# Extending Description Logics to Support Statistical Information

Terrance Swift[a], Gregory Wheeler[a]

[a]CENTRIA, Departamento de Informática, FCT, Universidade Nova de Lisboa 2829-516 Caparica, Portugal



**Abstract**

A core problem for knowledge representation and reasoning (KRR) systems is to infer from a knowledge base $\Gamma$ which properties are true for a given individual object $o$. Classical and non-monotonic logics are suitable for such inference when $\Gamma$ contains only non-probabilistic information. Alternately, Bayesian reasoning can be used if $\Gamma$ contains sufficient information about a single joint distribution, or perhaps a convex set of distributions. However, if the domain of $\Gamma$ contains incomplete or even contradictory probabilistic information, as is common in epidemiological and actuarial knowledge bases, then neither of these orthodox KRR methods will do. Evidential Probability addresses such cases by weighing the quality of the probabilistic information and the relevance of that information to $o$. However, in [15] and elsewhere, Evidential Probability has been formulated abstractly, without regard to decidability or to complexity. As a result, Evidential Probability has not been used within KRR systems. This paper shows how the inference procedure of Evidential Probability, when grounded in a decidable description logic, can be implemented when $\Gamma$ is a *Minimal Knowledge with Negation as Failure (MKNF)* knowledge base. In such cases, we show correctness of our implementation, demonstrate its low-order polynomial complexity, and provide an open-source prototype implementation.

A core problem for knowledge representation and reasoning (KRR) systems is to infer from a knowledge base which properties are true for a given individual object. Classical and non-monotonic logics are suitable for such inference when the knowledgebase contains only non-probabilistic information. Alternately, Bayesian reasoning can be used if there is sufficient information about a single joint distribution, or perhaps a convex set of distributions. But, if a knowledgebase contains incomplete, imprecise, and conflicting probabilistic information, orthodox approaches break down. Henry Kyburg's theory of *Evidential Probability* (EP) was specifically designed to address reasoning with imperfect statistical information.

At the heart of Evidential Probability are principles for identifying and pruning away irrelevant statistical information, and those principles essentially amount to a solution to the *problem of the reference class*[20]. The theory has been developed within the traditions of the philosophy of science and the foundations of statistics. Accordingly, the mathematical development of EP has aimed at a general theory of EP inference, rather than a concrete KRR system, thus has tended to be abstract. Even so, there have been implementations of the theory, principally CCRS [17], AEPS [14], and SDL [25]. SDL is a qualitative representation of defeasible reasoning common to inferential statistics [24], but it does not address reference class reasoning. As for CCRS and AEPS, both are based on a classical version of EP [11] that has since been abandoned for reasons to do with an unwarranted assumption about the structure of statistical grounds. The classical approach, introduced in [10] and fully developed in [11, 12], sought to bound probability assignments by a single set of reference statistics. But statistical evidence may not always yield a single set of reference statistics, so the classic approach was replaced by one where probability is bound by sets of reference statistics [15]. The semantics for this more general conception of EP inference is based on the notion of 'partial proof', developed by Kyburg in the 1990s, but fully presented in [15], where the inference procedure for partial proof is shown to be sound and a sketch for a completeness proof (provided there is some unspecified representation language) is given. Further refinements of EP inference based on the partial proof semantics appear in [13, 22] and [7], but these formulations did not focus representation languages either, thus they did not address decidability or complexity.

This paper shows how the inference procedure of Evidential Probability, when grounded in a decidable description logic, can be implemented when statistical information is represented by a *Minimal Knowledge with Negation as Failure (MKNF)* knowledge base. Section 1 introduces the theory of Evidential Probability, which relies on new results that help present the theory in a simplified and efficient form. Examples are also given, and an extended example is presented in section 2. Section 3 introduces our implementation of Evidential Probability, including a proof of correctness, and demonstrates the algorithm's low-order polynomial complexity. Section 4 discusses the KRR resources required to support our algorithm by revisiting the extended example of section 2 and showing how features of *Minimal Knowledge with Negation as Failure (MKNF)* provide the support the algorithm requires. Finally, section 5 describes an open-source prototype implementation of the algorithm and we discuss future research.

## 1. Preliminaries

Whereas probabilistic logics use the calculus of probability as a consistency maintenance mechanism to reason probabilistically about logical statements, logics of probability devise logical machinery to reason with probability statements [7]. Evidential Probability [15] is a logic of probability rather than a probabilistic logic. The evidential probability of a sentence $\chi$ given a set of background evidence $\Gamma$, written $\mathrm{Prob}(\chi, \Gamma)$, is not a ratio but a 2-place metalinguistic operator on analogy with provability. A good introduction to EP is [15], and the collected papers in [8] include both recent developments as well as critical essays.

### *1.1. Evidential Probability*

We will follow the treatment of Evidential Probability in [15] and [22]. Let $\mathcal{L}$ be a first-order language; the domain of the probability function $\mathrm{Prob}$ is $\mathcal{L} \times \wp(\mathcal{L})$, and its range is intervals $[l, u]$. EP is designed to assign a probability to the claim that a particular object belongs to a particular class given known statistical evidence relevant to that object, and this evidence is expressed by *statistical probability statements* of the following form:

$$\%\vec{x}(\tau(\vec{x}), \rho(\vec{x}), [l, u]). \tag{1}$$

Schematic 1 says that given a sequence of propositional variables $\vec{x}$ that satisfies the reference class predicate $\rho$, the proportion of $\rho$ that also satisfies the target class predicate $\tau$ is between the real numbers $l$ and $u$ within the unit interval.

Syntactically, '$\tau(\vec{x}), \rho(\vec{x}), [l, u]$' is an open formula schema, where '$\tau(\cdot)$' and '$\rho(\cdot)$' are replaced by open first-order formulas, '$\vec{x}$' is replaced by a sequence of propositional variables, and '$[l, u]$' is replaced by a specific sub-interval of $[0, 1]$. The operator '%' is similar to the ordinary first-order operators ($\forall, \exists$) of first-order logic, except that '%' is a 3-place binding operator over the variables appearing within the *target formula* $\tau(\vec{x})$ and the *reference formula* $\rho(\vec{x})$, and it binds those formulas to an interval.

In practice an object $o$ may belong to several reference classes with known statistics, and selecting the appropriate statistical distribution among the class of potential probability statements is the *problem of the reference class* [20, 12]. In a nutshell, evidential probability is a theory of reference-class reasoning. The task of assigning probability to a statement $\chi$ relative to an evidence set $\Gamma$ relies upon a procedure for eliminating excess candidates from the set of potential probability statements.

**Definition 1 (Potential Probability Statement).** A *potential probability statement* for $\chi$ with respect to $\Gamma$ is a tuple $\langle o, \tau(o), \rho(o), [l, u]\rangle$, such that instances of $\chi \leftrightarrow \tau(o)$, $\rho(o)$, and $\%x(\tau(x), \rho(x), [l, u])$ are each in $\Gamma$.

For example, if it is known that object $o$ satisfies the reference formula $\rho$, and known that between .7 and .8 of $\rho$ are also $\tau$, then $\langle o, \tau(o), \rho(o), [.7, .8]\rangle$ is a potential probability statement for $\chi$ based on the knowledge base $\Gamma$.

It is conceivable that several potential statistical statements linking the reference formula $\rho$ to the target formula $\tau$ may be associated with different probability intervals. In such cases, we shall say that only the strongest potential statistical statements constitutes support for a probability assignment.

**Definition 2 (Support).** Let $\Delta$ be a set of potential probability statements formed by the same reference and target formulas. $\Delta$ provides $[l, u]$ *support* for $\chi$ with respect to $\Gamma$ iff there is a $\delta \in \Delta$ whose interval is $[l, u]$, and for all $\delta^* \in \Delta$, $[l, u] \subseteq [l^*, u^*]$. $\Delta$ provides *undefined support* for $\chi$ if there is no $\delta \in \Delta$ with interval $[l, u]$ such that, for all $\delta^* \in \Delta$, $[l, u] \subseteq [l^*, u^*]$.

Given $\chi$, there are possibly many potential probability statements. Selecting the appropriate probability interval for $\chi$ from this set of potential probability statements reduces to identifying and resolving conflicts among the statistical statements that are the basis for each statement. Thus, the notion of conflict is central to evidential probability.

**Definition 3 (Conflict).** Two intervals $[l, u]$ and $[l', u']$ *conflict*) iff neither $[l, u] \subseteq [l', u']$ nor $[l, u] \supseteq [l', u']$. Two statistical statements conflict iff their intervals conflict.

Given a set of conflicting potential statistical statements with respect to some event $\chi$, some statistical statements in this set may be less relevant than others in computing $\chi$'s probability. In such cases there may be grounds for ignoring less relevant information in favor of more relevant information, which is called *sharpening*. There are three ways that a larger body of information may obscure more accurate information within a smaller body of information. A set of potential probability statements may be sharpened by *richness*, which advises to select information provided by a joint distribution over information provided by a marginal distribution if the two conflict; the set may be sharpened by *specificity*, which advises to select information provided by a more specific class over information provided by a less specific class if the two conflict; and evidence may be sharpened by *precision*, which advises to prefer a statistical statement $S$ with probability $[l, u]$ to statement $S$ with probability $[l', u']$ if $[l, u]$ is a proper subinterval of $[l', u']$.

1. **[Richness]** If $\varphi$ and $\vartheta$ conflict, $\vartheta$ is based on a marginal distribution while $\varphi$ is based on a joint distribution, and $\varphi$ cannot be eliminated by other applications of richness, then eliminate $\vartheta$ from the set of potential probability statements.
2. **[Specificity]** If $\varphi$ and $\vartheta$ conflict and both survive the principle of richness, and if the reference class of $\varphi$ is a subset of the reference class of $\vartheta$, then eliminate $\langle \tau, \rho_\vartheta, [l, u] \rangle$ from the set of potential probability statements.
3. **[Precision]** For potential probability statements $\varphi$ with probability $[l, u]$ and $\vartheta$ with probability $[l', u']$, if $[l, u] \subsetneq [l', u']$ then $\varphi$ is more precise than $\vartheta$ and $\vartheta$ is eliminated from the set of potential probability statements.

