

REPORT DOCUMENTATION PAGE

Form Approved  
OPM No. 0704-0188

AD-A250 605



to average 1 hour per response, including the time for reviewing instructions, searching existing data sources gathering and  
in. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions  
strate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to  
and Budget, Washington, DC 20503.

REPORT DATE: 1989  
3. REPORT TYPE AND DATES COVERED: Unknown

4. TITLE AND SUBTITLE: A Statistical Solution to the Qualificaiton Problem and How it Also Solves the Frame Problem  
5. FUNDING NUMBERS: DAAB10-86-C-0567

6. AUTHOR(S): Josh Tenenberg/Jay Weber

DTIC ELECTE  
MAY 27 1992  
S D C

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES): University of Rochester, Department of Computer Science, Rochester, NY 14627

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES): U.S. Army CECOM Signals Warfare Directorate, Vint Hill Farms Station, Warrenton, VA 22186-5100

10. SPONSORING/MONITORING AGENCY REPORT NUMBER: 92-TRF-0013

1. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT: Statement A; Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words) A close relative of the frame problem is the qualification problem. This problem concerns what preconditions an agent considers sufficient for an action to achieve an effect. In general, the consideration of an ideal list of sufficient preconditions will be impossible or impractical, and as such, an agent reasoning about action success will be obliged to do so from incomplete evidence. Standard approaches to this problem have been to use non-monotonic or consistency based logical methods that assume those sufficient preconditions which are usually true are true by default. However, these approaches all suffer from a classic problem of default logic called the lottery paradox, as a result of the coarse way that defaults capture statistical properties of the domain. In contract, we present a novel method for solving the qualification problem using standard techniques for statistical inference. We take it that the gent qcquires statistics about the proportion of success of its actions, conditioned upon the existence of certain preconditions which hold just prior to the action.

1. SUBJECT TERMS: Artificial Intelligence, Data Fusion, Frame Problem, Qualification Problem, Algorithms Causal Reasoning

15. NUMBER OF PAGES: 14

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT: UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE: UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT: UNCLASSIFIED

20. LIMITATION OF ABSTRACT: UL

# A Statistical Solution to the Qualification Problem

and How it Also Solves the Frame Problem

Josh D. Tenenberg\*  
josh@cs.rochester.edu

Jay C. Weber†  
jay@cs.rochester.edu

Computer Science Department  
University of Rochester  
Rochester, New York U.S.A., 14627

March 1989

Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or
A-1	Special

## Abstract

A close relative of the frame problem is the *qualification* problem. This problem concerns what preconditions an agent considers sufficient for an action to achieve an effect. In general, the consideration of an ideal list of sufficient preconditions will be impossible or impractical, and as such, an agent reasoning about action success will be obliged to do so from incomplete evidence. Standard approaches to this problem have been to use non-monotonic or consistency based logical methods that assume those sufficient preconditions which are usually true are true by default. However, these approaches all suffer from a classic problem of default logics called the *lottery paradox*, as a result of the coarse way that defaults capture statistical properties of the domain.

In contrast, we present a novel method for solving the qualification problem using standard techniques for statistical inference. We take it that the agent acquires statistics about the proportion of success of its actions, conditioned upon the existence of certain preconditions which hold just prior to the action. From such statistics, the agent derives degrees of belief in the success of *particular* actions. Choosing among a set of applicable statistics is the familiar

\*supported in part by the Air Force Systems Command, Rome Air Development Center, Griffiss Air Force Base, New York 13441-5700, and the Air Force Office of Scientific Research, Bolling AFB, DC 20332 under Contract Number F30602-85-C-0008 which supports the Northeast Artificial Intelligence Consortium (NAIC).

†supported in part by U.S. Army contract no. DAAB10-87-K-022.

