is wet, then conclude it rains".

r_2: "If this bottle leaks, the grass will get wet".

Finding the bottle leaking, we do not wish to conclude from these two rules that it rains, nor that it will rain.

Approaches based on conditional interpretation of causal rules also produce undesirable effects. Consider, for example, the following two rules, both pointing from cause to effect and both containing a voluntary act in their antecedents:

r_3: "If the ignition key is turned on, the car will start" $(tk \rightarrow cs)$.

r_4: "If the ignition key is turned on and the battery is dead, the car will not start" $(tk \wedge bd \rightarrow \neg cs)$.

From these two rules, all the entailment relationships discussed in Sections 3 and 4 produce the following pair of inferences:

$$tk \mid\!\sim \neg bd,$$

$$\neg tk \mid\!\not\sim \neg bd.$$

Taken together, these inferences suggest some strange connection between the act of turning the ignition key and the state of the battery: the battery is believed to be OK when we turn the ignition key on, but becomes suspect of being defective each time we turn the ignition key off. Such behavior is counterintuitive, because rules r_3 and r_4 intend merely to describe the response of the car to various combinations of key/battery conditions, and were not meant to imply any dependency among those conditions.

In general, if our understanding of the relationships between the ignition key and the battery is encoded in some knowledge base Δ, we certainly do not expect that adding r_3 and r_4 to Δ would modify those relationships. If the battery state is presumed to be independent from the ignition key prior to specifying rules r_3 and r_4, then it ought to remain independent from the ignition key after specifying those rules. This is a prevailing general pattern of causal reasoning: contemplating possible developments of future events should not affect our belief in past and present events. We call this pattern *modularity*, and it follows from the Markov property of causal organization, simply stated as: the occurrence of each event is independent of all past events, once we determine the direct causes of the event [54, 62, 70, 72]. The Markov property, which will be defined formally in Definition 32, together with the assumption that the rules in Δ specify *all* the causal relationships among the events of interest, will next be used to induce a new entailment relation, closely reflecting the common interpretation of causal knowledge bases. In the example above, the Markov property dictates that the relationship between tk and bd be determined solely by antecedents of rules leading to bd and tk and, if cs is not mentioned in any of these rules, that relationship will not be altered by invoking rules r_3 and r_4.

To incorporate the Markov property into the language of ranking functions, we will assume that all rules in Δ are causal, that is, each antecedent describes an event that is understood to have a direct causal influence on the consequent event. (Section 7.2 provides a definition of causal influence in terms of hypothetical interventions.) Additionally, we will assume a finite language $\mathcal{L}$, having n atomic literals just as we did in Section 2. Given these assumptions it is convenient to characterize the rules in Δ in terms of a graph $\Gamma(\Delta)$:

Definition 30 (*Networks*). Let Δ be a set of causal rules, and let $\Gamma(\Delta)$ be a graph where each node in $\Gamma(\Delta)$ corresponds to atomic symbols $x_1, x_2, \ldots$ in $\mathcal{L}$, and in which an arc is drawn from node x_i to node x_j just in case there is a rule in Δ such that x_i or $\neg x_i$ appears in the antecedent and x_j or $\neg x_j$ appears in the consequent. If the resulting graph is a directed acyclic graph (DAG), we will call $\Gamma(\Delta)$ a *causal network* (or *network* for short).[30]

For convenience in this section we restrict the rules to be of the form $x_1 \wedge \cdots \wedge x_m \rightarrow x_n$.[31] Note that being a truth-value assignment to propositions x_i ($1 \leqslant i \leqslant n$) in $\mathcal{L}$, any world ω can be represented by a conjunction $l_1 \wedge \cdots \wedge l_n$, $l_i \in \{x_i, \neg x_i\}$, of literals. If $x_r, \ldots, x_s$ are the parents of x_t in $\Gamma(\Delta)$, then the set $\{x_r, \ldots, x_s\}$ is called the *parent set* of x_t and will be represented by π_t. Intuitively, the parent set of x_t represents all the direct causes for x_t. Fig. 7 depicts the DAG Γ for the network containing rules r_3 and r_4 in the car-starting example, with $\pi_{cs} = \{tk, bd\}$ and $\pi_{tk} = \pi_{bd} = \{\}$ (the empty set).

To avoid excessive notation, we will use the symbols $x_1, x_2, \ldots, x_n$ both as propositional symbols and as variable names. It is understood, though, that whenever these

[30] If all rules in Δ are causal, it is reasonable to assume that $\Gamma(\Delta)$ will be indeed a DAG.

[31] The form $x_1 \wedge \cdots \wedge x_m \rightarrow x_n$ does not restrict the development of this paper but it clarifies the exposition. A causal rule may take on the general form $\alpha(x_1, \ldots, x_m) \rightarrow \beta(y_1, \ldots, y_n)$ where α and β are any boolean formulas. Any $\alpha(x_1, \ldots, x_m)$ can be simulated by a set of simpler rules, each containing a conjunction of atomic antecedents. Moreover, any rule $\alpha(x_1, \ldots, x_m) \rightarrow \beta(y_1, \ldots, y_n)$ can be represented by the following set of rules: $\alpha(x_1, \ldots, x_m) \rightarrow d'$, $\beta(y_1, \ldots, y_n) \Rightarrow d'$, and $\neg\beta(y_1, \ldots, y_n) \Rightarrow \neg d'$, where d' is a dummy variable and $\Rightarrow$ is a *strict conditional*, as defined in [33].

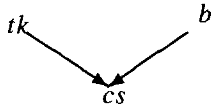

Fig. 7. Graph Γ for the network $\{tk \to cs, tk \wedge bd \to \neg cs\}$.

symbols appear in an equation, the intent is to assert a *set* of equations, one for every truth-value assignment that the propositional symbols can take. For example, the equality

$$k(x_i|x_j) = k(x_i)$$

stands for four equalities,

$$\begin{aligned} &k(x_i|x_j) = k(x_i), \qquad && k(\neg x_i|x_j) = k(\neg x_i), \\ &k(x_i|\neg x_j) = k(x_i), && k(\neg x_i|\neg x_j) = k(\neg x_i). \end{aligned}$$

To account for the distinct character of causal conditionals, we will impose the Markovian property on the admissible ranking functions. The Markov property can best be imposed through a process of stratification—the ranking is constructed in layers such that the addition of each new layer does not introduce new dependencies among variables in existing layers, similar to the construction of Bayesian networks [54].

Definition 31 (*Stratified rankings*). A ranking function $\kappa(\omega)$ is said to be *stratified* relative to a DAG $\Gamma(\Delta)$ if

$$\kappa(\omega) = \sum_i \kappa(x_i(\omega)|\pi_i(\omega)) \tag{39}$$

where π_i are the parents of x_i in Γ and $x_i(\omega)$ and $\pi_i(\omega)$ are both evaluated according to ω.[32]

Given a DAG Γ, it is easy to construct a ranking function stratified relative to Γ; for each parents–son family (X_i, π_i) in Γ, we assign an arbitrary integer-valued function $f_i(x_i, \pi_i)$, such that if $\min_{x_i} f_i(x_i, \pi_i) = 0$ then we sum up these functions over i. It is also straightforward, in principle, to check whether a given ranking function $\kappa(\omega)$ is stratified relative to a given graph Γ. Using Eqs. (5) and (6), we compute from $\kappa(\omega)$ the terms $\kappa(x_i(\omega)|\pi_i(\omega))$, $i = 1, \dots n$, we form the sum in (39) and check for equality.

Any ranking function satisfying the decomposition in Eq. (40) also satisfies a desirable modularity property: If we arrange the literals along an order that agrees with the directionality of the arrows in Γ, then the ranking associated with the first j literals $\kappa(x_1, \dots, x_j) = \sum_{i=1}^{j} \kappa(x_i|\pi_i)$ is not constrained by any rule residing outside the subgraph corresponding to those literals. Moreover, in a stratified ranking, each parents–son

[32] For every proposition ϕ, $\phi(\omega)$ is defined by $\phi(\omega) \triangleq \phi$, if $\omega \models \phi$, $\phi(\omega) \triangleq \neg\phi$ if $\omega \models \neg\phi$.

Table 7
Stratified ranking κ^s and minimal ranking κ^+ for the network $\{tk \to cs, tk \wedge bd \to \neg cs\}$

κ	Worlds in stratified rank κ^s	Worlds in κ^+ rank
0	$(\neg bd, tk, cs)$, $(\neg bd, \neg tk, \neg cs)$, $(\neg bd, \neg tk, cs)$	$(\neg bd, tk, cs)$, $(\neg bd, \neg tk, \neg cs)$, $(\neg bd, \neg tk, cs)$, $(bd, \neg tk, cs)$, $(bd, \neg tk, \neg cs)$
1	$(bd, tk, \neg cs)$, $(bd, \neg tk, \neg cs)$, $(\neg bd, tk, \neg cs)$, $(bd, \neg tk, cs)$	$(bd, tk, \neg cs)$, $(\neg bd, tk, \neg cs)$
2	(bd, tk, cs)	(bd, tk, cs)

relationship is treated as an independent autonomous process, since each variable is *conditionally independent* of all its non-descendants, given its parents

$$\kappa(x_i|x_{i-1} \wedge \cdots \wedge x_1) = \kappa(x_i|\pi_i), \tag{40}$$

that is, the degree of surprise of an event x_i given all *prior* events (according to the ordering mentioned above) must be equal to the degree of surprise of x_i given just those events that directly cause x_i, namely, its parent set. This ensures that rule violations that occur in different families accumulate surprise, in much the same way that the simultaneous occurrence of independent errors becomes less likely as the number of errors increases.