A corollary of the precision rule extends precision to sets of potential probability statements.

**Corollary 1.** *For all potential probability statements $\vartheta$ with intervals $[l, u]$, if there exists a statement $\varphi^*$ such that its associated interval $[l^*, u^*] \subsetneq (\bigcap\{[l, u]\})$, then all statements are eliminated except $\varphi^*$.*

Evidence for $\chi$ given background $\Gamma$ is *sharpened* just in case a potential probability statement is removed from consideration by either precision, richness, or specificity. However, there is no guarantee that exhaustively sharpening a set of potential probability statements will yield a unique statistical statement: sometimes the best available evidence is conflicting evidence. So, to ensure a unique probability assignment, evidential probability will need to identify the smallest cover of a set of conflicting intervals.

**Definition 4 (Cover).** An interval $[l, u]$ *covers* a set of intervals $\mathcal{I}$ iff for every $[l', u'] \in \mathcal{I}$, $l \leq l'$ and $u' \leq u$. A cover $[l, u]$ of $\mathcal{I}$ is the *smallest cover* iff, for all covers $[l^*, u^*]$ of $\mathcal{I}$, $l^* \leq l$ and $u \leq u^*$.

Now we can describe the basic idea behind evidential probability. The theory aims to specify norms for identifying and discarding irrelevant information, and it does so in part by maintaining a running account of what statements remain in conflict as statements are pruned away.

Kyburg and Teng simplified the algorithm for sharpening appearing in [15], and this new algorithm is described in [22]. The new algorithm incorporates *support* and *precision* into a *strength* rule, which relies upon sets of conflicting statements constructed by *closure under difference*. We first present evidential probability in terms of Teng's construction, then we introduce an equivalent but more efficient alternative.

**Definition 5 (Closure Under Difference (Teng 2007)).** Given a set of intervals $\mathcal{I}$, a set of intervals $\mathcal{I}'$ is *closed under difference* with respect to $\mathcal{I}$ iff $\mathcal{I}' \subseteq \mathcal{I}$ and $\mathcal{I}'$ contains every interval in $\mathcal{I}$ that conflicts with an interval in $\mathcal{I}'$.

With this notion of conflict for sets of intervals, we may turn our last principle, *strength*. The principle of strength may be thought of as a generalization of the Precision Corollary 2 in that, rather than $[l^*, u^*]$ being an interval associated with a particular statement in the set of potential statistical statements, the strength rule treats $[l^*, u^*]$ as the smallest cover of a set of conflicting intervals. Since an interval is the smallest cover of itself, Teng's formulation of the strength rule incorporates precision as a special case.

4. **[Strength] (Teng 2007)** Let $\Gamma^{RS}$ be the set of *relevant statistical statements* for $\chi$ with respect to $\Gamma$, defined as those statements not discarded by richness and specificity. The evidential probability of $\chi$ is the shortest among all covers of non-empty sets of statistical statements closed under difference with respect to the set $\Gamma^{RS}$; alternatively, it is the intersection of all such covers.

Thus, given a sentence $\chi$ and background theory $\Gamma$, the evidential probability of $\chi$ is calculated by (1) identifying the set of potential probability statements, (2) applying the principle of richness, (3) then specificity, and, finally, (4) strength.

### *1.2. Interval Management*

Whereas the execution of both richness and specificity depends upon orderings of statistical classes associated with potential probability statements, the execution of the precision and strength rule depends on the properties of the intervals associated with potential probability statements. Since our goal is to devise an efficient algorithm for computing evidential probability, we will need an efficient representation of both types of rules. First, we focus on the interval-management rules.

The idea behind Definition 5 is to formulate the strength rule as a generalization of the precision rule, and we incorporate this approach in our definition of *closure under conflict*.

**Definition 6 (Closed Under Conflict).** Let $\mathcal{I}$ be a non-empty finite set of intervals arranged in a sequence $\langle X_1, X_2, \ldots, X_n \rangle$, where $l_i$ and $u_i$ are the lower bound and the upper bound, respectively, of the $i$th interval $X_i$ in $\mathcal{I}$. Define

($i$) $L = \langle X'_1, X'_2, \ldots, X'_n \rangle$ to be a permutation of $\mathcal{I}$ such that for all $j > i$, either $l'_i < l'_j$ or ($l'_i = l'_j$ and $u'_i \geq u'_j$).

($ii$) $U = \langle X''_1, X''_2, \ldots, X''_n \rangle$ to be a permutation of $\mathcal{I}$ such that for all $j > i$, either $u''_i > u''_j$ or ($u''_i = u''_j$ and $l''_i \leq l''_j$).

Then:

1. If $|\mathcal{I}| = 1$, the closure under conflict of $\mathcal{I} = \mathcal{I}$, i.e., $CC(\mathcal{I}) = \{X_1\}$.
2. Otherwise, define $S_0 = \{X'_n, X''_n\}$ and the smallest cover of $S_0$ as $[l_0, u_0]$. Let $i = 0$.
   (a) Then define $S_{i+1}$ as the set $S_i \cup \{X_k \in \mathcal{I}$ : either the lower bound of $X_k$ is greater than the lower bound of the cover of $S_i$, $l_k > l_i$, or the upper bound of $X_k$ is less than the upper bound of the cover of $S_i$, $u_k < u_i\}$.
   (b) Repeat until $S_{i+1} = S_i$, in which case denote by $S$.
   (c) Then, $CC(\mathcal{I}) = S$.

**Example 1.** Let $\mathcal{I} = \{[.25, .35], [.4, .4], [.2, .45], [.3, .3], [.2, .5]\}$. Then, with respect to $\mathcal{I}$,

- $L = \langle [.2, .5], [.2, .45], [.25, .35], [.3, .3], [.4, .4] \rangle$, and
- $U = \langle [.2, .5], [.2, .45], [.4, .4], [.25, .35], [.3, .3] \rangle$.

The $CC(\mathcal{I}) = \{[.25, .35], [.4, .4], [.3, .3]\}$.

**Theorem 1.** *If a non-empty set of intervals $\mathcal{I}$ is closed under conflict, then $\mathcal{I}$ satisfies the Precision corollary 2.*

PROOF. Suppose that $\mathcal{I} = \langle X_1, \ldots, X_n \rangle$ is a set of intervals such that there exists an $X_i \in \mathcal{I}$ such that $X_i \subsetneq \bigcap\{X_1, \ldots, X_{i-1}, X_{i+1}, X_n\}$. By precision, $X_i$ should be preserved and $\{X_1, \ldots, X_{i-1}, X_{i+1}, X_n\}$. To show that this holds, observe that by Definition 6, $S_0 = \{X_i, X_i\}$ and, by hypothesis, there is no $X_k \in \mathcal{I}$ such that either the lower bound $l_k > l_i$, or such that the upper bound $u_k < u_i$. Hence, $S_{i+1} = S_i = S_0$. So, $CC(\mathcal{I}) = \{X_i\}$. $\dashv$

**Theorem 2.** *Let $\mathcal{I}$ and $S$ be be non-empty, finite sets of intervals. If $CC(\mathcal{I}) = S$, then the shortest cover of $S$ is equal to the shortest among all covers of non-empty sets closed under difference with respect to $\mathcal{I}$.*

PROOF. Suppose $|\mathcal{I}| = n$ and $CC(\mathcal{I}) = S$. We show that if $CC(\mathcal{I}) = S$, then the shortest cover of $S$ is equal to the shortest among all covers of non-empty sets closed under difference with respect to $\mathcal{I}$ by induction over $S$.

1. $S = S_0$. By Definition 6, $S_0 = \{X'_n, X''_n\}$ and for all $X_k \in \mathcal{I}$, both $l_k \leq l_n$ and $u_k \geq u_n$ hold. So $X'_n = X''_n$ and $CC(\mathcal{I}) = \{X_n\}$. Thus, (i) $S_0$ is a subset of $\mathcal{I}$, (ii) $S_0$ is closed under difference with respect to $\mathcal{I}$, and (iii) among all covers of non-empty sets closed under difference with respect to $\mathcal{I}$, no cover is shorter than the shortest cover of $S_0$. Therefore, the shortest cover of $\{X_n\}$ is the shortest among all covers of non-empty sets closed under difference with respect to $\mathcal{I}$.
2. $S = S_{i+1}$: By Definition 6, $S_{i+1} = S_i \cup \emptyset$, so for all $X_k \in \mathcal{I}$, both $l_k \leq l_i$ and $u_k \geq u_i$ hold, and $CC(\mathcal{I}) = S_i$. Suppose $[S_i]$ denotes the shortest cover of the set of intervals $S_i$, and $\min([\mathcal{X}])$ denotes the shortest of all covers of non-empty sets closed under difference with respect to $\mathcal{I}$. Observe that both $\mathcal{X}$ and $S_i$ are subsets of $\mathcal{I}$. Now suppose that $l_i < l_{\mathcal{X}}$, where $l_i$ is the lower bound of $[S_i]$ and