92-13686



92 0 22 0 0 0

*reference class* problem, which is the subject of much work on statistical inference. We demonstrate that each qualification serves to define a more specific reference class, and that although no completely specified reference class will be obtainable, more general classes are nonetheless useful because they capture statistical generalizations about the omitted qualifications. Since this approach quantitatively captures the statistical properties of the domain, it does not suffer from the lottery paradox as do default approaches. Lastly, we demonstrate that this approach also solves the frame (persistence) problem.

## 1 Introduction

McCarthy [1977] introduced the well-known example of the qualification problem, called the “potato in the tailpipe” scenario. Suppose you get in your car, and turn the key. You might expect the car to start, despite not knowing whether the battery is still charged, whether the ignition system is intact, whether there is a potato in the tailpipe, or whether any of a potentially infinite set of *qualifications* are known to hold. It should be clear that the problem of specifying sufficient preconditions for the success of an action is not specific to car-startings, but arises in reasoning about *any* action. The qualification problem asks formally how a reasoner can make a causal prediction based on a *reasonable* body of evidence.

There are actually several problems here, all of which can broadly be thought of as some aspect of the qualification problem. Firstly, as pointed out in [Elgot-Drapkin *et al.*, 1987], it may be *impossible* to write down all of the qualifications that are sufficient to guarantee action success — the complexity of the domain might be beyond human ability to completely formalize. Even were this not the case, some domains require unmanageably large sets of qualifications for every action/effect pair. Even if there were a tractable set of qualifications for an action/effect pair, there remains the problem of guaranteeing that all of the qualifications are satisfied in a given situation. For instance, when attempting to start a car, people don’t have guarantees that the car has gas, that the battery is charged, that the fuel pump is intact — all important features of a working car. Thus, in any representation of the car-starting problem, it is equally unlikely (and, in fact, suspect) that one will be able to guarantee the truth of the qualifications.

Despite these problems, automated agents in complex worlds will still be obliged to choose an appropriate course of action. Agents will thus be required to reason under uncertainty. Previous work on the qualification problem [McCarthy, 1980, McCarthy, 1984, Lifschitz, 1987, Ginsberg and Smith, 1987] has used some form of non-monotonic logical inference to manage this uncertainty by assuming that those sufficient preconditions which are usually true (a.k.a. qualifications) are true by default. As we shall see in Section 2, these approaches all suffer from a classic problem

of default logics called the *lottery paradox*, as a result of the coarse way that defaults capture statistical properties of the domain.

In this paper, we manage the uncertainty heralded by the qualification problem using the principles of *statistical inference* [Reichenbach, 1949, Kyburg, 1974, Pollock, 1984]. We start by reviewing the strengths and weaknesses of the default logic approach to the qualification problem in Section 2. In Section 3, we present an informal view of our approach. Section 4 provides a brief review of the statistical and probabilistic formalisms that we employ. In Section 5, we apply these formalisms to the task of causal reasoning, and demonstrate how we solve the qualification problem, and Section 6 demonstrates how we solve the frame problem as a special case. We close by discussing work on some additional issues, and summarizing our results.

## 2 Default Rules and the Lottery Paradox

Previous work on the the qualification problem [McCarthy, 1980, McCarthy, 1977, Ginsberg and Smith, 1987] has tried to capture performance issues, e.g. by writing the axioms so that a mixture of forward and backward chaining can be used, but it is not clear if abstract measures such as numbers of facts actually dominate the computational properties at the unspecified algorithmic level. It is more appropriate to view logical solutions to the qualification problem as representational rather than computational, especially when semantic notions such as possible worlds or model elimination are used to capture defaults. For this reason, we will not discuss the particulars of these solutions, but instead present a more general view of the use of default logic to solve the qualification problem.