The summation in Eq. (39) parallels the product decomposition of Bayesian networks [54]

$$P(x_n, \ldots, x_1) = \prod_{i=1}^{n} P(x_i|\pi_i), \tag{41}$$

which embodies the probabilistic version of the Markov property: the parent set of any given proposition x_i renders x_i *probabilistically independent* of all its predecessors (in the given ordering set by the DAG Γ). Stratified rankings can in fact be regarded as an order-of-magnitude abstraction of Bayesian networks, where numerical probabilities are replaced by integer-valued levels of surprise (κ), addition is replaced by minimization, and multiplication is replaced by addition.[33]

In addition to stratification, given a network Δ, the rankings of interest should also be admissible with respect to Δ, that is (see Definition 2): $\kappa(\neg\psi_i|\varphi_i) > 0$ for every rule $\varphi_i \to \psi_i \in \Delta$. The following example illustrates the difference between the requirements of minimality and stratification.

Example 32. Table 7 contrasts a stratified ranking κ^s with the minimal ranking κ^+ for the network $\Delta = \{tk \to cs, tk \wedge bd \to \neg cs\}$ in Fig. 7. Consider the world $\omega_1 \models \neg tk \wedge bd \wedge \neg cs$. The stratified ranking κ^s assigns to ω_1 the ranking $\kappa^s(\neg tk \wedge bd \wedge \neg cs) = 1$, while the minimal ranking κ^+ assigns $\kappa^+(\neg tk \wedge bd \wedge \neg cs) = 0$. Thus, κ^s is not minimal. In contrast, it is easy to verify that κ^+ is not stratified: $\kappa^+(bd) + \kappa^+(tk) = 0$,

[33] An even coarser abstraction of Eq. (41) in the context of relational databases can be found in [19], where the stratification condition is imposed on relations and then used in finding backtrack-free solutions for constraint satisfaction problems.

but $\kappa^{+}(bd \wedge tk) = 1$ in clear violation of Eq. (39), which dictates $\kappa(bd \wedge tk) = \kappa(bd) + \kappa(tk)$ (since $\pi_{bd} = \pi_{tk} = \{\}$). We see that the counterintuitive behavior shown in the beginning of this section emanates from a spurious dependency that k^{+} induces between bd and tk. This dependency is non-existent in stratified rankings.

In the next section we introduce a notion of entailment tied to stratified rankings and show formally that stratification endows this entailment relation with Markov and modularity properties (see Theorems 34 and 35 below).

7.1. c-entailment: consequences of causal rules

Definition 33 (*c-entailment*). Given a knowledge base Δ of causal rules, σ is *c-entailed* by ϕ in the context of Δ, written $\phi \mid\!\mid\!\!\sim_c \sigma$, if $\kappa_s(\sigma \wedge \phi) < \kappa_s(\neg\sigma \wedge \phi)$ in every ranking κ_s *stratified* relative to $\Gamma(\Delta)$, and *admissible* with respect to Δ.

The expression $\phi \mid\!\mid\!\!\sim_c \sigma$ is not to be interpreted as stating that ϕ is believed to *cause* σ but rather that σ is a plausible conclusion of ϕ under the causal interpretation given to the conditionals in Δ.

Theorems 34 and 35 illustrate how the Markov property imposed by the stratification (Definition 32) shapes the entailment relationship introduced in Definition 33.

Theorem 34. *Let $\{x_r, \ldots, x_s\}$ be of the parent set of x in $\Gamma(\Delta)$. Let $\mathcal{Y} = \{y_1, \ldots, y_m\}$ be a set of atomic propositions such that no $y_i \in \mathcal{Y}$ is a descendant of x in $\Gamma(\Delta)$, and let ϕ_Y be any wff built only with elements from $\mathcal{Y}$.*[34] *If $x_r \wedge \cdots \wedge x_s \mid\!\mid\!\!\sim_c x$, then $\phi_Y \wedge x_r \wedge \cdots \wedge x_s \mid\!\mid\!\!\sim_c x$.*

As a corollary to Theorem 34 it is easy to see that c-entailment is insensitive to *irrelevant* propositions. In particular, the sets of consequences induced by two disconnected networks will be independent of each other.

Theorem 35. *Let $\Gamma'(\Delta')$ be a subgraph of $\Gamma(\Delta)$ such that if x' is a node in $\Gamma'(\Delta')$ then all the rules in Δ with x' (or $\neg x'$) as their consequent are also in Δ'. Let φ and ψ be two wffs built with propositions corresponding to nodes in $\Gamma'(\Delta')$. If $\varphi \mid\!\mid\!\!\sim_c \psi$ in the context of Δ', then $\varphi \mid\!\mid\!\!\sim_c \psi$ in the context of Δ.*

Theorem 35 reflects the modularity of causal knowledge bases as it permits us to add rules to the network while preserving all consequences derivable from the original subnetwork.

Entailment relations are often characterized using a set of axioms, which can also be used as inference rules. For example, the following axioms completely characterize p-entailment [1,25,43]:

(1) (*Defaults*) If $\varphi \rightarrow \psi \in \Delta$, then $\varphi \mid\!\mid\!\!\sim_c \psi$.

(2) (*Deduction*) If $\models \varphi \supset \psi$, then $\varphi \mid\!\mid\!\!\sim_c \psi$.

[34] We also require that $\phi_Y \wedge x_r \wedge \cdots \wedge x_s$ be satisfiable.

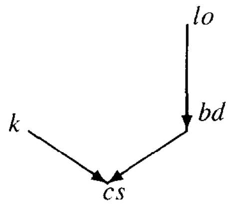

Fig. 8. Network Γ for the causal conditionals $\{k \to cs, k \wedge bd \to \neg cs, lo \to bd\}$.

(3) (*Augmentation*) If $\varphi \mathrel{|\!\!\approx_c} \psi$ and $\varphi \mathrel{|\!\!\approx_c} \gamma$, then $\varphi \wedge \gamma \mathrel{|\!\!\approx_c} \psi$.
(4) (*Cut*) If $\varphi \mathrel{|\!\!\approx_c} \gamma$ and $\varphi \wedge \gamma \mathrel{|\!\!\approx_c} \psi$, then $\varphi \mathrel{|\!\!\approx_c} \psi$.
(5) (*Disjunction*) If $\varphi \mathrel{|\!\!\approx_c} \psi$ and $\gamma \mathrel{|\!\!\approx_c} \psi$, then $\varphi \vee \gamma \mathrel{|\!\!\approx_c} \psi$.
(6) (*And*) If $\varphi \mathrel{|\!\!\approx_c} \psi$ and $\varphi \mathrel{|\!\!\approx_c} \gamma$, then $\varphi \mathrel{|\!\!\approx_c} \psi \wedge \gamma$.

Since the set of stratified rankings for a given network $\Gamma(\Delta)$ is a subset of the admissible rankings for Δ, these six axioms must also be sound with respect to c-entailment. The requirement of stratification allows us to augment this set with an additional rule, which we call *Markov*, that takes advantage of the structure of the graph Γ, and essentially follows Theorem 34.[35]

(7) (*Markov*) If $x_r \wedge \cdots \wedge x_s \mathrel{|\!\!\approx_c} x$, then $\phi_Y \wedge x_r \wedge \cdots \wedge x_s \mathrel{|\!\!\approx_c} x$ whenever
- $\{x_r, \ldots, x_s\}$ is the parent set of x in $\Gamma(\Delta)$, and
- ϕ_Y is a wff built only with elements from $\{y_1, \ldots, y_m\}$ such that no $y_i \in \mathcal{Y}$ is a descendant of x in $\Gamma(\Delta)$.[36]

We conjecture that inference rules (1)–(7) constitute a *complete* system of inference with respect to stratified rankings in the following sense: given a set of rules Δ (and the corresponding graph $\Gamma(\Delta)$), if $\varphi \mathrel{|\!\!\approx_c} \psi$ holds in every stratified ranking with respect to Δ, then $\varphi \mathrel{|\!\!\approx_c} \psi$ can be derived by the successive application of rules (1)–(7).

Example 36. Consider rules r_3 and r_4, introduced at the beginning of this section, augmented with a third causal rule (r_5) (see Fig. 8):

r_3: "If I turn the ignition key, the car will start" ($tk \to cs$).
r_4: "If I turn the ignition key and the battery is dead, the car will not start" ($tk \wedge bd \to \neg cs$).
r_5: "If I leave the headlights on for 12 hours the battery is dead" ($lo \to bd$).

Intuitively, these three rules should be sufficient to draw the conclusion that, if we left the headlights on for 12 hours and then turn the ignition key, the car will not start (i.e., $lo \wedge tk \mathrel{|\!\!\approx_c} \neg cs$). Yet, all the entailment relationships discussed in Sections 3 and 4 will allow the unintended scenario in which the car engine actually starts and the battery

[35] Note that Theorem 34 establishes the soundness of the Markov rule.
[36] As in the case of Theorem 34, we also require that $\phi_Y \wedge x_r \wedge \cdots \wedge x_s$ be satisfiable.

is not dead after all.[37] The stratification required by c-entailment assigns (at least) a unit degree of surprise ($\kappa = 1$) to this scenario (since it violates rule r_5), and zero surprise ($\kappa = 0$) to the intuitive scenario (which does not violate any rule, as if bd is "explained" by lo). A syntactic derivation of $lo \wedge k \mid\!\mid\!\!\sim_c \neg cs$ using inference rules (1)–(7) is given in the Appendix (see Proposition A.7). Rule (7) (Markov) allows the inference $lo \wedge tk \mid\!\mid\!\!\sim_c bd$ since tk is not a descendant of bd in $\Gamma(\Delta)$ and lo is its parent (see Fig. 8). Similarly, we can use the Markov rule for inferring $lo \wedge tk \wedge bd \mid\!\mid\!\!\sim_c \neg cs$ from $tk \wedge bd \mid\!\mid\!\!\sim_c \neg cs$, since tk and bd are the only direct causes for cs (i.e., its parent set) and lo is a non-descendant of cs.

The next example presents a simple abduction (or backward projection) problem and serves to contrast the behavior of c-entailment with that of chronological minimization [69].