$l_{\mathcal{X}}$ is the lower bound of $\min([\mathcal{X}])$. Then, either (i) $u_i > u_{\mathcal{X}}$ or (ii) $u_i \leq u_{\mathcal{X}}$. If (i) were true, then it is false that for all $X_k \in \mathcal{I}$, both $l_k \leq l_i$ and $u_k \geq u_i$ hold. So, $CC(\mathcal{I}) \neq S_i$, which contradicts our hypothesis. If (ii) were true, then $\mathcal{X}$ would conflict with $S_i$, so the shortest cover of $\mathcal{X}$ would not be the the shortest of all covers of non-empty sets closed under difference with respect to $\mathcal{I}$. An symmetrical argument holds supposing that $u_i > u_{\mathcal{X}}$. Hence,

$$[S_i] \subseteq \min([\mathcal{X}]), \tag{2}$$

since otherwise there would be a pair of conflicting intervals in $S_i \subseteq \mathcal{I}$ that were not covered by the shortest of all covers of non-empty sets closed under difference with respect to $\mathcal{I}$. So, $S_i$ is closed under difference with respect to $\mathcal{I}$. Furthermore,

$$[S_i] \not\subset \min([\mathcal{X}]), \tag{3}$$

since otherwise either the lower-bound of $\min([\mathcal{X}])$ would be greater than the lower bound of the cover of $S_i$, or the upper bound of $\min([\mathcal{X}])$ would be less than the upper-bound of $S_i$. But, either case describes a nonempty set of intervals which conflict with $S_i$. But, then $S_{i+1} \neq S_i \cup \emptyset$, hence $S \neq S_{i+1}$, contradicting our supposition. Therefore,

$$[S_i] = \min([\mathcal{X}]), \tag{4}$$

which concludes the proof. ⊣

**Corollary 2.** *Precision is a special case of closure under conflict, and a singleton set sharpens itself by precision:*

1. *If $CC(\mathcal{I}) = S_0$, then $S_0$ sharpens $\mathcal{I}$ by Precision.*
2. *If $\mathcal{I} = \{X_1\}$, then both*
   (a) *$CC(\mathcal{I}) = \{X_1\}$, and*
   (b) *$\{X_1\}$ is the only non-empty set closed under difference with respect to $\mathcal{I}$.*

PROOF. Corollary 2.2 is immediate. For Corollary 2.1, recall that $CC(\mathcal{I}) = S_0$ holds iff, for $|\mathcal{I}| = n$, $S_0 = \{X'_n, X''_n\}$ and $X'_n = X''_n$. ⊣

### *1.3. EP Examples*

We now turn to some examples.

**Example 2.** Suppose that a statement $\chi$ about object $o$ has these five potential probability statements:

1. $\langle o, A(o), P(o), [.25, .35]\rangle$
2. $\langle o, B(o), Q(o), [.4, .4]\rangle$
3. $\langle o, C(o), R(o), [.2, .25]\rangle$
4. $\langle o, D(o), S(o), [.3, .3]\rangle$
5. $\langle o, E(o), T(o), [.2, .5]\rangle$

Call this set of potential probability statements $\Gamma$. Now consider three different evidential situations involving $\Gamma$

- *Total knowledge*: If $\Gamma$ is all that is known about $\chi$ and there is no additional information about the relationships between the statistical classes underlying these statements, then the set of conflicting statements is $\{1, 2, 4\}$, which is identical to $CC(\mathcal{I})$ in Example 1. Neither richness nor specificity succeeds in pruning away potential probability statements, since there is no additional information. So, Strength assigns the shortest cover, $[.25, .4]$, as the probability that $\chi$

- *Specific information on reference classes*: Consider now the following two relationship among reference classes:

  a) $S \subsetneq Q$
  b) $Q \subsetneq S$

Consider the evidential situation in which a) is known. Then, the initial set of conflicting statements is $\{1, 2, 4\}$ as before. However, the clause a) allows for the specificity rule to be applied, yielding the set $\{1, 4\}$. But notice that the corresponding set of intervals for this set is not closed under conflict:

$$CC(\{[.25, .35], [.3, .3]\}) \neq \{[.25, .35], [.3, .3]\}$$

Instead, $CC(\{[.25, .35], [.3, .3]\}) = \{[.3, .3]\}$. So, by strength (trivially), the probability is .3

Compare now the evidential situation in which b) is known instead. The initial set of conflicting statements is $\{1, 2, 4\}$, but the application of specificity removes $[.3, .3]$ in this case rather than $[.4, .4]$. But notice that in this case there is no change, that is

$$CC(\{[.25, .35], [.4, .4]) = CC(\mathcal{I})$$

So, by strength, the probability is $[.25, .4]$.

**Example 3.** Suppose the proportion of white balls ($W$) in an urn ($U$) is known to be within $[.33, 4]$, and that ball $o$ is drawn from $U$. These facts are represented in $\Gamma$ by the sentences, $\%x(W(x), U(x), [.33, .4])$ and $U(o)$.

1. If these two statements are all that we know about $o$, i.e., they are the only statements in $\Gamma$ pertaining to $o$, then $\text{Prob}(W(o), \Gamma) = [.33, .4]$.
2. Suppose additionally that the proportion of plastic balls ($P$) that are white is observed to be between $[.31, .36]$, $o$ is plastic, and that every plastic ball is a white ball. That means that $\%x(P(x), U(x), [.31, .36])$, $P(o)$, and $\forall x.P(x) \rightarrow W(x)$ are added to $\Gamma$ as well. Then there is conflicting statistical knowledge about $o$, since either:
   (a) the probability that ball $o$ is white is between $[.33, .4]$,
   by reason of $\%x(W(x), U(x), [.33, .4])$, or

(b) the probability that ball $o$ is white is between $[.31, .36]$,
by reason of $\%x(W(x), P(x), [.31, .36])$,

may apply. There are several ways that statistical statements may conflict and there are rules for handling each type, which we will discuss in the next section. But in this particular case, because it is known that the class of plastic balls is more *specific* than the class of balls in $U$ and we have statistics for the proportion of plastic balls that are also white balls, the statistical statement in (2) dominates the statement in (1). So, the probability that $o$ is white is between $[.31, .36]$.

3. Adapting an example from [15, 216], suppose $U$ is partitioned into three cells, $u_1$, $u_2$, and $u_3$, and that the following compound experiment is performed. First, a cell of $U$ is selected at random. Then a ball is drawn at random from that cell. To simplify matters, suppose that there are 25 balls in $U$ and 9 are white such that 3 of 5 balls from $u_1$ are white, but only 3 of 10 balls in $u_2$ and 3 of 10 in $u_3$ are white. The following table summarizes this information.

Table 1: Compound Experiment

| | $u_1$ | $u_2$ | $u_3$ | |
|---|---|---|---|---|
| $W$ | 3 | 3 | 3 | 9 |
| $\overline{W}$ | 2 | 7 | 7 | 16 |
| | 5 | 10 | 10 | 25 |

We are interested in the probability that $o$ is white, but we have a conflict. Given these over all precise values, we would have $\mathrm{Prob}(W(o), \Gamma) = \frac{9}{25}$. However, since we know that $o$ was selected by performing this compound experiment, then we also have the conflicting direct inference statement

$$\%x, y(W^*(x, y), U^*(x, y), [.4, .4]),$$

where $U^*$ is the set of compound two stage experiments, and $W^*$ is the set of outcomes in which the ball selected is white.[1] We should prefer the statistics from the compound experiment because they are *richer* in information. So, the probability that $o$ is white is $.4$.

4. Finally, if there happens to be *no* statistical knowledge in $\Gamma$ pertaining to $o$, then we would be completely ignorant of the probability that $o$ is white. So in the case of total ignorance, $\mathrm{Prob}(W(o), \Gamma) = [0, 1]$.

The principle of specificity says that if it is known that the reference class $\rho_\vartheta$ is included in the reference class $\rho_\varphi$, then eliminate the statement $\varphi$. The statistical statements that survive the sequential application of the principle of richness followed by the principle of specificity are called *relevant statistics*.

[1] $\Gamma_\delta$ should also include the categorical statements $\forall x, y(U^*\langle x, y\rangle \rightarrow W(y))$, which says that the second stage of $U$ concerns the proportion of balls that are white, and three statements of the form $W^*(\mu, o) \leftrightarrow W(o)$, where $\mu$ is replaced by $u_1, u_2, u_3$, respectively. This statement tells us that everything that's true of $W^*$ is true of $W$, which is what ensures that this conflict is detected.

## 2. An Extended Example

Pig farmers have a strong preference for white pigs and have selectively bred for white coat color since the practice began in medieval Europe [26]. White coat color in domestic pigs is due to two mutations in the *KIT* gene [18]. Although *KIT* mutations control coat color in other mammals, they are often responsible for pigmentation disorders. *KIT* mutations are often lethal in mice, for instance, having a pleiotropic effect on cell development concerning skin, blood, and the small intestines, and may affect hearing, too. The dominant white allele in domestically bred pigs has a stronger effect on pigmentation than it does in mice, but without the selective disadvantage observed in mice. Indeed, there appears to be a selective *advantage* for white domestic pigs over non-white domestic pigs.

Imagine that a laboratory has conducted a study of the selective disadvantage of non-white domestic pigs to white domestic pigs.[2] Coat color in domestic pigs, in our toy theory, is controlled by two alleles at a single locus, the recessive $i$ for color and $I$ allele for the dominant white phenotype. The quantities that interest us include:

1. **Color**: The color of a pig's coat is represented by the binary variable $C$, which takes the value 1 for white pigs and 0 for non-white pigs.
2. **Genotype**: The discrete variable *Gene* has the value 0 for $II$, 1 for $Ii$ (or, equivalently, $iI$), and 2 for $ii$.
3. **Generation**: The discrete variable $G$ has the value $i - 1$ for the $i$th generation, for $i \in \{1, 2, 3, 4, 5\}$.
4. **Frequency of white pigs**: $Freq_{C=1}$ gives the proportion of white pigs in a specified set of pigs.

In addition to these quantities, we shall use $N_1, \ldots, N_{15}$ to denote the 15 pens in which a pig can live, $o_1, ..., o_{1500}$ as the set of randomly and uniquely assigned identification numbers for all 1500 pigs in the study, and the relation $Lives(x, y)$ which holds when $x$ is a pig, $y$ is a pen, and $x$ lives in $y$. The study began with a population of 100 pigs and data was collected for five generations. Each generation increased by 100 pigs, and each pen contained 100 pigs of only one generation.