We add a defeasible version of the implication operator to first-order logic, so that we may write the sentence  $\phi \rightsquigarrow \psi$  to mean “the sentence  $\phi$  defeasibly implies  $\psi$ ”. The formal properties of such a task have been investigated in numerous places [Loui, 1987, Nute, 1986, Reiter, 1980]; here we will rely on the following informal definition: if  $\phi \rightsquigarrow \psi$  is a default rule,  $\phi$  is considered true by the reasoner, and  $\psi$  is consistent with everything considered true, then  $\psi$  is also considered true. This operator is used to solve the qualification problem by writing causal rules which capture reasonable conclusions from appropriate evidence. e.g. for car starting:

$$\neg\text{running}(t) \wedge \text{turn-key}(t) \rightsquigarrow \text{running}(t+1) \quad (1)$$

$$\neg\text{running}(t) \wedge \text{turn-key}(t) \wedge \text{dead-battery}(t) \rightsquigarrow \neg\text{running}(t+1) \quad (2)$$

In words, if the car is not running and the key is turned, then it is reasonable to assume the car will be running at the next moment; however, if the battery is dead, then it is reasonable to assume that the car will *not* be running at the next moment. Since the conclusions of these two rules conflict, there are situations where the prediction about the car starting depends on the order that the rules are applied. Several authors

[Etherington, 1987, Touretzky, 1984] have advocated choosing the more *specific* rule, e.g. if both antecedents above are known to be true, then the second rule will be used since it involves properly more information about the time in question. Therefore, if both  $\neg\text{running}(0)$  and  $\text{turn-key}(0)$  are known but  $\text{dead-battery}(0)$  is not, the reasoner will conclude  $\text{running}(1)$  since the second rule is not applicable. However, if  $\text{dead-battery}(0)$  becomes known, then both rules are applicable and the second is chosen by specificity, concluding  $\neg\text{running}(1)$ .

Note that default rule (1) *implicitly* captures that the car's battery is seldom dead. Otherwise, if the battery often died, then (1) and (2) could not both be reasonable default rules. Thus a default rule makes implicit default assertions about the sufficient preconditions that do not appear in the antecedent. This is the essence of all the default logic solutions to the qualification problem. An analogue of this feature will also be crucial to our statistical solution to the qualification problem, discussed in Section 3.

A classic problem of default logic approaches is the *lottery paradox* [Kyburg, 1970]. Imagine reasoning about a lottery where one of a large number of people will win a prize. The chance that any particular person will win is certainly very small, making it a reasonable default that the person will not win. Since this can be uniformly applied to all people involved, the conjunction of all the defaults implies that no person will win. This contradicts the definition of the lottery, which says that there will be a winner. The reason for this contradiction is that defaults to not lose force when composed, yet the statistical facts they reflect do.

The lottery paradox appears in the use of defaults for causal reasoning. The above car-starting default rules allow the reasoner to assume, for a particular car-starting instance, that the car will start and therefore the battery will not be dead. This seems perfectly reasonable. However, when taken together, these particular assumptions imply that since the reasoner has no future counter-evidence, the car will *always* start and the battery will *never* be dead. This would indicate that battery cables or preventive replacement of batteries are superfluous, and other conclusions we know to be wrong in practice. To correct this problem, it has been suggested that facts in the evidence set must be indefeasibly true; this would prevent far-reaching predictions, but also prevent reasonable predictions even two moments ahead. Shoham [1988] suggested attaching a number to each default rule (actually the projection of a "potential history") that bounds the number of times the rule can be applied in a chain. This solution is too *ad hoc* for examples such as the car battery, where there is no most reasonable cutoff period (although things do tend to break as soon as the warranty expires).

### 3 Our Approach: Informal View

A default rule works well when its antecedent implies that its consequent is almost always or almost never true, because it assigns one of two truth values, true or false. Default rules do not handle the middle ground, where the antecedent provides information about the consequent, but not to the point of declaring it true or false. For example, suppose the car is known to have difficulty starting in cold weather. Certainly cold(0) is an important fact to consider when predicting running(1), but to say that the car will start anyway or definitely not start is inappropriate.