Example 37. Consider a sequence of rules $\{l_0 \to l_1, l_1 \to l_2, \ldots, l_{n-1} \to l_n\}$ standing for the various instances of "if a gun is loaded at time t_i, then it is expected to remain loaded at time t_{i+1}" ($0 \leqslant i < n$). Given that the gun is loaded at t_0 and that it is found unloaded at time t_n (i.e., $l_0 \wedge \neg l_n$ is true), the scheme of chronological minimization will favor the somewhat counterintuitive inference that the gun remained loaded until t_{n-1} (i.e., $l_1 \wedge \cdots \wedge l_{n-1}$ is true). c-entailment, however, only yields the weaker, but intuitive, conclusion that the gun must have been unloaded at some time between t_1 and t_{n-1}, i.e., $\neg(l_1 \wedge \cdots \wedge l_n)$, but it is unable to pinpoint the precise time of the unloading (see Proposition A.8 in the Appendix).[38] On the one hand, c-entailment and chronological minimization are expected to yield the same conclusions in problems of pure prediction, since enforcing ignorance of future events is paramount to the principle of modularity, which is inherent in c-entailment. On the other hand, they differ in tasks of abduction, as demonstrated in this example.

Computationally, we can take advantage of the relation between stratified rankings and probability distributions represented in Bayesian networks, to compute default conclusions from stratified rankings. Hunter [39], for example, has shown that the polytree algorithm for probabilistic belief update [54] can be modified slightly[39] and become applicable to a network of belief quantified with a stratified ranking. This result generalizes to our formalism in the following way: the polytree algorithm for computing probabilistic belief revision on Bayesian networks, including its variants based on joint trees and cutset conditioning [54], can be used to compute c-entailment as long as Δ defines a unique stratified ranking. Fortunately, the condition of stratification allows us to specify a unique ranking modularly: if each family composed of a node x and its parents π_x specifies a conditional ranking $\kappa(x|\pi_x)$ for all instantiations of x and π_x, their resulting ranking (Eq. (39)) is guaranteed to be unique and stratified.

[37] This problem is reminiscent of the one pointed out by Hanks and McDermott regarding causality in the Yale Shooting Problem (YSP) [37]. In fact, rules r_3–r_5 are isomorphic to the YSP [25].

[38] This example is isomorphic to the "stolen car problem" [42].

[39] Namely, the replacement of multiplication by addition, and summation by minimization.

Table 8
Two minimal stratified rankings for the network $\{a \to c, b \to \neg c\}$

κ	Rank 1	Rank 2
0	$(\neg a, b, \neg c)$, $(\neg a, \neg b, c)$, $(\neg a, \neg b, \neg c)$	$(a, \neg b, c)$, $(\neg a, \neg b, c)$, $(\neg a, \neg b, \neg c)$
1	$(\neg a, b, c)$, (a, b, c), $(a, \neg b, c)$	$(\neg a, b, \neg c)$, $(a, b, \neg c)$, $(a, \neg b, \neg c)$
2	$(a, b, \neg c)$, $(a, \neg b, \neg c)$	(a, b, c), $(\neg a, b, c)$

In the case where the specification of Δ does not constitute a unique ranking, techniques for rank completion such as the one resulting in the Z^+ ranking of Section 3 can be adopted. The difference is that stratified rankings allow us to perform the completion locally, on a family by family basis. Yet, contrary to the case studied in Section 3 the stratified completion of a set Δ may not result in a *unique minimal*[40] ranking for every Δ. Consider for example the following network $\Delta = \{a \to c, b \to \neg c\}$. This set admits the two minimal stratified rankings depicted in Table 8.

In [29] we study a class of networks, which we call *stratifiable*, that admits unique minimal rankings. We also provide an effective procedure to build this ranking. In essence, Δ will be stratifiable if it allows recursive construction of a minimal stratified ranking from the minimal stratified rankings that are admissible with respect to (ordered) subsets of Δ. Thus, let $\{x_1, \ldots, x_n\}$ be an ordered set of the variables in Δ according to $\Gamma(\Delta)$.[41] Let $(\Delta_1, \ldots, \Delta_n)$ be an ordered partition of Δ, such that Δ_i contains all the rules that have x_i (or $\neg x_i$) as their consequence. In a stratifiable network, the minimal ranking for $\bigcup_{i=1}^{j} \Delta_i$, namely $\kappa_s^*(x_j \wedge \cdots \wedge x_1)$, is built from the minimal ranking for $\bigcup_{i=1}^{j-1} \Delta_i$ $(\kappa_s^*(x_{j-1} \wedge \cdots \wedge x_1))$, and a minimal conditional (and admissible) ranking for Δ_j following Eq. (39):

$$\kappa_s^*(x_j \wedge \cdots \wedge x_1) = \kappa'(x_j | \pi_j) + \kappa_s^*(x_{j-1} \wedge \cdots \wedge x_1). \tag{42}$$

As an example, the network $\Delta = \{tk \to cs, tk \wedge bd \to \neg cs\}$ is not stratifiable. The reason is that its minimal (and unique) ranking, depicted in Table 7, is not minimal for $\Delta_{tk} \cup \Delta_{bd}$. The minimal stratified ranking for $\Delta_{tk} \cup \Delta_{bd}$ is $\kappa(tk \wedge bd) = \kappa(\neg tk \wedge bd) = \kappa(tk \wedge \neg bd) = \kappa(\neg tk \wedge \neg bd) = 0$,[42] whereas the ranking in Table 7 has $\kappa(tk \wedge bd) = \kappa(\neg tk \wedge bd) = 0$, and $\kappa(tk \wedge \neg bd) = \kappa(\neg tk \wedge \neg bd) = 1$. The network $\Delta' = \{true \to \neg bd, tk \to cs, tk \wedge bd \to \neg cs\}$ is, on the other hand, stratifiable. Its minimal (and unique ranking) is the one depicted in Table 7. Note that the only difference between the networks Δ and Δ' is the rule $true \to \neg bd$. We remark that networks other than the class of stratifiable networks admit minimal (and unique) stratified rankings, yet, stratifiable networks allow for the recursive construction of such ranking. Automatic procedures for completions of arbitrary networks (including the addition of rules to make a network stratifiable) is a subject of current research.

[40] Minimal in the same sense as in Definition 15 with the additional requirement that the comparison is among stratified rankings.
[41] The order of the variables can be any topological sort of the nodes in $\Gamma(\Delta)$.
[42] Since there are no rules that have either tk or bd as consequents, the minimal ranking for these two propositions is the trivial ranking where $\kappa(\omega) = 0$ everywhere.

Table 9
Admissible ranking for $\{tk \rightarrow cs, tk \wedge bd \rightarrow \neg cs, tk \rightarrow x, x \rightarrow bd\}$

κ	Worlds in an admissible, non-stratified ranking
0	$(\neg tk, x, bd, \neg cs)$
1	$(tk, x, \neg bd, cs)$
2	$(tk, x, bd, \neg cs)$
3	Rest of the ω's

Finally, parallel to the notion of consistency in Definition 2, it is important to define a notion of a *c-consistent* Δ by requiring the existence of at least one stratified ranking relative to $\Gamma(\Delta)$. Knowledge bases that are consistent in the probabilistic sense of Definition 2 but yet cannot be given a causal interpretation do exist. For example, $\Delta = \{tk \rightarrow cs, tk \wedge bd \rightarrow \neg cs, tk \rightarrow x, x \rightarrow bd\}$ [43] is c-inconsistent because, if we accept that tk causes cs, we should expect $\neg bd$ to hold by default when tk is true. On the other hand if there is a causal linkage from tk to bd, we should expect bd to hold when tk is true, contrary to our previous expectation. Note that this contradictory knowledge base is admissible in the sense of Definition 2, as shown by the ranking in Table 44. [44] This ranking allows turning the key to protect the battery against the damage inflicted by x, but such a flow of events would be contrary to the causal reading of Δ, but shows no direct linkage from tk to db.

7.2. Actions and observations

Although c-entailment reflects inferences that are typical of causal knowledge, it is still a relation between an observation ϕ and a conclusion σ. As such, it does not exploit the full power of causal models, which reins in the management of actions and other external changes. In Fig. 8, for example, c-entailment would not distinguish between the context ϕ = "the battery was found to be dead" and ϕ' = "someone deliberately drained the battery dead", although the two sentences should trigger different inferences altogether. The first describes a change in our knowledge about an unchanging world, and would normally suggest an explanation within the system, for example, that the headlights were left on. The second triggers no such inferences, as it describes a change in the world due to external intervention, one that occurred regardless of the normal processes listed in the system (e.g., the tendency of batteries to remain charged).

The network $\Gamma(\Delta)$ provides a sufficient code for specifying how external actions would influence the agent's belief ranking $\kappa(\omega)$, including actions that were not anticipated during the construction of the network and, hence, do not possess explicit symbolic representation in Δ. The causal content of the rules in Δ provides a license to treat actions as modalities over the atomic propositions in L by ascribing meaning to new propositions of the type $do(x_i)$ or $do(\neg x_i)$ [35]. The key idea is that each

[43] This is the car example augmented with $k \rightarrow x$ and $x \rightarrow bd$, introducing an intermediate variable x through which turning the key causes the battery to die.

[44] This ranking is not stratified for Δ since $\kappa(bd \wedge x \wedge tk) = 2$, but $\kappa(bd|x) + \kappa(x|tk) + \kappa(tk) = 1$ which contradicts Eq. (39).

child–parents family in the network, by virtue of representing a collection of causal rules, stands for an autonomous physical mechanism in the domain, one that remains invariant to changes, unless specifically altered. An action that imposes a truth value on an atomic proposition (e.g., $do(db)$ or $do(\neg db)$) is viewed as an external intervention which overrules just one mechanism (e.g., the tendency of batteries to remain charged) while leaving all other mechanisms unaltered. The specification of composite actions then requires only the identification of the mechanisms that are overruled by those actions. Once these mechanisms are identified, the effect of the action (or combinations thereof) can be computed from the constraints imposed by the remaining mechanisms.