Now let us specify the knowledge base, $\Gamma$. In addition to the genetic information just mentioned, we know that in the initial drove of 100 pigs the distribution of the genotypes is: 0.09 $II$, 0.42 $Ii$, and 0.49 $ii$. It follows from this distribution that the frequency of $I$ alleles is 0.3, the frequency of $i$ alleles is 0.7, and the proportion of white pigs is $51/100$. Let's suppose that in any pig population the distribution of genotypes is multinomial, $M(p_I^2, 2p_I(1 - p_I), (1 - p_I)^2)$, where $p_I$ is the proportion of $I$ alleles. The data covers five generations of random mating starting with our initial drove: $p_I$ is 0.3 for *G* = 0, within [0.348, 0.367] for *G* = 1, within [0.401, 0.439] for *G* = 2, within [0.450,0.512] *G* = 3, and within [0.502,0.576] for *G* = 4. We use '$S$' to denote the entire population in the study. Also included in $\Gamma$ are the logical consequences of this information. For example, it follows that $p_i$ when $G = 2$ is between $[.561, .599]$. The following table collects the general information in $\Gamma$ that we'll use in our example.

[2] From this point on, our example uses fictitious data and replaces the actual genetic mechanism for coat color with a simplified mechanism.

Table 2: General Knowledge in $\Gamma$

| $G$ | $p_I$ | $p_{II}$ | $p_{iI}$ | $Freq_{C=1}$ | $\|G\|$ |
|---|---|---|---|---|---|
| 0 | .3 | .09 | .42 | .51 | 100 |
| 1 | $[.348, .367]$ | $[.121, .135]$ | $[.441, .479]$ | $[.562, .614]$ | 200 |
| 2 | $[.401, .439]$ | $[.161, .193]$ | $[.450, .526]$ | $[.611, .719]$ | 300 |
| 3 | $[.450, .512]$ | $[.203, .262]$ | $[.439, .563]$ | $[.642, .825]$ | 400 |
| 4 | $[.501, .576]$ | $[.252, .332]$ | $[.426, .574]$ | $[.678, .906]$ | 500 |
| $S$ | $[.433, .485]$ | $[.193, .245]$ | $[.436, .538]$ | $[.628, .782]$ | 1500 |

In addition to this general knowledge in $\Gamma$, there is also specific knowledge about some of the pigs. Pig $o_{18}$ lives in $N_{12}$ and $G(o_{18}) = 1$. Pig $o_{1351}$ lives in some pen or another. Pig $o_{662}$ lives in $N_{10}$. The set of pigs $\{o_{1428}, o_{1366}, o_{1250}, o_{986}\}$ are all members of the second generation. Pig $o_{333}$ lives in $N_6$. Pig $o_{333}$ is White.

While it is immediately clear how to represent specific information in logical form, let's pause to consider how general knowledge is represented in $\Gamma$. Quantity (4), the frequency of white pigs, implicitly refers to different reference classes of pigs. For instance, the frequency of white pigs in the fourth generation is between $0.642$ and $0.825$, and between $0.628$ and $0.782$ in the total population of 1500 pigs in the study. Within EP, each claim is represented by

$\%x(Freq_{C=1}(x), G_{=1}(x), [.642, .825])$ and

$\%x(Freq_{C=1}(x), S(x), [.628, .782])$, respectively.

$G = 1$ and $S$ are reference formulas, denoting reference classes for which we have statistics for the target formula, $Freq_{C=1}$. The fact that $G = 1$ is a proper subset of $S$ is also represented in $\Gamma$.

With this setup, we may illustrate some assignments of evidential probability relative to $\Gamma$:

1. $\text{Prob}(Gene(o_{18}) = 0, \Gamma) = [0.203, 0.261]$. The frequency of the $II$ genotype in generation 4 ($G = 3$) is [0.203, .261], but the overall frequency of the $II$ genotype in the full population of the study is [0.193, 0.245], which conflicts with the frequency in generation 4. However, the specificity rule advises that we ignore the general statistic for $S$, $[0.193, 0.245]$, in favor of the statistic for the class $G = 1$, $[0.121, 0.135]$.

2. $\text{Prob}(White(o_{18}) = 1, \Gamma) = [.562, .614]$. Even though the overall frequency of white pigs is [.628, .782], the probability that $o_{18}$ is a white pig is between $[.562, .614]$. Since $\{x|x \in G = 1\} \subsetneq S$, $\%x(Freq_{C=1}(x), G_{=1}(x), [.562, .614])$, and $\%x(Freq_{C=1}(x), S(x), [.628, .782])$ are in $\Gamma$ *and no other information relevant to* $o_{18}$ is in $\Gamma$, then by specificity we select the more specific reference class statistic.

3. $\text{Prob}(White(o_{100}) = 1, \Gamma) = [.628, .782]$. All that we know about pig $o_{100}$ is that it is in the experiment. Since there is no information in $\Gamma$ about $o_{100}$ belonging to a subclass which dominates the frequency information for the population, the frequency information for $S$ is used.

4. $\text{Prob}(Gene_{=0}(o_{333}), \Gamma) = [.307, .313]$. We know that $o_{333}$ lives in pen $N_6$, but there is no information about the proportion of white pigs in $N_6$. So, although $N_6 \subsetneq S$ and $o_{333} \in N_6$, there is no direct inference statement in $\Gamma$ recording the proportion of white pigs in $N_6$. We do know that $o_{333}$ is white, however. Since White pigs are a sub-class for which we have statistics, specificity advises we use the estimate of the proportion of those that are of genotype $II$, [.307,.313], rather than the estimate for the proportion of $S$ that are $II$, $[.193, .245]$.

5. We select 100 pigs from the full population $S$ and observe that 17 are afflicted by polydactyly. If this is all that we know, we may perform a confidence interval analysis and assign a probability of 0.95 that the rate of polydactyly in $S$ is between $12\%$ and $21\%$. If $0.95$ counts as a level of practical certainty, we will simply accept *by default* that between $12\%$ and $21\%$ of the pigs in $S$ have dewclaws.

   Nevertheless, suppose that the rate of polydactyly in domestic pigs is very well understood in the literature to be ($A$) between $0\%$ and $11\%$ about $80\%$ of the time, ($B$) between $12\%$ and $21\%$ about $15\%$ of the time, and ($C$) greater than $21\%$ no more than $5\%$ of the time. In fact, we can see that the observed frequency of 17 in 100 of observed cases ($E$) given $A$, $B$, and $C$ is .10, .50, and .40, respectively. With this (approximate) distribution of long-run frequencies of polydactyly in domestic swine specified, we can apply a *Bayesian analysis* of our estimate of polydactyly in our population,

$$P(E \mid B) = \frac{P(B \mid E)P(B)}{P(B \mid E)P(B) + P(A \mid E)P(A) + P(C \mid E)P(C)} \approx 0.429.$$

   $P(E|B) \approx 0.429$ conflicts with the interval $[0.95, 1]$ for the hypothesis that between 12% and 21% of the pigs in $S$ have dewclaws, which we constructed by confidence methods. The richness rule specifies that the probability based on prior frequency knowledge be preferred to the $[0.95, 1]$ interval, which would *defeat* accepting that between 12% and 21% of the pigs in $S$ have dewclaws.

6. Suppose that a pen is chosen at random and a pig from that pen, $o^*$, is selected from that pen at random. If this is all that is known, then $\text{Prob}(White(o^*), \Gamma) = [.601, .871]$ is the probability that a pig is born white. The interval $[.601, .871]$, calculated by taking the average of the probabilities for being white in each generation, does not conflict with the overall proportion of white pigs in the population, $[.628, .782]$. So $[.628, .782]$ is not eliminated in favor of $[.601, .871]$ by the richness rule, because the two statistics are not in conflict. Both remain relevant.

Now suppose that in addition we know two things about $o^*$. First, $o^*$ is the off-spring of $o_{1428}$ and $o_{986}$, both of which are in $G = 1$. So, $o^* \in G = 2$ and the frequency of white pigs in $G = 2$ is $[.611, .719]$. Second, $o^*$ has the genotype $Ii$, and suppose that the proportion of pigs of genotype $Ii$ that are white is exactly .57. Both $[.57, .57]$ and $[.611, .719]$ are in conflict with each other, and with $[.601, .871]$. But, by specificity, $[.601, .871]$ is ignored. Then by strength, $\mathrm{Prob}(White(o^*), \Gamma) = [.57, .57]$.

## 3. Computing Evidential Probability

The computation of Evidential Probability makes use of concepts both from description logics and from logic programming, although no detailed knowledge of these areas is assumed in this paper [3]. However, we do review aspects of the notation of logic programming and of description logics that are used. In logic programming notation

- Constants, function symbols and predicate symbols all begin with a lower-case letter or a number, while variables begin with a capital letter. Predicates and functions may have any arity.
- Conjunction is denoted via the comma ',', while a clause has the form

  $$Head \text{ :- } Body.$$

  where *Head* is an atom and *Body* is a conjunction of literals, where a literal has the form *Atom* or the form *not Atom* for some atom *Atom*.

Thus, in logic programming an assertion for a unary predicate would be denoted as e.g. *grandmother(rose)*, for a binary predicate as e.g. *parentof(rose,mary)*, etc.

Description Logics (DLs) vary in the types of operators they allow. However, in the most common DLs,

- Only constant functions (objects) and unary and binary predicates are allowed. A symbols for a constant function begins with a non-capitalized letter; a symbol for a predicate begin with a capitalized letter.
- Logical quantifiers and operators are employed to construct sentences. Because only certain patterns of quantification are allowed, DL formulas do not require the explicit use of variables.

Thus, that rose is a grandmother would be denoted as $Rose \in Grandmother$. Further details can be found in e.g. [3].

In general, our pseudo-code makes use of logic programming notation unless otherwise specified.

[3] Background on Description Logics can be found in [3], while background on logic programming can be found in [16].