Therefore, the first step of our solution is to generalize the default approach to derive a *degree of belief* in the consequent (effect) given the knowledge that the antecedent (precondition list) is true. This degree can be based on actual domain observations about the relative frequency of the effect given the preconditions, e.g. the car starts 70% of the time in cold weather. This allows for much more expressive relationships between preconditions and effects. It also captures the “non-monotonic flavor” of the qualification problem. For example, the car may start 95% of the time that the key is turned, only 70% of the time the key is turned and the weather is cold, but back up to 90% of the time the key is turned, the weather is cold, and the anti-freeze is extra strength.

There may be multiple relative frequencies applicable to the same prediction. In the percentages above, if the reasoner ignores the fact that the weather is cold, it may derive a 95% degree of belief instead of the 70% degree of belief. This is the *reference class problem* [Kyburg, 1983b] that has concerned philosophers and statisticians for several decades. A reference class is simply a body of knowledge a reasoner uses to make a prediction about a degree of belief, analogous to the antecedent of a default rule. We will say more about this later, but the most commonly accepted principle for preferring one reference class to another is to choose the most specific of the two. This is the analogue of the rule preference from default logic described in Section 1; in fact, specificity was discussed for degrees of belief before default logics were even invented.

This approach does not suffer from the lottery paradox as does the default logic approach, since degrees of belief do not need to be “subsidized” in order to be as strong as true or false. For example, given the belief that the car starts 95% of the time, the reasoner derives that there is a  $95\% \cdot 95\% \approx 90\%$  chance that the car will start twice<sup>1</sup>, a 85% chance for three times, etc. This captures the intuition that the more times an action is attempted, the more likely it will fail during one of those attempts. This commonsense fact is totally lost in the default logic approach.

As with default logic, the essence of how our approach solves the qualification problem comes from what a rule says about those preconditions that do *not* actually

---

<sup>1</sup> Assuming that the car starting trials are independent. If they are not, information about how they are dependent can be used to combine the individual chances.

appear in the rule. For example, if it is believed that the car starts 95% of the time, and it is believed that a dead battery prevents the car from starting (all of the time), then it must also be believed that the battery is dead less than  $100\% - 95\% = 5\%$  of the time. In fact, all the situations that would prevent the car from starting are believed to total exactly 5%. Thus our causal rules capture *statistical generalizations* about the combined effect of an action's qualifications.

## 4 Statistical Inference

Having informally described the features of our approach, we now turn to a more detailed exposition. We have adopted Kyburg's representation, and provide a brief description here. More thorough descriptions can be found in [Kyburg, 1983a, Kyburg, 1974]. Kyburg makes an important distinction (which is often ignored in AI work [Dean and Kanazawa, 1988]) between a *statistic*, which captures the relative frequency between sets of objects, and a *probability* that some object is a member of a set with respect to some knowledge about that object. These two concepts are often confused because they both assign a value to similar syntactic objects from the interval  $[0, 1]$  within a field of numbers. However, it is important to bear in mind the difference; a statistic assigns a relative frequency to a pair of sets, and a probability assigns a degree of belief to a pair of ground formulas (no free variables).

A statistic is written using the “%” predicate, e.g. “%(fliers | birds) = .95” means that 95% of the objects in the set of birds also belong to the set of fliers (we use the conditional bar (|) instead of comma (,) for this predicate to keep the order of arguments clear). Actually, Kyburg uses a more general form of statistical assertion of the form:

$$\%(z | y) \in [p, q]$$

which is taken to mean “the proportion of  $y$ 's that are  $z$ 's is in the interval from  $p$  to  $q$ ”, where  $p \leq q$ . Statistical assertions of this form can be used in the obvious way to say that the proportion of birds that fly is between .9 and .95. Two special cases of this form are important: when  $p = q$  the assertion is equivalent to asserting a point value, and when  $p = 0$  and  $q = 1$  the assertion is a tautology and therefore conveys no information.