Formally, intervention is interpreted as conditioning in a larger probability space which includes hypothetical action variables. The effect of an atomic action $do(x_i)$ or $do(\neg x_i)$ is represented by adding to Γ a link $DO_i \longrightarrow x_i$, where DO_i is a new variable taking values in $\{do(x_i), do(\neg x_i), idle\}$, which represents the external intervention. Thus, the new parent set of x_i in the augmented network is $\pi'_i = \pi_i \cup \{DO_i\}$ and it is related to x_i by

$$\kappa(x_i|\pi'_i) = \begin{cases} \kappa(x_i|\pi_i), & \text{if } DO_i = idle, \\ \infty, & \text{if } DO_i = do(\neg x_i), \\ 0, & \text{if } DO_i = do(x_i). \end{cases} \tag{43}$$

The effect of performing action $do(x_i)$ is to transform $\kappa(\omega)$ into a new belief ranking, $\kappa_{x_i}(\omega)$, given by

$$\kappa_{x_i}(\omega) = \kappa'(\omega|do(x_i)) \tag{44}$$

where κ' is the ranking dictated by the augmented network $\Gamma \cup \{DO_i \longrightarrow x_i\}$ and Eqs. (39) and (43).

This yields a simple and direct transformation of pre-action and post-action rankings:

$$\kappa_{x_i}(\omega) = \begin{cases} \kappa(\omega) - \kappa(x_i|\pi_i(\omega)), & \omega \models x_i, \\ \infty, & \omega \models \neg x_i, \end{cases} \tag{45}$$

where $\pi_i(\omega)$ stands for the values that ω assigns to the parents of x_i. This formula reflects the removal of the term $\kappa(x_i|\pi_i)$ from the sum of Eq. (39), since π_i no longer influence x_i. Graphically, the removal of this term is equivalent to removing the links between π_i and x_i while keeping the rest of the network intact. We see that, unlike Bayesian conditioning $\kappa(\omega|x_i)$ (see Eq. (37)) the effect of action $do(x_i)$ is to shift the ranking of each world consistent with (x_i) by a different amount, $\kappa(x_i|\pi_i(\omega))$, depending on the contribution of x_i to the pre-action ranking $\kappa(\omega)$. Worlds in which the occurrence of x_i is a serious possibility given the state $\pi(\omega)$ of its parents, retain their rankings, while those in which the occurrence of x_i is surprising, experience a reduction in κ. This reduction exonerate those latter worlds from the blemish of predicting the exceptional event x_i, since x_i is explained away by an external intervention.

The transformation in Eq. (45) is the ranking-equivalent of a class of probability operators which Lewis named *imaging* [49]. Whereas Bayes conditioning $P(w|e)$ uniformly transfers probability mass from each world excluded by e to the remaining worlds (in proportion to their current $P(w)$), imaging works differently; each world w excluded by e transfers its mass individually to a select set worlds deemed "closest" to w. Causal knowledge imposes specific preferences as to which worlds should be considered closest to any given world. A world w_1 is "closer" to w than w_2 is, if the set of atomic actions needed for transforming w into w_1 is a proper subset of those needed for transforming w into w_2. Considering, for example, the causal chain $x_1 \to x_2 \to \cdots \to x_n$, among all worlds satisfying $x_i = true$, the one closest to $w = \neg x_1 \wedge \neg x_2 \wedge \cdots \wedge \neg x_n$ is $w_1 = \neg x_1 \wedge \cdots \wedge \neg x_{i-1} \wedge x_i \wedge x_{i+1} \wedge \cdots \wedge x_n$.

Imaging is the probabilistic basis for Winslett-Katsuno–Mendelzon [41,73] possible semantics of "belief update", as distinct of "belief revision", and is tacitly invoked whenever actions are specified in terms of *transition probabilities*, as in stochastic control and Markov decision processes (see [61] for more detailed discussion of these aspects). Also, in [35] it is shown that the shift in Eq. (45) conforms to the semantics of belief update introduced by Katsuno and Mendelzon [41].

The transformation defined in (45) exhibits the following features [60]:

(1) An action $do(x_i)$ can affect only the descendants of x_i in Γ.

(2) The variables associated with the root nodes of Γ (often called "exogenous") possess the unique property

$$\kappa_{x_i}(\omega) = \kappa(\omega | x_i).$$

In other words, acting $do(x_i)$ and has the same effect on the agent's belief as passively observing x_i.

(3) The effect of a compound action $A = \bigwedge_{j \in J} A_j$, where each A_j stands for either $do(x_j)$ or $do(\neg x_j)$, can be defined as a sequential application of the transformations associated with the atomic components. The order is irrelevant, since the transformation in Eq. (45) is commutative and associative.

(4) For every variable x_i there exists a unique minimal set of other variables π_i (corresponding to the direct causes of x_i) having the following property. For any two actions, $do(A)$ and $do(B)$, such that neither A nor B logically entails x_i or $\neg x_i$, we have

$$\kappa_A(\pi_i) = \kappa_B(\pi_i) \Rightarrow \kappa_A(x_i) = \kappa_B(x_i).$$

In other words, establishing the impact of an action on the direct causes of x_i is sufficient for determining its impact on x_i as well.

This last property reflects the invariance (or "autonomy") of the linkage between x_i and π_i relative to external interventions, and can in fact be taken as the operational definition of "direct causes". The modeler imagines a hypothetical, ideal laboratory where every compound action can be realized, and envisions the effects of such actions on various variables in the system. The direct causes π_i of x_i exhibit (minimally) a unique behavior in this laboratory: once we fix their values no (indirect) action can ever effect our belief in x_i. Note that this definition embodies a basic asymmetry between causes and effects; fixing the consequences of x_i does not provide x_i any protection against further interventions while fixing the causes of x_i does.

Eq. (45) was derived under the assumption that $\kappa(\omega)$ is given by the sum of Eq. (39), which reflects generic state of knowledge prior to making any specific observations. To define the effect of action $do(A)$ on ranking functions that result from updating $\kappa(\omega)$ by some observations, one must invoke a persistence model to determine which beliefs will persist and which will be "clipped" by the influence of action $do(A)$. If we assume that only those properties should persist which are not under any causal influence to terminate, the following result obtains:

Theorem 38 (see [58]). *Let A be a conjunction of atomic actions, $A = \bigwedge_{j\in J} A_j$, where each A_j stands for either $do(x_j)$ or $do(\neg x_j)$, and let $\kappa(\omega|C)$ be the pre-action ranking after observing C. Then the post-action ranking $\kappa_{A|C}(\omega)$ is given by the formula*

$$\kappa_{A|C}(\omega) = \kappa(\omega) - \sum_{i\in J\cup R} \kappa(x_i(\omega)|\pi_i(\omega)) + \min_{\omega'} \left[\sum_{i\notin J} S_i(\omega,\omega') + \kappa(\omega'|C) \right] \tag{46}$$

where R is the set of root nodes of Γ, ω and ω' are the post-action and pre-action states, respectively, and $S_i(\omega,\omega')$ plays the role of a state transition probability

$$S_i(\omega,\omega') = \begin{cases} s_i, & \text{if } x_i(\omega) \neq x_i(\omega') \text{ and } x_i \in R, \\ s_i, & \text{if } x_i(\omega) \neq x_i(\omega'), x_i \notin R \text{ and } \kappa(\neg x_i(\omega)|\mathbf{pa}_i(\omega)) = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$

$S(\omega,\omega')$ represents persistence assumptions: It is surprising (to degree $s_i \geqslant 1$) to find x_i change from its pre-action value of $x_i(\omega')$ to a post-action value of $x_i(\omega)$ if there is no causal reason for the change.

Eq. (46) implies that belief changes due to long streams of observations and actions can be computed as successive updating operations on epistemic states, these states being organized by a fixed causal network, in which the only varying element is the belief ranking κ.

8. Conclusion

We have presented a qualitative, order-of-magnitude approximation of probability theory, where knowledge is encoded by linguistic expressions which distinguish the typical from the surprising, and the answers identify the set of beliefs induced by any given evidence. Like system-Z, its predecessor, system-Z^+ maintains a clear separation between generic knowledge (in the form of rules) and contingent knowledge (in the form of boolean sentences summarizing observations) and manages the revision of deductively closed beliefs using a qualitative version of probabilistic conditioning.

System-Z^+ augments system-Z with the capability of handling variable-strength rules as well as expressions of imprecise observations. These capabilities are useful in applications such as diagnosis and class-property hierarchies, where rule firmness can be obtained from either statistical information or a general understanding of the domain.

Imprecise observations occur where objects are vaguely defined or when definitions are only partially satisfied.

The system presented is semi-tractable in the sense that it is tractable for every sublanguage in which propositional satisfiability is polynomial (Horn expressions, network theories, acyclic expressions, etc.). To the best of our knowledge, this is the first system that reasons with approximate probabilities and offers such broad guarantees of tractability. Whereas most tractability results exploit the topological structure of the knowledge base [18,45,54] (e.g., trees or hypertrees), ours are topology independent. These results should carry over to possibility theory as formulated by Dubois and Prade [22], which has similar features to Spohn's system except that beliefs are measured on the real interval $[0,1]$. In addition, as Section 5 shows, the system can also accommodate expressions of imprecise observations without loss of tractability, thus providing a good model for weighing the impact of evidence and counterevidence on our beliefs. Finally, the enterprise of belief revision, as formulated in [2,24], can find a tractable and natural embodiment in system-Z^+, unhindered by difficulties that plagued earlier systems.

The incorporation of actions into the language of ranking functions, in Section 7.2, introduces a natural realization of belief *update* which, unlike belief revision, invokes a theory modification, to mimic the change produced by the action. In this realization, revisions and updates are unified through the conditioning operator; revision results from conditioning on observations and update from conditioning on actions [35].

However, unlike the general and abstract formulation of belief revision and update in the literature [3,41], our formulation is tailored specifically to support causal reasoning. The topology of the causal networks plays a crucial role in this update process because it is only by examining the network that we can identify the mechanism (set of parents) which is overruled by any given update and this information, in turn, is necessary for the correct prediction of indirect ramifications of the update. It is our contention that a device similar to a causal network is a necessary component in any formalism of actions and change. Extensions of this representation of actions and their applications involving temporal reasoning, conditional actions and counterfactual queries are explored in [5,6,15,59].