*3.1. Background Knowledge*

The representation of a background knowledge supporting the algorithms for computing Evidential Probability will be described in Section 4.1. For now, we simply review what sort of information that representation must provide. In this section, we assume that all background knowledge is provided by a DL theory $\Gamma$. For a statistical statement

$$\%(\mathit{Target}, \mathit{Class}, \mathit{Lower}, \mathit{Upper}).$$

we require that both $\mathit{Target}$ and $\mathit{Class}$ are named classes in $\Gamma$, rather than class expressions. For instance, $\mathit{Target}$ would be required to be the class name $\mathit{Grandmother}$ rather than a class expression

$$\mathit{Female} \cap \mathit{Person} \cap (\exists \mathit{ParentOf}(\exists \mathit{ParentOf}))$$

that expresses the intersection of $\mathit{Female}$, $\mathit{Person}$, and of objects that are a $\mathit{ParentOf}$ some other object that is also a $\mathit{ParentOf}$ an object. The restriction to class names per se does not restrict expressivity, as a class name can always be associated with a class expression via an axiom such as:

$$\mathit{Grandmother} \equiv \mathit{Female} \cap \mathit{Person} \cap (\exists \mathit{ParentOf}(\exists \mathit{ParentOf}))$$

At the same time, the use of (named) class expressions in a DL implicitly restricts $Target$ and $Class$ to be formulas with exactly 1 open variable, as opposed to the general form of Schematic 1 of Section 1..

Given the above form of statistical statements, the requirements on the reasoning mechanism for background knowledge $\Gamma$ are as follows. Given a target class name $Target$ and an object $o$, there must be an effective procedure to determine

- Those statistical statements about $Target$ that are *relevant* to $o$. I.e. all statements of the form $\%(T', C, l, u)$ such that $\Gamma \vdash (Target \equiv T')$ and $\Gamma \vdash (o \in C)$.
- For two classes $C_1$ and $C_2$, whether $C_1 \subseteq C_2$: in other words, whether
$$\forall X[C_1(X) \Rightarrow C_2(X)].$$
- For two classes $C_1$ and $C_2$, whether $C_1$ is richer than $C_2$

*3.2. Overview of the Algorithm*

Figure 3.2 shows EPInfer, the overall procedure for determining the probability that a given object $o$ belongs to a target class $\mathit{Target}$, assuming as above that there is an effective procedure for obtaining relevant statistical statements, and for querying about the inclusion and richness relations for classes.

First, the set of statistical statements relevant to $o$ is obtained using the procedure getRelevantStatements, which is further discussed in Section 4.1. We note however, that this set is assumed to be finite. Once the relevant statements have been determined, conflicting statements are sharpened by richness and then sharpened by specificity as described in Section 3.3. Closure under conflict is then applied to the surviving statements, which combines the principles of precision and strength to select the strongest evidence for a probability assignment. Then, the minimum cover of the strongest evidence is taken by $lower$ and $upper$.

```
/* Assumes background knowledge of
      1) Evidential Probability statements;
      2) a binary inclusion relation for classes; and
      3) a binary richness relation for classes. */

EPInfer(Target, o)
      getRelevantStatements(Target, o, RelevantStatements)
      sharpenByRichness(RelevantStatements, RichStatements)
      sharpenBySpecificity(RichStatements, RichSpecificStatements)
      closeUnderConflict(RichSpecificStatements, Statements)
      lower = the minimum p such that %(t, c, p, q) ∈ Statements
      upper = the maximum q' such that %(t', c', p', q') ∈ Statements
      return [lower, upper]
```

Figure 1: The Top-Level Algorithm for Inferring Evidential Probability

*3.3. Sharpening by Specificity and by Richness*

From a computational point of view, sharpening by specificity and sharpening by richness are analogous procedures: each uses a binary relation, representing either richness or specificity, to reduce a set of statistical statements. In this section, we focus our attention on sharpening by specificity and later sketch how the algorithm can be applied to richness. Figure 3.3 shows pseudo-code for sharpenBySpecificity. The procedure takes as input a set $Input$ of statistical statements that are relevant to a given object $o$ and target $Target$, and that are assumed to have been sharpened by richness. It outputs the subset of $Input$ that is further sharpened by specificity.

The first step of sharpenBySpecificity is to construct a specificity graph $SpecG$ from its input and its background knowledge in the following manner. From $Refs$, the set of reference classes occurring in some formula in $Input$, a quotient set (with respect to $\Gamma$-equivalence), $Classes$, is chosen. In other words, each element of $Refs$ is $\Gamma$-equivalent to some element of $Classes$, but no element in $Classes$ is $\Gamma$-equivalent to any element in $Classes$ except itself. The nodes of $SpecG$ are $o$ plus the elements of $Classes$. There is a directed edge in $SpecG$ between $o$ and all $C \in Classes$ of which it is an immediate member, i.e. those elements in $Classes$ that are not proper superclasses of any other elements in $Classes$. Further, for $C_1, C_2 \in Classes$, there is a directed edge between $C_1$ and $C_2$ if $C_1$ is a proper subclass of $C_2$. Since $Classes$ is a quotient set with respect to $\Gamma$-equivalence, $SpecG$ will be a directed acyclic graph. Within $SpecG$, the length of a path between two nodes is simply the number of edges between them. Using this definition, the maximal path length is taken between $o$ and all other nodes in $SpecG$: because $SpecG$ is acyclic, all such paths will be finite.

Using $SpecG$ to structure a search, sharpening proceeds by traversing $SpecG$ from the most specific classes to the least specific classes, potentially sharpening the set of input statements at each stage, and maintaining the sharpened set in $statements(k)$, where $k$ can be thought of as a counter of the sharpening actions. Specifically, the

```
/* Assumes a global target class Target and object o */
sharpenBySpecificity(Input, Output)
    let Refs = {R|%(Target, R, l, u) ∈ Input}
    let Classes be a Γ-equivalent quotient set of Refs
    construct SpecG from o and Classes as described in
                                the accompanying text
    compute the maximal path length in SpecG between o
                                and each C ∈ Classes
    set Max to the maximal path length between o and any C ∈ Classes
    statements(0) = Input; k = 0
    for (i = 1 to Max)
        frontier(i) = {C ∈ Classes|i is the maximal path length in SpecG
                        between o and C and there is at least one statement
                        with reference class C in statements(k)}
        while (frontier(i) ≠ ∅)
            choose N_F ∈ frontier(i); frontier(i) = frontier(i) − {N_F};
            for each N′ such that N_F ≡ N′
                if Support(N′, statements(k)) does not exist, abort; otherwise
                for (C ∈ Refs|C ≡ C′ and there is an edge between N′ and C′)
                    sharpen(B′, C)
    Output = statements(k)

sharpen(Node, Class)
for each %(Target, C, l, u) ∈ statements(k)
if conflicts(Support(Node, statements(k)), [l, u])
    statements(k) = statements(k) − %(Target, C, l, u); k = k + 1
```

Figure 2: Pseudo-code for Sharpening by Specificity

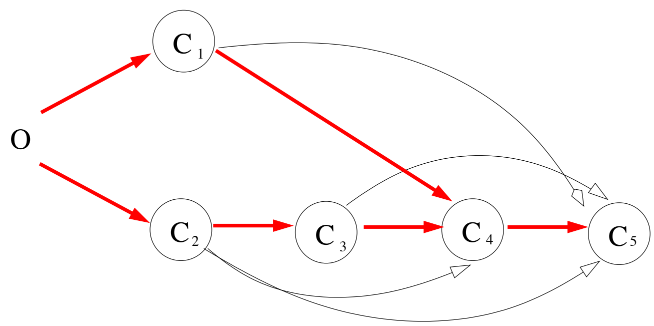


Figure 3: $SpecG$ for the class $\mathcal{C}_Q$ of classes

algorithm starts with $statements(0)$ set to $Input$. At each iteration, indexed by $i$, a set of reference classes $frontier(i)$ is chosen whose maximal path from $o$ in $SpecG$ is $i$ and whose support is non-trivial; the support of these reference classes is used to sharpen away statistical statements whose reference classes are less specific than those in $frontier(i)$. For the first iteration, $frontier(1)$ consists of those classes in $SpecG$ maximal path from $o$ is 1, i.e. those classes of which $o$ is an immediate member. For each $N_F$ in $frontier(1)$, the algorithm iterates through each reference class $N'$ that is $\Gamma$-equivalent to $N_F$. $N'$ is first checked to ensure it has a well-defined support: that is, that there are not statements $\%(Target, N', l_1, u_1)$ and $\%(Target, N', l_2, u_2)$ that conflict. If $N'$ has a well-defined support, then the support for $N'$ in $statements(k)$ is used to sharpen any statements for reference classes to which that are proper supersets of $N'$ and that conflict with the support of $N'$. Once the edges of frontier nodes at iteration step 1 have been fully explored, $i$ is incremented and the algorithm proceeds to use as a frontier those nodes in $SpecG$ whose maximal path from $o$ is $i$. For each new node $N_F$ sharpening proceeds as before, terminating when $SpecG$ has been fully explored.

**Example 4.** Suppose an objet $o$ is a member of all classes in

$$\mathcal{C} = \{C_1, C_2, C_3, C_4, C_5, C_6\}$$

for which the inheritance relation can be axiomitized as

$$C_1 \subset C_4 \quad C_2 \subset C_3 \quad C_3 \subset C_4 \quad C_4 \subset C_5 \quad C_3 \equiv C_6$$

Then the set $\mathcal{C}_Q = \{C_1, C_2, C_3, C_4, C_5\}$ forms a $\Gamma$-equivalent quotient set for $\mathcal{C}$. $SpecG$ for $\mathcal{C}_Q$ is shown in Figure 4. In Figure 4 there is an edge between $O$ and the most specific classes to which it belongs: $C_1$ and $C_2$. There is also an edge from each class $C_j$ and each of its proper supersets.

sharpenBySpecificity traverses the graph of Figure 4 as follows. In iteration 1, $frontier(1)$ consists of $C_1$ and $C_2$, iteration 2 of $C_3$, iteration 3 of $C_4$, and iteration

4 of $C_5$. As $C_6$ is equivalent to $C_3$, it is also used for sharpening in iteration 2. Note that although there is a path of length 2 between $O$ and $C_4$ the *maximal* path length is 3, so that $C_4$ is not chosen until iteration 3. To see the importance of this, consider a situation where

$$conflicts(C_3, C_4) \qquad conflicts(C_4, C_5)$$

If $C_4$ were chosen as a frontier node in iteration 2, its support could be used to remove the statistical statement for $C_5$ *before* $C_4$ was removed due to its conflict with $C_3$. Using of maximal path lengths to chose frontier nodes avoids situations in which a class whose support is used to sharpen away statistical statements later has its own support removed.