A set of these statistical assertions, plus other sentences in first-order logic (with axiomatic set theory, identity, and arithmetic operations) comprise an agent's *rational corpus*. The elements of this corpus must be “practically certain,” that is, believed with sufficient strength so as to be treated as truths<sup>2</sup>. This corpus can be used to derive a probability for a ground sentence that is equivalent to asking if an object is a member of a set, e.g. we might need a degree of belief for the sentence  $\text{Tweety} \in$

---

<sup>2</sup>There are additional syntactic constraints on rational corpi, which will not be discussed further here, but can be found in [Kyburg, 1983a].

**fliers.** Formally, if  $\phi$  is a sentence and  $KB$  is a rational corpus,  $Prob(\phi | KB) = [p, q]$  iff there exist terms  $x, y, z$  such that:

1. " $\phi \equiv x \in z$ " is a sentence in  $KB$ ,
2. " $x \in y$ " is a sentence in  $KB$ ,
3. " $\%(z | y) \in [p, q]$ " is a sentence in  $KB$ , and
4.  $x$  is a random element of  $y$  w.r.t  $z$  relative to  $KB$ .

For example, suppose an agent's  $KB$  contains only the logical consequences of the following facts:

$$\%(fliers | birds) \in [.90, .95]$$

$$Tweety \in birds$$

then  $Prob(Tweety \in fliers | KB) = [.90, .95]$  since condition 1 is trivially satisfied, binding  $z$  to **fliers**, condition 2 binds  $y$  to **birds**, condition 3 defines the interval result, and condition 4 is true because *as far as the agent knows*,  $x$  could be any element of  $y$ . The set  $y$  is what we have been calling the "reference class". As remarked earlier, the choice of reference class can change the resulting degree of belief.

It is interesting to note that condition 4 above subtly imposes the principle of specificity, i.e. a statistic with a more specific reference class is preferred when assigning probabilities. To see how, consider adding the following sentences to the  $KB$ :

$$\%(fliers | penguins) \in [0, .005]$$

$$Tweety \in penguins$$

$$penguins \subset birds$$

Now our previous derivation of  $[.90, .95]$  no longer holds since condition 4 is violated. **Tweety** is no longer an *epistemically random* element of **birds**, since it is now a belief that **Tweety** is in **penguins**. Condition 4 is satisfied for the reference class **penguins**, so the resulting degree of belief is  $[0, .005]$ .

Kyburg also presents a stronger version of specificity, where the relative vagueness of the intervals is considered in addition to the relative specificity of the reference classes. For example, if instead I told you that **Tweety** was a "speckled nuthatch" you might not have a statistic about flying that has a reference interval more informative than  $[0, 1]$ , and it would be better to just use the statistic conditioned on birdhood. A full discussion of principles of mediation is unnecessary for our present purposes, the following principle will be used: statistics for a reference set  $Y$  will always be preferred to a reference set  $W$  whenever  $Y \subset W$ , and the reference interval of  $Y$  is no weaker than (is contained within) the reference interval of  $Z$ .



## 5 Statistical Inference for Causal Reasoning

Having now outlined the basics of statistics and probabilities, we turn to their application to causal reasoning.

### 5.1 Statistical causal rules

In our discussion about causal default rules, we employed a simple model of time, which we now make more explicit: the time line is divided up into a set of *moments*, which are totally ordered by a successor function (the successor of moment  $t$  is  $t'$ , a.k.a.  $t + 1$ ). Since moments are isomorphic with the integers, we will use integer constants as moment constants.

A *fluent* [McCarthy and Hayes, 1981, Lifschitz, 1987] in our approach is simply a set of moments. For example, *turn-key* is the set of all moments at which the car's key is turned. We represent that a fluent is true of a particular moment through set membership, e.g.  $t \in \text{turn-key}$ . This notation is similar to the common reified approach using literals of the form  $\text{holds}(\text{turn-key}, t)$  [Lifschitz, 1987], but more convenient for statistical assertions. An agent's *domain knowledge* is a set of these set membership assertions.