The introduction of a DAG for representing causal relations (Section 7) forms a qualitative counterpart to Bayesian networks [54], thus providing a framework for transporting capabilities between probabilistic reasoning and nonmonotonic logics. For example, Darwiche and Goldszmidt [16] and Henrion et al. [38] compare the performance of stratified rankings with respect to probabilistic Bayesian networks in a diagnosis application. A principled formalization of causal knowledge in ATMS based, again, on a qualitative counterpart of Bayesian networks, is reported by Darwiche [14]. More recently, the connection between stratified rankings and Bayesian networks has given rise to faster algorithms for reasoning under uncertainty in both the probabilistic and nonmonotonic approaches [30].

Using stratified rankings and their probabilistic origins, a qualitative bridge can be established to another stronghold of probability theory—decision analysis. By combining order-of-magnitude specifications of probabilities and utilities, we can compare and rank the expected utilities of actions and consequences, conditioned on observations.

This forms the basis for a *qualitative decision theory* [58] which, in turn, provides a framework for qualitative planning under uncertainty.

Acknowledgements

We thank Ernest Adams, Craig Boutilier, Nir Friedman, Hector Geffner, and Kurt Konolige for insightful discussions on related topics and an anonymous reviewer for helpful comments on a previous version of this paper. Funding was provided in part by NSF grant #IRI-9200918, AFOSR grant #94-1-0173, and MICRO grant #93-124. In addition Moisés Goldszmidt acknowledges the support of an IBM Graduate Fellowship 1990–1992, and the Rockwell Science Center, where part of this was performed.

Appendix A. Proofs of main results

Theorem 4 was proven in [1] where rules are interpreted directly in terms of high probability conditional statements. A proof in terms of ranking functions runs along the lines of the proof of Theorem 19 below. Theorem 5 and Corollary 6 appear in [57] and proofs may be found in [33]. A proof for Theorem 9 can be found in [28, 31]. Theorem 16 is shown in [57] and its proof is similar to that of Theorem 21 (below), its parallel for the case of non-flat sets of rules.

Theorem 19. *Δ^{+} is δ-consistent iff Δ is consistent (in the sense of Definition* 2*).*

Proof. If Δ is consistent (in the sense of Definition 2), we know that there exists a tolerated rule in each non-empty subset Δ' of Δ, and furthermore, we can construct an ordered partition $(\Delta_0, \Delta_1, \ldots, \Delta_n)$ of Δ where: rules in Δ_0 are tolerated by Δ, rules in Δ_1 are tolerated by $\Delta - \Delta_0$ and so on. By Definition 3, for each one of these Δ_j, there must exist a corresponding non-empty subset Ω_j of Ω (the set of all possible worlds), such that for each rule $r_j \in \Delta_j$ there exists a $\omega_j \in \Omega_j$, where ω_j verifies r_j and ω_j satisfies all the rules in the set that results from the union $\bigcup_{i=j}^{n} \Delta_i$ of members of the partition of Δ. Using these possible worlds, we define a partition $(\Omega_0, \Omega_1, \ldots, \Omega_n, \Omega_{n+1})$ of Ω, where each Ω_j contains possible worlds with the characteristics mentioned above, and Ω_{n+1} contains the possible worlds necessary to complete the partition of Ω. Let δ_i^* denote the highest δ among rules in set Δ_i. We now build, in a recursive fashion, an admissible ranking κ relative to Δ^+ based on these two partitions in the following manner: if $\omega_0 \in \Omega_0$, set $\kappa(\omega_0) = 0$; else if $\omega_j \in \Omega_j$, set $\kappa(\omega_j) = \kappa(\omega_{j-1}) + \delta_{j-1}^* + 1$ (where ω_{j-1} is an arbitrary element of the set Ω_{j-1}). Note that the κ-minimal possible world 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 relative to Δ^+, it is enough to show that for an arbitrary pair of possible worlds $\omega_j \in \Omega_j$ and $\omega_{j+1} \in \Omega_{j+1}$ the following relation holds:

$$\kappa(\omega_j) + \delta_j < \kappa(\omega_{j+1}) \tag{A.1}$$

where δ_j can be any δ among the rules in Δ_j. But this relation is guaranteed by the construction of κ since $\kappa(\omega_j) + \delta_j^* + 1 = \kappa(\omega_{j+1})$, where δ_j^* is the highest δ among the rules in Δ_j. Therefore κ is admissible relative to Δ^+, and it follows that Δ^+ is δ-consistent.

For the converse, note that if Δ^+ is δ-consistent, then by Definition 18, there exists a ranking κ such that for every rule $\varphi_i \xrightarrow{\delta_i} \psi_i$ in Δ^+

$$\kappa(\varphi_i \wedge \psi_i) < \delta_i + \kappa(\varphi_i \wedge \neg\psi_i). \tag{A.2}$$

Since δ_i is an integer bigger or equal to zero it follows that

$$\kappa(\varphi_i \wedge \psi_i) < \kappa(\varphi_i \wedge \neg\psi_i), \tag{A.3}$$

and thus by Definition 2, Δ is consistent. □

Theorem 21. *Every consistent Δ^+ has a unique minimal ranking given by κ^+.*

Proof. We need some intermediate results. First we show that κ^+ is admissible. We then define *compactness* (Definition A.2), and show that κ^+ is compact and unique (Lemmas A.3 and A.4). Theorem 21 will follow from these results.

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

Proof. Given that $Z^+(r_i) = \min_j\{\kappa^+(\omega_j) \mid \omega_j \models \varphi_i \wedge \psi_i\} + \delta_i$, we can rewrite the conditions for admissibility (Eq. (14)) as

$$Z^+(r_i) < \min_j\{\kappa^+(\omega_j) \mid \omega_j \models \varphi_i \wedge \neg\psi_i\}. \tag{A.4}$$

Since $\kappa^+(\omega) = \max\{Z^+(r_i) \mid \omega \models \varphi_i \wedge \neg\psi_i\} + 1$, it follows that the right-hand side of Eq. (A.4) is at least $Z^+(r_i) + 1$ and κ^+ is admissible. □

Having proved that κ^+ is admissible, we now prove that κ^+ is compact given a set of rules Δ. A ranking κ will be said to be compact with respect to Δ if lowering the rank of any world w in κ without modifying the rank of the rest of the worlds would make κ inadmissible with respect to Δ. Formally,

Definition A.2 (*Compact rankings*). Given Δ, an admissible ranking κ is said to be *compact* with respect to Δ if for every ω' any ranking κ' satisfying

$$\begin{aligned} \kappa'(\omega) &= \kappa(\omega), \quad \omega \neq \omega', \\ \kappa'(\omega) &< \kappa(\omega), \quad \omega = \omega' \end{aligned}$$

is *not* admissible with respect to Δ.

Lemma A.3. *The ranking κ^+ is compact with respect to a given Δ.*

Proof. By contradiction. Assume it is possible to lower $\kappa^+(\omega')$ of some possible world ω', where $\kappa^+(\omega') > 0$. From the definition of κ^+ (Definition 20), there must be a rule $r : \varphi \xrightarrow{\delta} \psi \in \Delta$ such that $\kappa^+(\omega') = Z^+(r) + 1$ and moreover $\omega' \models \varphi \wedge \neg\psi$ (see Eq. (15)). This implies that

$$\kappa^+(\omega') = \min\{\kappa^+(\omega) \mid \omega \models \varphi \wedge \psi\} + \delta + 1. \tag{A.5}$$

Lowering the value of $\kappa^+(\omega')$ will imply that

$$\kappa^+(\omega') \leqslant \min\{\kappa^+(\omega) \mid \omega \models \varphi \wedge \psi\} + \delta, \tag{A.6}$$

which, since w' falsifies rule r, clearly violates the condition of admissibility in Eq. (14) with respect to rule r. □

Lemma A.4. *Every consistent Δ^+ has a unique compact ranking given by κ^+.*

Proof. By Lemma A.3, 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 possible world. We will show that if there exists an ω' such that $\kappa(\omega') < \kappa^+(\omega')$ then κ cannot be admissible, where if $\kappa(\omega') > \kappa^+(\omega')$, then κ cannot be compact.

Assume first that $\kappa(\omega') < \kappa^+(\omega')$, and let I be the lowest κ-value for which such inequality holds. In other words, for every world ω such that $\kappa(\omega) < I$, $\kappa^+(\omega) = \kappa(\omega)$. Let $\kappa^+(\omega') = J$, where $J > I$. Note that J must be strictly greater than zero, and as a consequence, by the definition of κ^+ (Definition 20), we know that there is a rule $r : \varphi \xrightarrow{\delta} \psi$ such that $\omega' \models \varphi \wedge \neg\psi$. This implies that Eq. (A.5) holds, and as a consequence

$$\min\{\kappa^+(\omega) \mid \omega \models \varphi \wedge \psi\} = J - \delta - 1. \tag{A.7}$$

Since κ is assumed to be admissible, we have that the following must also hold for rule r in κ, this implies that[45]

$$\kappa(\omega') \geqslant \min\{\kappa(\omega) \mid \omega \models \varphi \wedge \psi\} + \delta + 1. \tag{A.8}$$

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

$$J > \min\{\kappa(\omega) \mid \omega \models \varphi \wedge \psi\} + \delta + 1. \tag{A.9}$$

If we subtract $\delta + 1$ from both sides of this inequality and use Eq. (A.7) we get

$$\min\{\kappa^+(\omega) \mid \omega \models \varphi \wedge \psi\} > \min\{\kappa(\omega) \mid \omega \models \varphi \wedge \psi\}. \tag{A.10}$$

But this cannot be since I was assumed to be the minimal value of κ for which this inequality can occur, and if $\min\{\kappa(\omega) \mid \omega \models \varphi \wedge \psi\} > I$, then κ would violate Eq. (A.8) which in turn would imply that κ was not admissible in the first place.

Now assume that there is a non-empty set of possible worlds for which $\kappa(\omega) > \kappa^+(\omega)$, and let I be the lowest κ^+ value in which $\kappa(\omega') > \kappa^+(\omega')$ for some possible world ω'. We will show that κ cannot be compact, since it will be possible to reduce

[45] Note that Eq. (A.8) implies that I must be bigger than zero, and an inductive proof may be constructed on this basis. We find the proof presented here shorter and clearer.