*Sharpening by Richness.* Sharpening by richness is performed in an analogous manner to sharpening by specificity, except that it is based on a partial order for richness rather than for class inclusion. Given an input set $Input$, a quotient set is first chosen based on equivalence with respect to richness. A graph $RichG$ is then constructed so that there is an edge between a given object $o$ and those reference classes of that are maximal with respect to richness. In addition, if $C_1$ is strictly richer than $C_2$ a directed edge from $C_1$ to $C_2$ is also constructed. Once $RichG$ is constructed, it is traversed exactly as $SpecG$ to sharpen a set of statistical statements.

### *3.4. Correctness of* sharpenBySpecificity

To demonstrate the correctness of sharpenBySpecificity, we begin with a few lemmas that characterize its behavior. The first lemma shows that if a given reference class is used for sharpening, its support will not change as the algorithm proceeds.

**Lemma 1.** *In the execution of* sharpenBySpecificity*, suppose that for $F, Class \in Refs$,* sharpen$(F, Class)$ *is called with $k = k_1$. Then for any $k_n > k_1$,*

$$support(F, statements(k_1)) = support(F, statements(k_n))$$

PROOF. Note that the support of $F$ will be changed between $k_1$ and $k_n$ only if there is a statement of the form $\%(T, F, l, u)$ in $statements(k_1)$ that is not contained in $statements(k_n))$. For this to occur sharpen$(C, F)$ would need to be called for some $C \in Refs$ when $k > k_1$. Because of the structure of $SpecG$, this means that $C$ must be a subclass of $F$. Let $i_1$ be the value of $i$ when $k = k_1$ – i.e. the iteration number in which sharpen$(F, Class)$ is called. Next, let $i_n$ be the value of $i$ at the time of the call to sharpen$(C, F)$. Clearly $i_n \geq i_1$. If $i_n = i_1$, then $F$ is $\Gamma$-equivalent to a node $N$ in $SpecG$ whose maximal distance from $o$ is $i_1$; however, since $C \subset F$, $C$ is also $\Gamma$-equivalent to a node $N'$ in $SpecG$ whose maximal distance from $o$ is $i_1$. By construction of $SpecG$, $N$ cannot equal $N'$ as $SpecG$ is a DAG. There is therefore a path from $o$ to $F$ through $C$, contradicting the fact that the maximal distance from $o$ to $F$ is $i_1$. A similar argument shows that it cannot be the case that $i_n > i_1$.$\dashv$

Recall that $SpecG$ represents $o$ together with all of the statistically relevant classes to which it belongs, along with the proper subclass relation on those classes. At each iteration $i$, a portion of $SpecG$ has been traversed. This portion consists of all nodes that have been chosen previously in $frontier(i'), i' \leq i$ along with all edges originating in one of these nodes. This traversed portion, can be viewed as a subgraph, $SpecG(i)$, of $SpecG$ The following lemma states that $SpecG(i)$ directly determines the portion of sharpening that has been performed through iteration $i$.

**Lemma 2.** *Let $i$ be an iteration of* SharpenBySpecificity*, and $k_i$ be the value of $k$ at the end of iteration $i$. Then at the end of iteration $i$, $SpecG(i)$ contains no two nodes $C_{sub}$ and $C_{sup}$ such that*

- $C_{sub} \equiv_\Gamma C'_{sub}$ *and* $C_{sup} \equiv_\Gamma C'_{sup}$*;*
- $\%(T', C'_{sub}, l_{sub}, u_{sub}), \%(T'', C'_{sup}, l_{sup}, u_{sup}) \in statements(k_i)$*; and*
- $conflicts([l_{sub}, u_{sub}], [l_{sup}, u_{sup}])$

PROOF. This is a straightforward induction on $i$. At $i = 0$, $SpecG(0)$ contains $o$ together with those classes $frontier(i)$ of which $o$ is an immediate member. In addition, $SpecG(0)$ contains edges from each $F \in frontier(0)$ to all classes of which $F$ is a proper subclass. When each edge $(F, C)$ is followed, all statistical statements in $statements(k)$ with a reference class $\Gamma$-equivalent to $F$ are checked against all those with a reference class $\Gamma$-equivalent to $C$, so that the statement holds. For $i > 0$ the statement holds by the induction hypothesis for nodes that are not in $frontier(i)$. For $frontier(i)$, the argument is the same as for $frontier(0)$.$\dashv$

The next step is to show that when sharpeningBySpecificity removes a statement from $statements(k)$, the removal is actually a sharpening by specificity. We begin by restating the classic notion of sharpening by specificity [15] using the terminology of this section.

**Definition 7.** Let $\Delta, \Delta'$ be sets of statistical statements for a given theory $\Gamma$ and target class $Target$. $\Delta$ sharpens by specificity $\Delta'$ iff

$$\begin{aligned}
\Delta' = \Delta - \%(T, C, l, u)& \\
\wedge \exists C_{sub} \in \Delta[C_{sub} \subseteq C \wedge conflicts(Support(C, \Delta), [l, u])& \\
\wedge \neg \exists C_{sub_{sub}} \in \Delta[C_{sub_{sub}} \subseteq C_{sub}& \\
\wedge conflicts(Support(C_{sub_{sub}}, \Delta), Support(C_{sub}, \Delta)]]&
\end{aligned}$$

**Theorem 3.** *In* sharpenBySpecificity*,*

$$\forall k \geq 0[statements(k+1) \textit{ sharpens by specificity } statements(k)]$$

PROOF. This statement is again proved by induction on the iteration number $i$, where we assume that sharpening is performed by a call sharpen($N$,$C$). For $i = 1$, it is immediate that $N \subseteq C$, and

$$conflicts(Support(N, statements(k_i), Support(C, statements(k_i))).$$

Also at $i = 1$, by construction of $SpecG$, there is no $C' \in Refs$ such that $C' \subset N$, so

$$\neg \exists C'[C' \subseteq N \wedge conflicts(Support(C', statements(k_i)), Support(N, statements(k_i)))]$$

For $i > 0$, it is again immediate that $N \subseteq C$, and

$$conflicts(Support(N, statements(k_i), Support(C, statements(k_i))).$$

To show

$$\neg\exists C'[C' \subseteq N \wedge \mathit{conflicts}(Support(C', statements(k_i)), Support(N, statements(k_i)))]$$

first note that by the induction hypothesis this statement was true at the beginning of iteration $i$, for some $k' \leq k_i$. However, by Lemma 1,

$$Support(C', statements(k')) = Support(C', statements(k_i))$$

and

$$Support(N, statements(k')) = Support(N, statements(k_i))$$

so that the statement holds.⊣

Theorem 3 shows that sharpenBySpecificity is correct in the sense that each removal from *Input* is an actual sharpening; the following theorem shows that all possible sharpenings are performed.

**Theorem 4.** *Let $k_{fin}$ be the final value of $k$ after applying* sharpeningBySpecificity *to a set* Input *of statistical statements over a theory* $\Gamma$. *Then*

$$\neg\exists\Delta[\Delta \subset statements(k_{fin}) \wedge \Delta \textit{ sharpens by specificity } statements(k_{fin})]$$

PROOF. The proof follows almost immediately from Lemma 2, which effectively states that for an edge $(C_1, C_2)$ in $SpecG(i)$ there is no conflict between statements for reference classes equivalent $C_1$ and those equivalent to $C_2$. First, note that at the end of the final iteration $i_{fin}$ of shapeningBySpecificity, $SpecG(i) = SpecG$. Since by construction $SpecG$ fully encapsulates the specificity and $\Gamma$-equivalence relations among reference classes, the theorem holds.⊣

Next, we strengthen Theorem 4 to show that the result of sharpeningBySpecificity is unique in the following sense.

**Theorem 5.** *Let* Output *be the set of statements produced by applying* sharpeningBySpecificity *to a set* Input *of statistical statements for a given target* Target *over a theory* $\Gamma$. *Furthermore, suppose* Input *does not contain any statements with conflicting intervals for the same reference class. Then*

$$\forall\Delta \subset Input[(\Delta \textit{ sharpens by specificity } Input) \Rightarrow (Output \subseteq \Delta)]$$

PROOF. Consider a statement $S = \%(T, C, l, u)$ contained in $Output - \Delta$. Then $S$ must have been removed using a reference class $C_{sub}$ with a conflicting statistical statement. There are two cases. First, note that if $C_{sub} \equiv_\Gamma C$ a statement for $C_{sub}$ cannot sharpen by specificity $C$ (cf. Definition 7). Therefore, $C_{sub}$ must be a proper subset of $C$. However, in this case it is straightforward application of Lemma 2 that sharpeningBySpecificity will remove $S$ leading to a contradiction.⊣

### *3.5. Correctness of* EPInfer

To prove correctness of EPinfer, we first observe that some components of the algorithm are verified directly by definition. Specifically, getRelevantStatements is verified directly by Definition 1, and closedUnderConflict, *lower*, and *upper* correspond to the strength rule, which is verified by Definitions 5 and 6 and Theorem 2.

Regarding the two rules which manage conflicts among statements ordered by class attributes, SharpenByRichness and SharpenBySpecificy, Section 3.4 provides a proof of correctness for SharpenBySpecificy that also verifies, *mutatis mutandis*, SharpenByRichness.

With the correctness results for these components, and the observation that EPInfer codifies the correct order of execution of each component of steps, we then have the following representation result.

**Theorem 6.** *Given a knowledge base* $\Gamma$ *and statement* $\chi$ *such that* $\chi \equiv o \in Target$, *the algorithm* EPInfer*(Target, o) yields* $[l, u]$ *iff* $\Gamma \vdash_{[l,u]} \chi$.