A *statistical causal theory* is a set of statistical assertions about fluents. In particular, we are interested in statistical assertions about how fluents that contain a moment  $t$  influence whether other fluents contain  $t'$ , e.g.

$$\%(\{t : t' \in \text{running}\} \mid \text{turn-key}) \in [p, q]. \quad (3)$$

That is, the proportion of time points in which the car is running preceded by a time point in which the key was turned is in the interval  $[p, q]$ . As shorthand, we use the notation  $X^*$  for the set  $\{t : t' \in X\}$ . In the discussion to follow, we will use the standard set operators  $\overline{X}$  (complement),  $X \subset Y$  (subset) and  $X \cap Y$  (intersection).

The union of the domain knowledge with the statistical causal theory yields a rational corpus from which we may derive interval degrees of belief that moments belong to fluents. For example, suppose the KB contains only the logical consequences of the following sentences:

$$\begin{aligned} \%(\text{running}^* \mid \text{turn-key} \cap \overline{\text{running}}) &\in [.95, .97] \\ 0 &\in \text{turn-key} \end{aligned} \quad (4)$$

According to principles described in Section 4, this KB derives a degree of belief of  $[.95, .97]$  for " $0 \in \text{running}^*$ ", which by the definition of our shorthand is equivalent to " $1 \in \text{running}$ ". This mechanism allows the agent to make predictions between certainty and falsehood, a capability that standard non-statistical approaches do not have. The agent can use this added power to make quantitative judgments about the relative merits of different courses of action, e.g. to compare the expected utilities of starting one car versus another.

## 5.2 Potatoes in tailpipes

This approach handles exceptions to action success by the way that new reference classes are chosen when additional sentences are added to the KB. For example, we add the following additional sentences to the KB (where "potato-in-tailpipe" is the intended interpretation of *p-in-t*):

$$\%(\text{running}^* \mid \text{turn-key} \cap \overline{\text{running}} \cap \text{p-in-t}) \in [0, .1] \quad (5)$$

$$0 \in \text{p-in-t} \quad (6)$$

This new KB derives the degree of belief of  $[0, .1]$  for  $1 \in \text{running}$  rather than the previous interval  $[.95, .97]$ . Remember, however, that if the interval in (5) were less focused than that of (4) (e.g.  $[.1, .98]$ ), then the previous interval would actually be preferred.

## 5.3 Exceptions to exceptions

Ginsberg and Smith [1987], noted that a problem with McCarthy's default solution to the qualification problem is that exceptions to action success, such as potatoes in tailpipes, may themselves have exceptions, such as a special exhaust system with two tailpipes. In other words, determining the sufficient conditions for concluding that a qualification defeats an action also gives rise to the qualification problem. This is not a problem in our approach: exceptions to exceptions simply give rise to more specific reference classes. Suppose that we know that our car has a special twin exhaust system. Given the following statistic:

$$\%(\text{running}^* \mid \text{turn-key} \cap \overline{\text{running}} \cap \text{p-in-t} \cap \text{twin-exh}) \in [.8, .9] \quad (7)$$

our degree of belief that the car runs will be based upon this new reference class. Our approach simply treats exceptions (and exceptions to exceptions, *ad infinitum*) precisely in the same fashion as it treats all qualifications: as evidence for choosing the appropriate reference class.