$\kappa(\omega')$ to $\kappa^+(\omega')$ while keeping constant the κ of all other possible worlds. From $\kappa^+(\omega') = I$ we know that ω' 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 possible world ω, such that ω verifies r and $\kappa^+(\omega) < I$. Since for all these possible worlds κ is assumed to be equal to κ^+ it follows that none of the rules verified by these worlds, can be violated by reducing $\kappa(\omega')$ to $\kappa^+(\omega')$. □

The proof of Theorem 21 follows immediately from Lemma A.4 and Definition 15.

Theorem 22. *The function Z computed by procedure* Z^+_order *satisfies Definition* 20 (*i.e., the output of procedure* Z^+_order *is* Z^+).

Proof. We first show that the relevant steps in procedure Z^+_order 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^+. For similar reasons, Δ^* 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 non-empty subset of Δ^+. Finally, in the computation of Eq. (17), since ω only falsifies rules in $\mathcal{R}Z^+$, the Z-ranking for each of these rules is 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 possible world ω_0 (for each one of these rules), such that ω_0 verifies r_0 and ω_0 satisfies Δ^+. Thus, each one of these possible worlds does not falsify any rules in Δ^+, and $\kappa^+(\omega_0) = 0$. According to Eq. (16) in Definition 20, $Z^+(r_0) = \delta_0$ for those rules and that is precisely what is computed in step 2.

The proof proceeds by induction on the iterations of loop 3; we show that for every rule $r \in \mathcal{R}Z^+$, $Z(r) = Z^+(r)$ holds as an invariant. 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 r^* inserted into $\mathcal{R}Z^+$ at step 3(c). Note that since all the values $\kappa^+(\omega_r)$ for ω_r in every Ω_r are computed from Z^+-values of rules in $\mathcal{R}Z^+$ (step 3(b), Eqs. (17) and (18)), they must be equal to $\kappa^+(\omega)$. Let a characteristic possible world for a rule r be the possible world ω_r^* with minimal ranking κ^+ verifying r. Thus, $Z^+(r) = \min_{\omega \models \varphi \wedge \psi} \kappa^+(\omega) + \delta = \kappa^+(\omega_r^*) + \delta$. We claim that $\kappa^+(\omega_{r^*}^*)$ [46] is a characteristic possible world for the rules outside $\mathcal{R}Z^+$ it verifies. Suppose not: assume that there is a possible world ω_{r^*} such that ω_{r^*} verifies a rule r^* (that is inserted into $\mathcal{R}Z^+$ in step 3(c)), and $\kappa^+(\omega_{r*}) < \kappa^+(\omega_{r^*}^*)$. Note that ω_{r*} must falsify a rule $r' \notin \mathcal{R}Z^+$. Otherwise the computation in Eq. (17) would not have used $\omega_{r^*}^*$ but ω_r^* instead. Let $\omega_{r'}$ be a characteristic possible world for r', then

$$\kappa^+(\omega_{r'}) < \kappa^+(\omega_{r^*}). \tag{A.11}$$

Note that $\omega_{r^*}^*$ cannot verify r', since otherwise

$$\kappa^+(\omega_{r^*}^*) < \kappa^+(\omega_{r^*}), \tag{A.12}$$

[46] Recall that r^* is a rule inserted into $\mathcal{R}Z^+$ in step 3(c).

a contradiction. If $\omega_{r'}$ does not verify the same rule r^* that $\omega^*_{r^*}$ verifies, then $Z(r') \geqslant Z(r^*)$ by step 3(c), and then by Eq. (18), $\kappa(\omega_{r^*}) > \kappa(\omega^*_{r^*})$ which is a contradiction. Therefore $\omega_{r'}$ verifies the same r^*, and by the minimality of $\omega^*_{r^*}$ among the worlds in Ω_{r^*}, $\omega_{r'}$ must falsify a rule $r'' \notin \mathcal{RZ}^+$. If $\omega_{r''}$ is a characteristic possible world for r'' we have that

$$\kappa^+(\omega_{r''}) < \kappa^+(\omega_{r'}) < \kappa^+(\omega_{r^*}). \tag{A.13}$$

ω_{r^*} cannot verify r''; otherwise we get the contradiction

$$\kappa^+(\omega_{r^*}) < \kappa^+(\omega_{r'}) < \kappa^+(\omega_{r^*}) \tag{A.14}$$

and if $\omega^*_{r^*}$ verifies r'' we get the contradiction of Eq. (A.12). By similar arguments as before $\omega_{r''}$ must falsify another rule outside $\mathcal{RZ}^+$. However, given that Δ^+ is finite, we cannot extend the "chain" of Eq. (A.13) indefinitely, and therefore we are bound to get a contradiction in the form of Eq. (A.12) or Eq. (A.14). Since our only assumption was that $w^*_{r^*}$ is not a characteristic possible world for the rules it verifies, that assumption must be wrong. It follows then that the value of $Z(r^*)$ computed in step 3(b) (Eq. (17)) must be equal to Z^+. For the induction step assume that the invariant holds up till the nth iteration. Then by the same argument used in the basis of the induction, the $\kappa(\omega_r)$ for $\omega_r \in \Omega_r$ are equal to $\kappa^+(\omega_r)$, the minimal $\kappa^+(\omega^*_{r^*})$ in Eq. (17) must be a characteristic possible world for the rules r^* outside of $\mathcal{RZ}^+$ that it verifies, and thus $Z(r^*) = \kappa^+(\omega^*_{r^*}) + \delta_{r^*} = Z^+(r^*)$. □

The proof of Theorem 23 is made easier by the following pair of intermediate results (Lemmas A.5 and A.6).

Lemma A.5. *Let $\Delta^+ = \{r_i \mid r_i = \varphi_i \xrightarrow{\delta_i} \psi_i\}$, $1 \leqslant i \leqslant n$, be a consistent knowledge base in which rules are sorted according a priority function $Z(r_i)$. Let $\kappa(\omega)$ be defined as in Eq.* (15):

$$\kappa(\omega) = \begin{cases} 0, & \text{if } \omega \text{ does not falsify any rule in } \Delta^+, \\ \max\limits_{\omega \models \varphi_i \wedge \neg\psi_i} [Z(r_i)] + 1, & \text{otherwise.} \end{cases} \tag{A.15}$$

Then, for any wff ϕ, $\kappa(\phi)$ can be computed in $O(\log n)$ *propositional satisfiability tests.*

Proof. The idea is to perform a binary search on Δ^+ to find the lowest $Z(r)$ such that there is a model for ϕ that does not violate any rule r' with priority $Z(r') \geqslant Z(r)$. We first divide Δ^+ into two roughly equal sections: top-half (r_{mid} to r_{high}) and bottom-half (r_{low} to r_{mid}). Then we examine the top-half: If the wff $\alpha = \phi \bigwedge_{j=\text{mid}}^{n} \varphi_j \supset \psi_j$ is satisfiable, then there exists a model for ϕ that does not violate any rule in this top-half. It follows that $Z(r_{\text{mid}})+1$ is an upper bound on the value of $\kappa(\phi)$, and the binary search is continued iteratively in the bottom-half. If, on the other hand, α is not satisfiable, then the maximum $Z(r_i)$ for any model for ϕ must be in the top-half, and the search is continued there. Eventually, the set in which the search is conducted is reduced to one rule, and we can determine the value of $\kappa(\phi)$ with one more satisfiability test. □

Lemma A.6. *The value of $Z(\phi \xrightarrow{\delta} \sigma)$ in Eq. (17) can be computed in* $O(\log|\mathcal{R}Z^+|)$ *satisfiability tests.*

Proof. Let Δ' in step 3(a) be equal to $\{\varphi_i \xrightarrow{\delta_i} \psi_i\}$, and let the wff α be equal to $\sigma \wedge \phi \bigwedge_i \varphi_i \supset \psi_i$ where i ranges over all the rules in Δ'. Note that since any world ω_r in Ω_r is a model for $\sigma \wedge \phi$ and does not violate any rule in Δ', it follows that $\omega_r \in \Omega_r$ iff $\omega_r \models \alpha$. Then, since $\kappa(\alpha) = \min_{\omega_r \in \Omega_r} \kappa(\omega_r)$, $Z(\phi \xrightarrow{\delta} \sigma)$ must be equal to $\kappa(\alpha) + 1 + \delta$. Thus, once Δ' is sorted, by Lemma A.5, $\kappa(\alpha)$ can be computed in $O(\log|\mathcal{R}Z^+|)$ satisfiability tests which proves Lemma A.6. □

Theorem 23. *Given a consistent $\Delta^+ = \{r_i \mid r_i = \varphi_i \xrightarrow{\delta_i} \psi_i\}$, $1 \leqslant i \leqslant n$, the computation of the ranking Z^+ requires $O(n^2 \times \log n)$ satisfiability tests.*

Proof. Step 1 requires at most n satisfiability tests and is performed once, while step 2 takes at most n data assignments. Step 3(a) again requires $O(n)$ satisfiability tests. Computing Eq. (17) in Step 3(b) can be done in $O(\log|\mathcal{R}Z^+|)$ satisfiability tests according to Lemma A.6,[47] and since it will be executed at most $O(n)$ times, it requires a total of $O(n \times \log n)$ satisfiability tests. Step 3(c) is a minimum search which can be done in conjunction with the computation of Eq. (17), since we only need to keep the minimum of such values (this involves $|\Delta^*|$ data comparisons). Loop 3 is performed at most $n - |\Delta_0|$ times, hence the whole computation of the priorities Z^+ on rules requires a total of $O(n^2 \times \log n)$ satisfiability tests. □

Theorem 25. *Let $r_1 : \varphi \xrightarrow{\delta_1} \psi$ and $r_2 : \phi \xrightarrow{\delta_2} \sigma$ be two rules in a consistent Δ such that*

(1) $\varphi \mathrel{|\!\!\sim_p} \phi$ (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 $Z^+(r_1) > Z^+(r_2)$ independently of the values of δ_1 and δ_2.