PROOF. The sequence of sharpensByRichness, sharpensBySpecificity, and closeUnderConflict defines sharpening of evidence [15, p. 219]. The RHS of this theorem is the RHS of the partial proof theorem in [15, p. 220]. By definition of partial proof, $\Gamma$ provides a partial proof of $\chi$ to degree $[l, u]$ if and only if there is a set of potential statistical statements $\Delta \subset \Gamma$ such that the degrees of support provided to $\chi$ by $\Delta$ are $l$ and $u$, and no alternative $\Delta'$ sharpens $\Delta$, and any alternative $\Delta'$ that contains statistical statements in conflict with $\Delta$ is sharpened by $\Delta$. $\dashv$

| | |
|---|---|
| %(stolen,redMountainDomestic,0.0279,0.971) | %(stolen,redMountainImported,0,1) |
| %(stolen,redTouringDomestic,0,0.062) | %(stolen,redTouringImported,0,0.05672) |
| %(stolen,redRacingDomestic,0,0.0558) | %(stolen,redRacingImported,0,0.0638) |
| %(stolen,redTouring,0,0.0454) | %(stolen,redRacing,0,0.0454) |
| %(stolen,redMountain,0.0279,0.0971) | %(stolen,redImported,0,0.0467) |
| %(stolen,redDomestic,0.0123,0.0572) | %(stolen,racingImported,0,0.0582) |
| %(stolen,touringImported,0,0.055) | %(stolen,mountainImported,0,0.1218) |
| %(stolen,racingDomestic,0,0.0574) | %(stolen,touringDomestic,0,0.0453) |
| %(stolen,mountainDomestic,0,0342,0.786) | %(stolen,red,0.0084,0.0476) |
| %(stolen,racing,0,0.0467) | %(stolen,touring,0.0052,0.0407) |
| %(stolen,mountain,0.0352,0.0774) | %(stolen,imported,0.0042,0.049) |
| %(stolen,domestic,0.0217,0.0506) | %(stolen,top,0.0212,0.0455) |

Figure 4: Statistical Statements about Stolen Bicycles

**Example 5.** We illustrate the above algorithms using a set of statistical statements from Section 9.8 of [15]) about the probabilities that various types of bicycles will be stolen. Figure 3.5 lists some of these statements, and as background knowledge we assume a common-sense inclusion relation: e.g., redRacingImported is a subclass of redImported, redRacing, and racingImported, and so on. Let $b_1$ be a red, imported, racing bicycle. In this case, getRelevantStatements of Figure 3.2 is called with input arguments *Target* = *stolen* and *Object* = $b_1$, and sets *RelevantStatements* to the set of all statistical statements whose target is Γ-equivalent to stolen and whose reference class is equivalent to some class of which $b_1$ is a member. *RelevantStatements* thus contains the following statements.

%(stolen,redRacingImported,0,0.0638)
%(stolen,redRacing,0,0.0454)
%(stolen,redImported,0,0.0467)
%(stolen,racingImported,0,0.0582)
%(stolen,red,0.0084,0.0476)
%(stolen,racing,0,00467)
%(stolen,imported,0.0042,0.0490)
%(stolen,top,0.0212,0.0455)

We note that the background knowledge for example does not contain any relations between marginal and joint distributions, so that *sharpenByRichness* has no effect. Accordingly, we next execute sharpenBySpecificity. The first iteration begins with $frontier(1)$ equal to redRacingImported. Since this class is a proper subclass of all other classes, it is checked for conflict against each of them, but since it does not conflict with any other classes, no sharpening is performed. In iteration 2, $frontier(2)$ is the set

{redRacing,redImported,racingImported}

We assume that redRacing is the first class in $frontier(2)$ to be chosen. Its support conflicts with the statistical statement with reference classes red and top. Since traversal of the edges for redRacing has sharpened away two statistical statements, when the next class in $frontier(1)$ is chosen, $k = 2$ and $statements(2)$ is the set

%(stolen,redRacingImported,0,0.0638)
%(stolen,redRacing,0,0.0454)
%(stolen,redImported,0,0.0467)
%(stolen,racingImported,0,0.0582)
%(stolen,racing,0,00467)
%(stolen,imported,0.0042,0.0490)

Assuming that redImported is chosen next, it conflicts with imported; so that at the end of the traversal of the edges in $SpecG$ arising from this node, $k = 3$ and $statements(3)$ is

%(stolen,redRacingImported,0,0.0638)
%(stolen,redRacing,0,0.0454)
%(stolen,redImported,0,0.0467)
%(stolen,racingImported,0,0.0582)
%(stolen,racing,0,00467)

Finally, racingImported is chosen, but produces no sharpening. Next, in iteration 3, $frontier(3)$ contains only racing whose support does not sharpen any other statement. After iteration 3, the algorithm terminates, setting $Output$ to the set of 5 statistical statements above. Applying closureUnderConflict and the principle of strength, EPInfer terminates with the interval $[0, 0.0454]$.

## 4. A KRR System for Evidential Probability

In order to understand the knowledge representation and reasoning capabilities that are needed to support EPInfer we consider the example about pigs from Section 2. First, we note that DLs handily capture subclass relations; they thus easily represent that generations 1-6 are subclasses of the class of all generations. or that pens 1-12 are subclasses of all pens. DLs also handle well information about specific pigs: that $o_{333}$ is a member of the class of white pigs, or that $o_{18}$ belongs to the class of pigs living in $N_{12}$ or in generation 1.

At the same time, expressing other types of information with DLs can be inconvenient or impossible. For instance, it is more concise to specify 1500 instances of pigs through a single rule than through 1500 facts. In Prolog syntax, such a rule would be

```
pig(o(X)):- between(1,1500,X).
```

where between/3 backtracks through all numbers in the range 1-1500. Furthermore, DLs cannot easily represent statistical statements, even when they are facts, such as

```
%(white,gen(2),0.611,0.719).
```

To take a more complex case, rules are especially helpful to determine statistical statemensts for the experiment of choosing a pen at random, and then choosing a pig from that pen. Such an experiment may be used to sample information about a target representing genotypic alleles or the phenotype white. These statements can be written as a rule that constructs the weighted average of all pigs in each generation having the target property. In Prolog this is represented as:

```
%(Target,pen_selected_pig,L,U):-
    %(Target,gen(0),L_0,U_0), %(Target,gen(1),L_1,U_1),
    %(Target,gen(2),L_2,U_2), %(Target,gen(3),L_3,U_3),
    %(Target,gen(4),L_4,U_4),
    pigsInGen(gen(0),N_0), pigsInGen(gen(1),N_1),
    pigsInGen(gen(2),N_2), pigsInGen(gen(3),N_3),
    pigsInGen(gen(4),N_4),
    Tot is N_0 + N_1 + N_2 + N_3 + N_4,
    L is N_0/Tot*L_0 + N_1/Tot*L_1 + + N_2/Tot*L_2 + N_3/Tot*L_3 + N_4/Tot*L_4,
    U is N_0/Tot*U_0 + N_1/Tot*U_1 + N_2/Tot*U_2 + N_3/Tot*U_3 + N_4/Tot*U_4.
```

Rules can therefore be convenient to represent statistical statements, especially for complex reasoning as with the example about pigs. However, rules can also be useful to support temporal or spatial information. This is because supporting such reasoning requires a much stronger deduction mechanism than most DLs support (the deduction must be extended to concrete domains, cf. [6]). Furthermore, since general arithmetic is undecidable it is not supported by DLs. And last, the important problem of reachability in a general binary relation cannot even be represented using first-order logic. All of these limitations can be overcome by extending DLs with a fixed-point logic such as is provided by logic programming. We now discuss a KRR formalism for representing and reasoning with Evidential Probabilities that combines the advantages of DLs as a background knowledge server, with the advantages of rules for representing both statistical statements and the EPInfer algorithm itself.

### *4.1. Extending MKNF Knowledge Bases with Evidential Probability*

As noted in Section 3.1 given a particular target $Target$ and object $o$, EPInfer requires as input

1. the set of classes, $\mathcal{C}$ to which $o$ belongs;

2. all statistical statements for $\Gamma \vdash (T' \equiv Target)$ with reference class $C \in \mathcal{C}$; and
3. the inclusion and richness relations among the relevant classes in $\mathcal{C}$.

Also as noted, while DLs are suitable for representing 1) and 3), rules are more suitable for representing 2). In addition, if Evidential Probability is to be represented within the KRR system itself, rather than as an addition, the KRR system must support enough procedurality to implement EPInfer. At the same time, we want the KRR system to have a clear semantics, to be decidable, and to be scalable depending on the type of reasoning it supports.

A system that combines the first-order reasoning capabilities of a DL (or other decidable fragment of first-order logic) along with the procedurality of rules would fulfill these desiderata. There is, in fact, an extensive literature concerning how to combine DLs with rules, and the approach we adopt here is of a hybrid knowledge base using Minimal Knowledge and Negation as Failure [19] (MKNF). The general advantages of MKNF have been well-studied. MKNF allows knowledge about objects to be fully inter-definable between rules, and a DL that is taken as a parameter of the formalism. Using this parameterized DL, MKNF is defined using monotonic fixed-point operators that compute in each iteration step, besides the usual immediate consequences from rules, the set of all atoms derivable from the DL whose derivations about objects (ABox) is augmented with knowledge already proven by rules and previous DL derivations.. In other words, the MKNF formalism allows knowledge about objects as used by EPInfer to be inter-definable between rules and a DL.