## 5.4 Causal generalizations

As we mentioned in Section 3, a crucial feature of our use of statistics for causal reasoning is how statistics generalize over what does *not* appear in the description of the reference class. For causal reasoning, this means that fluents that do not appear in a reference class are generalized over when that reference class is used in a statistic. For example, (3) embeds within it the relative impact of potatoes in tailpipes, as well as every other qualification that might defeat the car running. That is, 3 gives data about worlds in which potatoes are in the tailpipe, but it weights this data precisely

in proportion to the percent of worlds in which potatoes are in tailpipes, given that the key is turned. This fact is given by the following identity (which we take to be derivable in our rational corpus from axioms about proportion):

$$\begin{aligned} \%(\text{running}^* | \text{turn-key}) &= \%(\text{running}^* | \text{turn-key} \cap \text{p-in-t}) \%(\text{turn-key} | \text{p-in-t}) \\ &+ \\ &\%(\text{running}^* | \text{turn-key} \cap \overline{\text{p-in-t}}) \%(\text{turn-key} | \overline{\text{p-in-t}}) \end{aligned}$$

As a consequence, we get information about the possible effect of potatoes in tailpipes from the coarser statistic, even without considering potatoes in tailpipes. If potatoes in tailpipes occurred more frequently, this fact would be reflected in the coarse statistic, since the car would start less frequently when the key was turned. In fact, the general statistic embeds within it information about every exception, i.e. qualification, whether we know about such conditions and the problems that they might cause. This is how we solve the qualification problem.

## 6 Survivor Functions and The Frame Problem

Dean and Kanazawa [1988] developed a statistical solution to the frame (persistence) problem by using a concept from queuing theory called a *survivor function*. Using our notation and terminology, these survivor functions were expressed as statistics for each fluent set  $p$ :

$$f_p(\Delta) = \%(\{t : t + \Delta \in p\} | p)$$

Intuitively, these functions derive the probability that a fluent will remain true after a given length of time. Dean and Kanazawa observed that given the simpler statistic  $\%(p^* | p) = \lambda$ , under certain conditions (detailed in [Weber, 1989]), the survivor function is a geometric distribution  $f_p(\Delta) = \lambda^\Delta$  (or in the continuous case, an exponential distribution  $f_p(\Delta) = e^{-\lambda\Delta}$ ). For example, suppose there is a 90% chance that your car will not be towed from an illegal parking space during a given hour. Given certain assumptions about the independence of the trials, this means that there is a 81% chance that the car will not be towed after two hours, a 73% chance after three hours, etc. Therefore, the survivor function can be used to derive a degree of belief that a fluent persists over time, solving the frame problem.

Our approach subsumes the use of survivor functions, because the statistics they embody are a special case of the statistical causal rules we have described. In fact, our approach allows for much more flexible survivor functions that are sensitive to contextual information. For example, suppose the illegally parked car above is parked next to a fire hydrant, which defines the more specific reference class in the statistic:

$$\%(\text{not-towed}^* | \text{not-towed} \cap \text{by-hydrant}) \in [.85..85]$$

This statistic can be used to define a survivor function that incorporates this contextual information.

In addition to survivor functions, Dean and Kanazawa do allow an additional form of statistical assertion called a *projection rule*, which they use to derive a degree of belief that a fluent *changes*, e.g. in our notation:

$$\%(running^* | \overline{running} \cap turn-key \cap has-gas \dots) = x \quad (8)$$

In words, this statistic captures the proportion of times that the car will be running given that previously it was not running, its key was turned, it had gasoline, plus the truth of all other relevant preconditions. Projection rules are much the same as another special case of our statistical causal rules. The principal deficiency of their approach, however, is that they do not handle, or even recognize, the problem of conflicting reference classes. The reference class problem appears even within their restricted statistical assertions, when representing survivor functions and projection rules for both a fluent and its negation. Information about whether a fluent will change competes with information about whether the fluent's negation will persist. For example, consider the survivor function for the negation of running:

$$\%(\overline{running} | \overline{running}) = z$$

from which it follows that:

$$\%(running^* | \overline{running}) = 1 - z$$

which can be applied whenever (8) can be applied, leading to a choice between  $x$  and  $1 - z$  as the probability of a sentence such as " $1 \in running$ ". Therefore, since they do not provide a mechanism like specificity for resolving competing reference classes, they do not solve the frame problem in a consistent way. For the same reason, they also do not address the qualification problem at all. Our approach solves both problems by using the simple and powerful mechanisms of well-established techniques for statistical inference.