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

$$\kappa(\varphi \wedge \phi) < \kappa(\varphi \wedge \neg\phi). \tag{A.16}$$

Thus, any characteristic possible world $\omega^+_{r_1}$ for r_1 must render ϕ (the antecedent for r_2) true, and since there is no possible world such that both rules are verified (condition (2) in the theorem above), all $\omega^+_{r_1}$ must also falsify r_2. From Definition 8 (Eqs. (15) and (16)): $\kappa^+(\omega^+_{r_1}) \geqslant Z^+(r_2) + 1$, and $Z^+(r_1) = \kappa^+(\omega^+_{r_1}) + \delta_2$. It follows that $Z^+(r_1) > Z^+(r_2)$. Note that the characteristic possible world 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. □

---

[47] Note that we need $\mathcal{R}Z^+$ to be sorted, non-decreasingly, with respect to the priorities Z. This requires that the initial values inserted to $\mathcal{R}Z^+$ in step 2 of procedure Z^+_order be sorted—$O(|\Delta_0|^2)$ data comparisons—and that the new Z-value in step 3(c) be inserted in the right place—$O(|\mathcal{R}Z^+|)$ data comparisons. We are assuming that the cost of each of these operations is much less than that of a satisfiability test.

Theorem 34. *Let $x_r \wedge \cdots \wedge x_s$ be the parents of x in $\Gamma(\Delta)$. Let $\mathcal{Y} = \{y_1, \ldots, y_m\}$ be a set of atomic propositions such that no $y_i \in \mathcal{Y}$ is a descendant of x in $\Gamma(\Delta)$, and let ϕ_Y be any wff built only with elements from $\mathcal{Y}$.*[48] *If $x_r \wedge \cdots \wedge x_s \mid\!\mid\!\!\sim_c x$ then $\phi_Y \wedge x_r \wedge \cdots \wedge x_s \mid\!\mid\!\!\sim_c x$.*

Proof. We show the stronger result that

$$\kappa(x|x_r \wedge \cdots \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n) = \kappa(x|x_r \wedge \cdots \wedge x_s). \tag{A.17}$$

The statement in terms of consequence relations in the theorem is implied by Eq. (A.17), since the $\kappa(x|x_r \wedge \cdots \wedge x_s \wedge \phi_Y)$ can be computed from the $\kappa(x|x_r \wedge \cdots \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n)$, and the semantics of the consequence relation is expressed in terms of the conditional κ rankings.

By the conditions of stratification in Definition 32 (Eq. (39)), we have that

$$\begin{aligned}&\kappa(x \wedge x_r \wedge \cdots \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n \wedge \mathcal{P})\\ &\quad = \kappa(x|x_r \wedge \cdots \wedge x_s) + \kappa(x_r|\pi(x_r)) + \cdots\\ &\qquad + \kappa(x_s|\pi(x_s)) + \kappa(y_1|\pi(y_1)) + \cdots\end{aligned} \tag{A.18}$$

where $\mathcal{P}$ represents a conjunction of the parents of the set $\mathcal{S} = \{x_r, \ldots, x_s, y_1, \ldots, y_n\}$ without including any of the elements in $\mathcal{S}$. Once more by the requirements of stratification we can transform the right-hand side of Eq. (A.18) into

$$\begin{aligned}&\kappa(x \wedge x_r \wedge \cdots \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n \wedge \mathcal{P})\\ &\quad = \kappa(x|x_r \wedge \cdots \wedge x_s) + \kappa(x_r|\pi(x_r)) + \cdots\\ &\qquad + \kappa(x_s \wedge x_r \wedge y_1 \wedge \cdots \wedge y_n \wedge \mathcal{P}).\end{aligned} \tag{A.19}$$

We take the minimum on both sides over $\mathcal{P}$ and we have

$$\begin{aligned}&\kappa(x \wedge x_r \wedge \cdots \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n)\\ &\quad = \kappa(x|x_r \wedge \cdots \wedge x_s) + \kappa(x_r \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n)\end{aligned} \tag{A.20}$$

which is equivalent to

$$\begin{aligned}&\kappa(x \wedge x_r \wedge \cdots \wedge x_s \wedge y_1 \wedge \cdots \wedge y_n) - \kappa(x_s \wedge x_r \wedge y_1 \wedge \cdots \wedge y_n)\\ &\quad = \kappa(x|x_r \wedge \cdots \wedge x_s)\end{aligned} \tag{A.21}$$

and to Eq. (A.17). □

Theorem 35. *Let $\Delta' \subseteq \Delta$ and $\Gamma'(\Delta')$ a subgraph of $\Gamma(\Delta)$ such that if x' is a node in $\Gamma'(\Delta')$ then all the rules in Δ with x' (or $\neg x'$) as their consequent are also in Δ'. Let φ and ψ be two wffs built with propositions corresponding to nodes in $\Gamma'(\Delta')$. If $\varphi \mid\!\mid\!\!\sim_c \psi$ in the context of Δ' then $\varphi \mid\!\mid\!\!\sim_c \psi$ in the context of Δ.*

[48] We also require that $\phi_Y \wedge x_r \wedge \cdots \wedge x_s$ be satisfiable.

Proof. Note that any stratified ranking for Δ must also be a stratified ranking for Δ'. Therefore if $\kappa(\neg\varphi|\psi) > 0$ in every stratified ranking for Δ', $\kappa(\neg\varphi|\psi) > 0$ in every stratified ranking for Δ, and the theorem follows. □

The following two propositions prove results in reference to Examples 36 and 37:

Proposition A.7. *Given* $\Delta = \{tk \to cs, tk \wedge bd \to \neg cs, lo \to bd\}$, $lo \wedge tk \mathrel{|\!\!\!\sim_c} \neg cs$.

Proof. $\Gamma(\Delta)$ is shown in Fig. 8; then:

(1) $lo \mathrel{|\!\!\!\sim_c} bd$; by the *Defaults* rule.
(2) $tk \wedge lo \mathrel{|\!\!\!\sim_c} bd$; by (1) and the *Markov* rule.[49]
(3) $tk \wedge bd \mathrel{|\!\!\!\sim_c} \neg cs$; by the *Defaults* rule.
(4) $tk \wedge bd \wedge lo \mathrel{|\!\!\!\sim_c} \neg cs$; by (3) and the *Markov* rule.[50]
(5) $tk \wedge lo \mathrel{|\!\!\!\sim_c} \neg cs$; by (2), (4) and the *Cut* rule. □

Proposition A.8. *Given the network* $\{l_0 \to l_1, l_1 \to l_2, \ldots, l_{n-1} \to l_n\}$, $l_0 \wedge \neg l_n \mathrel{|\!\!\!\sim_c} \neg(l_1 \wedge \cdots \wedge l_n)$ *and it is not the case that* $l_0 \wedge \neg l_n \mathrel{|\!\!\!\sim_c} l_i$ *for* $1 \leqslant i < n$.

Proof. $l_0 \wedge \neg l_n \mathrel{|\!\!\!\sim_c} \neg(l_1 \wedge \cdots \wedge l_n)$ follows trivially from the *Deduction* rule. The fact that we cannot point out the exact moment in which the gun is unloaded follows from the ranking built using the following recipe: $k(\omega)$ = number of rules in Δ falsified by ω. It is easy to verify that this ranking is stratified, and that all formulas representing situations in which the gun is unloaded at different times have equal ranking. Thus, it is not the case that in all stratified and admissible rankings $\kappa(\neg l_i|l_0 \wedge l_n) > 0$ for any particular l_i, $1 \leqslant i < n$. □

References

[1] E.W. Adams, *The Logic of Conditionals* (Reidel, Dordrecht, Netherlands, 1975).

[2] C. Alchourrón, P. Gärdenfors and D. Makinson, On the logic of theory change: partial meet contraction and revision functions, *J. Symb. Logic* **50** (1985) 510–530.

[3] F. Bacchus, *Representing and Reasoning with Probabilistic Knowledge, A Logical Approach to Probabilities* (MIT Press, Cambridge, MA, 1991).

[4] F. Bacchus, A. Grove, J. Halpern and D. Koller, Statistical foundations of default reasoning indifference and irrelevance, in: *Proceedings Fourth International Workshop on Nonmonotonic Reasoning*, Vermont (1992) 1–12.

[5] A. Balke and J. Pearl, Counterfactual probabilities: computational methods, bounds, and applications, in: *Proceedings Tenth Annual Conference on Uncertainty in Artificial Ingelligence* (1994) 46–54.

[6] A. Balke and J. Pearl, Counterfactuals and policy analysis in structural models, in: *Proceedings Eleventh Annual Conference on Uncertainty in Artificial Ingelligence* (1995) 11–18.

[7] R. Ben-Eliyahu, NP-complete problems in optimal horn clauses satisfiability, Tech. Rept. R-158, Cognitive Systems Laboratory, UCLA, Los Angeles, CA (1990).

[8] C. Boutilier, Conditional logics for default reasoning and belief revision, Ph.D. Dissertation, University of Toronto, Toronto, Ont. (1992); also: Tech. Rept. 91-1, Department of Computer Science, University of British Columbia, Vancouver, BC (1991).

[49] Since tk is a non-descendant of bd, and lo is the parent of bd in $\Gamma(\Delta)$.

[50] Since lo is a non-descendant of cs, and $\{tk, bd\}$ is the parent set of cs in $\Gamma(\Delta)$.

[9] C. Boutilier, What is a default priority?, in: *Proceedings Ninth Biennial Conference of the Canadian Society for Computational Studies of Intelligence*, Vancouver, BC (1992) 140–147.

[10] C. Boutilier, Iterated revision and minimal revision of conditional beliefs, *J. Philos. Logic* (1995); Preliminary version in: *Proceedings IJCAI-93*, Chambery, France (1993) 519-525.

[11] C. Boutilier and M. Goldszmidt, Revision by conditional beliefs, in: *Proceedings AAAI-93*, Washington, DC (1993) 649-654.

[12] G. Brewka, Preferred subtheories: an extended logical framework for default reasoning, in: *Proceedings IJCAI-89*, Detroit, MI (1989).

[13] M. Dalal, Investigations into a theory of knowledge base revision: preliminary report, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 475–479.