As the MKNF semantics is based on a fixed point, MKNF knowledge bases can be defined using different fixed-point semantics for negation as failure: notably the stable model semantics and the well-founded semantics (WFS). The relationship between the two semantics has been fully explored. Of the two, WFS is weaker (in the sense that it is more skeptical), but has the clear advantage that its low polynomial complexity is more suitable for applications with large amounts of data than is the the NP-complete complexity of the stable model semantics. MKNF under WFS (termed $\mathit{MKNF}_{WFS}$) has well defined properties [9]. The least fixpoint of the $\mathit{MKNF}_{WFS}$ operators coincides with the original WFS [23] if the ontology is empty, and coincides with the semantics of the ontology if there are no rules; in addition, when dealing with polynomial DLs (e.g. $\mathcal{EL}^{+}$[2]), $\mathit{MKNF}_{WFS}$ retains a polynomial data complexity. Furthermore, there exists an algorithm, called **SLG($\mathcal{O}$)**, for querying a $\mathit{MKNF}_{WFS}$ knowledge base [1]. **SLG($\mathcal{O}$)** is sound and ideally complete for $\mathit{MKNF}_{WFS}$. Furthermore, when **SLG($\mathcal{O}$)** is extended to use a technique called *term-depth abstraction* [21] for subgoals it will terminate whenever the $\mathit{MKNF}_{WFS}$ model is finite. We summarize the results of the literature as follows.

**Theorem 7.** *[9, 1] Let $\mathcal{H}$ be a* $\mathrm{MKNF}_{WFS}$ *knowledge base, such that the least fixed-point model $\mathcal{M}$ of $\mathcal{H}$ is finite. Let $Q$ be a DL-safe conjunctive query* [4] *to $\mathcal{H}$. Then an* **SLG($\mathcal{O}$)** *evaluation of $Q$ to $\mathcal{H}$ that uses term-depth abstraction for calls will terminate. Furthermore, if the underlying DL, $\mathcal{O}$ has polynomial complexity, then the*

[4] A DL-safe conjunctive query is a conjunction of predicates with variables where queries have to be ground before being processed in the ontology

**SLG($\mathcal{O}$)***evaluation of $Q$ against $\mathcal{H}$ will also have polynomial complexity.*

The foregoing statement concerns only evaluation of a combination of rules and DL under *MKNF*$_{WFS}$ semantics, and it needs to be extended to include the EPInfer algorithm. First, note that querying the set of relevant statistical statementss for an object $o$, the set $\mathcal{C}$ to which $o$ belongs, and the richness and specificity relations among classes in $\mathcal{C}$ can be done in a straightforward manner using rules and DL-safe queries. Next, it is straightforward that if the set $Input$ is finite, then EPinfer will also terminate.

More specifically, note that the complexity of the routine SharpenBySpecificity (and SharpenByRichness) is determined as follows. The cost of determining the maximal path length between the object $o$ and all classes, is $\mathcal{O}(|Classes|^2)$, where $|Classes|$ is the number of nodes in the graph. This can be seen by considering first that $SpecG$ is acyclic, and all edge costs have positive weight. Accordingly, finding maximal paths in $SpecG$ is equivalent to finding minimal paths in a transformed graph, all of whose edges have negative weight, and a standard algorithm such as Dijkstra's [4] can be used. The rest of the algorithm traverses the graph, along with classes in $Refs$ that are equivalent to classes in the quotient set $Classes$, and is $\mathcal{O}(|Refs|^2)$ where $|Refs|$ is the size of the set $Refs$. Thus the worst-case complexity of sharpenByRichness and sharpenBySpecificity is quadratic in the set of relevant statements:

$$\mathcal{O}(|RelevantStatements|^2)$$

Furthermore, note that the complexity of determining closure under conflict (Definition 5) is dominated by the time to sort the sets of intervals, and so is

$$\mathcal{O}(|RelevantStatements| \times \log(RelevantStatements))$$

so that EPInfer as a whole has abstract complexity $\mathcal{O}(|RelevantStatements|^2)$. These observations, together with Theorem 7 imply the following theorem.

**Theorem 8.** *Let $\mathcal{H}$ be a* MKNF$_{WFS}$ *knowledge base, such that the model $\mathcal{M}$ of $\mathcal{H}$ is finite. Let $Q$ be a query to* EPInfer *to calculate the evidential probability for a given object $o$ and target $I$ that makes use of* **SLG($\mathcal{O}$)**. *Then $Q$ will terminate, and if the underlying DL, $\mathcal{O}$ has polynomial complexity, then $Q$ will also have polynomial complexity.*

## 5. Discussion

A preliminary prototype of EPInfer has been implemented in XSB Prolog. This prototype requires background knowledge to be stated directly as Prolog facts. Integration with a hybrid MKNF reasoner is underway, based on the query-driven MKNF implementation described in [5]. This MKNF implementation can work on two different types of ontologies: Type 0 ontologies, which are low complexity and resemble frame systems; and Type 1 ontologies, which are based on the expressive description logic $\mathcal{ALCQ}$. The MKNF system in its turn is based on XSB Prolog's ontology management system, the *Coherent Description Framework*, which is used commercially for semantic web and e-commerce applications. All of these systems are open-source, as will be the extension to evidential probability. As a result, full integration of EPInfer with the CDF-based MKNF reasoner will allow the use of Evidential Probability to be evaluated for research and commercial applications.

We anticipate that our implementation of EP will help advance the theory of Evidential Probability, too. Part of the controversy over Evidential Probability has been that the theory changed over time but those changes were not easy to track because of its abstract presentation. Fairly or not, that history has led some to conclude that at the bottom of Evidential Probability there is neither a logic nor a solution to reference class reasoning. Having a sound, strongly complete, and decidable version of EP should addresses the 'no logic' charge. But more importantly, our concrete implementation should also help address some of the philosophical issues raised by reference class reasoning in general, and the EP solution in particular. Evidential probability was originally conceived to be a type of one-shot inference, since any change to a knowledge base or to the target statement necessitated recomputing a new probability. But, because of the efficiency ofEPInfer, this may be less of an issue. MKNF provides resources for keeping track of iterated changes to an ontology, and therefore provides the expressive capacity to chain EP inferences. Another objection to Evidential Probability is that several of steps in the inference procedure appear difficult to justify. But, again, the efficiency of EPInfer may ameliorate this concern by allowing the option of varying parameters.





## 6. References

[1] J.J. Alferes, M. Knorr, and T. Swift. Queries to hybrid MKNF knowledge bases through oracular tabling. In *International Semantic Web Conference*, 2009.

[2] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *IJCAI'05: 19th Int. Joint Conf. on Artificial Intelligence*, pages 364–369. Morgan Kaufmann, 2005.

[3] F. Baader, D. Calvenese, D. McGuiness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[4] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[5] S. Gomes, J.J. Alferes, and T. Swift. Implementing query answering for hybrid MKNF knowledge bases. In *Theory and Practice of Logic Programming*, 2011. To appear.

[6] V. Haarslev, C. Lutz, and R. Moller. A description logic with concrete domains and a role-forming predicate operator. *Journal of Logic and Computation*, 9(3):351–384, 1999.

[7] Rolf Haenni, Jan-Willem Romeyn, Gregory Wheeler, and Jon Williamson. *Probabilistic Logic and Probabilistic Networks*. The Synthese Library. Springer, Dordrecht, 2010.

[8] William Harper and Gregory Wheeler, editors. *Probability and Inference: Essays In Honor of Henry E. Kyburg, Jr.* King's College Publications, London, 2007.

[9] Matthias Knorr, José Júlio Alferes, and Pascal Hitzler. A well-founded semantics for MKNF knowledge bases. In Diego Calvanese, Enrico Franconi, Volker Haarslev, Domeni co Lembo, Boris Motik, Anni-Yasmin Turhan, and Sergio Tessaris, editors, *Description Logics 2007*, pages 417–425. CEUR–WS, 2007.

[10] Henry E. Kyburg, Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT, 1961.

[11] Henry E. Kyburg, Jr. *The Logical Foundations of Statistical Inference*. D. Reidel, Dordrecht, 1974.

[12] Henry E. Kyburg, Jr. The reference class. *Philosophy of Science*, 50:374–397, 1983.

[13] Henry E. Kyburg, Jr. Bayesian inference with evidential probability. In William Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, pages 281–296. King's College, London, 2007.

[14] Henry E. Kyburg, Jr. and Bulent Murtezaoglu. A modification to evidential probability. In *Uncertainty in Artificial Intelligence: Proceedings (UAI-1991)*, pages 228–231, 1991.

[15] Henry E. Kyburg, Jr. and Choh Man Teng. *Uncertain Inference*. Cambridge University Press, Cambridge, 2001.

[16] J. W. Lloyd. *Foundations of Logic Programming*. Springer-Verlag, Berlin Germany, 1984.

[17] Ron Loui. Computing reference classes. In *Uncertainty in Artificial Intelligence: Proceedings (UAI-1986)*, pages 183–188, 1986.

[18] Stefan Marklund, James Lijas, Heriberto Rodriguez-Martinez, Lars Rönnstrand, Keiko Funa, Maria Moller, Dirk Lange, Inger Edfors-Lilja, and Leif Andersson. Molecular basis for the dominant white phenotype in the domestic pig. *Genome Research*, 8:826–833, 1998.

[19] B. Motik and R. Rosati. A faithful integration of description logics with logic programming. In *20th Int. Joint Conf on Artificial Intelligence (IJCAI)*, pages 477–482, Hyderabad, India, January 6–12 2007. AAAI Press.

[20] Hans Reichenbach. *The Theory of Probability*. University of California Press, Berkeley, 1949.

[21] H. Tamaki and T. Sato. OLDT resolution with tabulation. In *International Conference on Logic Programming*, pages 84–98. Springer-Verlag, 1986.

[22] Choh Man Teng. Conflict and consistency. In William L. Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* King's College Publications, London, 2007.

[23] A. van Gelder, K.A. Ross, and J.S. Schlipf. Unfounded sets and well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, 1991.

[24] Gregory Wheeler. A resource bounded default logic. In James Delgrande and Torsten Schaub, editors, *NMR 2004*, pages 416–422, 2004.

[25] Gregory Wheeler and Carlos Damásio. An implementation of statistical default logic. In Josés Alferes and João Leite, editors, *Logics in Artificial Intelligence: Proceedings of the 9th European Conference (JELIA 2004)*, Lecture Notes in Artificial Intelligence, pages 121–133, Berlin, 2004. Springer-Verlag.

[26] Julian Wiseman. *A History of the British Pig*. Ebenezer Baylis & Sons, Worcester, 1986.