[14] A. Darwiche, Conditional independence in ATMSs: independence-based algorithms for computing labels and diagnoses, in: *Proceedings 5th International Workshop on Principles of Diagnosis*, New Paltz, NY (1994) 70–80.

[15] A. Darwiche and M. Goldszmidt, Action networks: a framework for reasoning about actions and change under uncertainty, in: *Proceedings Tenth Annual Conference on Uncertainty in Artificial Intelligence* (1994) 136–144.

[16] A. Darwiche and M. Goldszmidt, On the relation between kappa calculus and probabilistic reasoning, in: *Proceedings Tenth Annual Conference on Uncertainty in Artificial Intelligence* (1994) 145–153.

[17] A. Darwiche and J. Pearl, On the logic of iterated belief revision, in: *Proceedings Fifth Conference on Theoretical Aspects of Reasoning about Knowledge*, Pacific Grove, CA (1994) 5–23.

[18] R. Dechter and J. Pearl, Network-based heuristics for constraint satisfaction problems, *Artif. Intell.* **34** (1987) 1–38.

[19] R. Dechter and J. Pearl, Directed constraint networks: a relational framework for causal modeling, in: *Proceedings IJCAI-91*, Sydney, Australia (1991) 1164–1170.

[20] J.P. Delgrande, An approach to default reasoning based on a first-order conditional logic: revised report, *Artif. Intell.* **36** (1988) 63–90.

[21] W. Dowling and J. Gallier, Linear-time algorithms for testing the satisfiability of propositional Horn formulae, *J. Logic Programming* **3** (1984) 267–284.

[22] D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty* (Plenum Press, New York, 1988).

[23] R. Fagin, J.D. Ullman and M. Vardi, On the semantics of updates in databases, in: *Proceedings Second ACM Symposium on Principles of Database Systems*, Atlanta, GA (1983) 352-365.

[24] P. Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (MIT Press, Cambridge, MA, 1988).

[25] H.A. Geffner, *Default Reasoning: Causal and Conditional Theories* (MIT Press, Cambridge, MA, 1992).

[26] H.A. Geffner and J. Pearl, A framework for reasoning with defaults, in: H. Kyburg, R. Loui and G. Carlson, eds., *Knowledge Representation and Defeasible Reasoning* (Kluwer Academic Publishers, Bosston, MA, 1990) 245–265.

[27] H.A. Geffner and J. Pearl, Conditional entailment: bridging two approaches to default reasoning, *Artif. Intell.* **53** (1992) 209-244.

[28] M. Goldszmidt, Qualitative probabilities: a normative framework for commonsense reasoning, Ph.D. Thesis, Computer Science Department, University of California, Los Angeles, CA (1992); also: Tech. Rept. TR-190, Cognitive Systems Laboratory.

[29] M. Goldszmidt, Building minimal stratified rankings, Technical Memorandum 150, Rockwell Science Center, Palo Alto, CA (1995).

[30] M. Goldszmidt, Fast belief update using order-of-magnitude probabilities, in: *Proceedings Eleventh Annual Conference on Uncertainty in Artificial Intelligence* (1995) 208–216.

[31] M. Goldszmidt, P. Morris and J. Pearl, A maximum entropy approach to nonmonotonic reasoning, *IEEE Trans. Pattern Anal. Mach. Intell.* **15** (1993) 220-232.

[32] M. Goldszmidt and J. Pearl, On the relation between rational closure and system Z, in: *Proceedings Third International Workshop on Nonmonotonic Reasoning*, South Lake Tahoe, CA (1990) 130-140.

[33] M. Goldszmidt and J. Pearl, On the consistency of defeasible databases, *Artif. Intell.* **52** (1991) 121-149.

[34] M. Goldszmidt and J. Pearl, System Z^+: a formalism for reasoning with variable strength defaults, in: *Proceedings AAAI-91*, Anaheim, CA (1991) 399-404.

[35] M. Goldszmidt and J. Pearl, Rank-based systems: a simple approach to belief revision, belief update, and reasoning about evidence and actions, in: *Proceedings Third International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA (1992) 661-672.
[36] M. Goldszmidt and J. Pearl, Stratified rankings for causal relations, in: *Proceedings Fourth International Workshop on Nonmonotonic Reasoning*, Vermont (1992) 99-110.
[37] S. Hanks and D. McDermott, Non-monotonic logics and temporal projection, *Artif. Intell.* **33** (1987) 379-412.
[38] M. Henrion, G. Provan, B. del Favero and G. Sanders, An experimental comparison of diagnostic performance using infinitesimal and numerical Bayesian belief networks, in: *Proceedings Tenth Annual Conference on Uncertainty in Artificial Intelligence* (1994) 319-326.
[39] D. Hunter, Parallel belief revision, in: R.D. Shachter, T.S. Levitt, L.N. Kanal and J.F. Lemmer, eds., *Uncertainty in Artificial Intelligence* **4** (North-Holland, Amsterdam, 1990) 241-252.
[40] R. Jeffrey, *The Logic of Decision* (University of Chicago Press, Chicago, IL, 1983); also: (McGraw-Hill, New York, 1st ed., 1965).
[41] H. Katsuno and A. Mendelzon, On the difference between updating a knowledge base and revising it, in: *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA (1991) 387-394.
[42] H. Kautz, The logic of persistence, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 401-407.
[43] S. Kraus, D. Lehmann and M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, *Artif. Intell.* **44** (1990) 167-207.
[44] H. Kyburg, *Probability and the Logic of Rational Belief* (Wesleyan University Press, Middletown, CT, 1961).
[45] S. Lauritzen and D. Spiegelhalter, Local computations with probabilities on graphical structures and their applications to expert systems, *Roy. Stat. Soc.* **50** (1988) 154-227.
[46] D. Lehmann, What does a conditional knowledge base entail?, in: *Proceedings First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Ont. (1989) 212-222.
[47] D. Lehmann and M. Magidor, What does a conditional knowledge base entail?, *Artif. Intell.* **55** (1992) 1-60.
[48] I. Levi, Iteration of conditionals and the Ramsey test, *Synthese* **76** (1988) 49-81.
[49] D. Lewis, Probabilities of conditionals and conditional probabilities, *Philos. Rev.* **85** (1976) 297-315.
[50] J. McCarthy, Applications of circumscription to formalizing commonsense knowledge, *Artif. Intell.* **28** (1986) 89-116.
[51] B. Nebel, Belief revision and default reasoning: syntax-based approaches, in: *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA (1991) 417-428.
[52] D. Nute, Conditional logic, in: S. Shapiro, ed., *Encyclopedia of Artificial Intelligence* (Wiley, New York, 2nd ed., 1992) 854-860.
[53] J. Pearl, Embracing causality in formal reasoning, *Artif. Intell.* **35** (1988) 259-271.
[54] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, CA, 1988).
[55] J. Pearl, Probabilistic semantics for nonmonotonic reasoning: a survey, in: *Proceedings First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, Ont. (1989) 505-516.
[56] J. Pearl, Jeffrey's rule, passage of experience and neo-Bayesianism, in: H. Kyburg, R. Loui and G. Carlson, eds., *Defeasible Reasoning and Knowledge Representation* (Kluwer Academic Publishers, Boston, MA, 1990) 121-135.
[57] J. Pearl, System Z: a natural ordering of defaults with tractable applications to default reasoning, in: *Proceedings Third Conference on Theoretical Aspects of Reasoning about Knowledge*, Pacific Grove, CA (1990) 121-135.
[58] J. Pearl, From conditional oughts to qualitative decision theory, in: *Proceedings Ninth Annual Conference on Uncertainty in Artificial Intelligence*, Washington, DC (1993) 12-20.
[59] J. Pearl, A probabilistic calculus of action, in: *Proceedings Tenth Annual Conference on Uncertainty in Artificial Intelligence* (1994) 454-462.
[60] J. Pearl, Causal diagrams for experimental research, *Biometrika* **28** (4) (1995) 669-710.

[61] J. Pearl, Causation, action and counterfactuals, in: A. Gammerman, ed., *Computational Learning and Probabilistic Reasoning* (Wiley, New York, 1995); also in: *Proceedings Sixth Conference on Theoretical Aspects of Rationality and Knowledge*, Netherlands (1996) 51–73.

[62] J. Pearl and T. Verma, A theory of inferred causation, in: *Proceedings Second International Conference on Principles of Knowledge Representation and Reasoning*, Cambridge, MA (1991) 441–452.

[63] D. Poole, On the comparison of theories: preferring the most specific explanation, in: *Proceedings IJCAI-85*, Los Angeles, CA (1985) 144–147.

[64] D. Poole, A logical framework for default reasoning, *Artif. Intell.* **36** (1988) 27–47.

[65] R. Reiter, A logic for default reasoning, *Artif. Intell.* **13** (1980) 81–132.

[66] H. Rott, A non-monotonic conditional logic for belief revision i, in: A. Fuhrmann and M. Morreau, eds., *The Logic of Theory Change*, Lecture Notes in Computer Science **465** (Springer, Berlin, 1991) 135–181.

[67] P. Shenoy, On Spohn's rule for revision of beliefs, *Int. J. Approx. Reasoning* **5** (1991) 149–181.

[68] Y. Shoham, Nonmonotonic logics: meaning and utility, in: *Proceedings IJCAI-87*, Milan (1987) 388–393.

[69] Y. Shoham, *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence* (MIT Press, Cambridge, MA, 1988).

[70] W. Spohn, Stochastic independence, causal independence, and shieldability, *J. Philos. Logic* **9** (1980) 73–99.

[71] W. Spohn, Ordinal conditional functions: a dynamic theory of epistemic states, in: W.L. Harper and B. Skyrms, eds., *Causation in Decision, Belief Change, and Statistics* (Reidel, Dordrecht, Netherlands, 1988) 105–134.

[72] P. Suppes, *A Probabilistic Theory of Causation* (North-Holland, Amsterdam, 1970).

[73] M. Winslett, Reasoning about action using a possible worlds approach, in: *Proceedings AAAI-88*, St. Paul, MN (1988) 89–93.