## 7 Objections to Statistical Inference

There are several objections that will likely be voiced concerning this general approach. One concerns the fact that agents cannot have all possible statistics, and as such, may not have the statistics required for "common-sense" inferences. This concern is partly addressed in Section 5.4, where we note that large reference classes provide useful information. In addition, by observing the statistical independence of different properties (that what I ate for breakfast does not affect whether my car starts), agents will not be obliged to keep statistics at a finer grain.

Since statistics are interval valued, one can argue that there always exists at least the trivial statistic for any set: the proportion of  $P$ 's in our universe is in the interval

[0,1]. As Loui remarks [Loui, 1987], Kyburg's system suggests "how to choose between evidence that is strong but ill-founded, and evidence that is well-founded but weak," where well-foundedness refers to the specificity of the reference class, and strength to the size of the interval.

Dennett [1987] raises the objection that one must "represent (so that it can be used) all that hard-won empirical information — a problem that arises independently of the truth value, probability, warranted assertability, or subjective certainty of any of it." That is, even if the agent had all of the appropriate statistics, using such statistics would be a rather nasty computational problem. But clearly, any solution to this problem will amount to a trade off, as Fodor points out [1987], between economy and warrant. One will trade the computational advantage of incomplete reasoning systems that jump (often over great distances) to conclusions without considering *all* available knowledge, against the likelihood that such conclusions will be false. We argue that by having access to the statistical basis for belief, agents will better be able to compare the value of knowledge with the cost of obtaining and referencing it. That is, each datum will be required to "pull its own weight" [Hartman and Tenenberg, 1987].

Loui [Loui, 1987], has made a serious attempt at computing reference classes, based upon Kyburg's system. He notes that "Finding candidate reference classes ... [is] intensively set-theoretic," and as such, incurs a high computational cost. However, using a constrained language and specialized inference procedures, implementations "proceeded at practically useable speeds." He concludes "The probability intervals produced are intuitively appealing. ... The next step is to see how well these ideas fare in applications." We view action reasoning as one such application.

An additional approach that one of the authors is taking [Weber, 1989], concerns the incremental computation of reference classes. Preconditions are added to the reference class in order of decreasing statistical impact on the degree of belief for the effect. Since the impact computations are not data dependent, this incremental refinement of belief can be performed quickly on parallel hardware, allowing the reasoner to trade off the accuracy of a prediction with the time it takes to derive it. This approach provides a novel answer to Dennett's objection discussed above.

## 8 Conclusion

The qualification problem concerns determining sufficient conditions to guarantee the success of an agent's actions. Because of an agent's resource limitations and the complexity of most interesting domains, reasoning about action success will have to be made under considerable uncertainty. We propose to solve this problem by associating a statistically founded probability with each proposition, thereby casting the qualification problem as an instance of the reference class problem. Since statistics capture knowledge obtained over large sets of events, an agent's ability to draw strong

conclusions about what will be true in the future is not based solely upon what is known at the current instant. This approach therefore provides a uniform basis not only for solving the qualification problem, but also the frame problem, and other problems of causal prediction.

## References

- [Dean and Kanazawa, 1988] Thomas Dean and Keiji Kanazawa. Probabilistic temporal reasoning. In *Proceedings of AAAI-88*, pages 524-528, August 1988.
- [Dennett, 1987] Daniel Dennett. Cognitive wheels: The frame problem in ai. In Pylyshyn, editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, pages 41-64. Ablex Publishing Company, Norwood, New Jersey, 1987.
- [Elgot-Drapkin et al., 1987] Jennifer Elgot-Drapkin, Michael Miller, and Donald Perlis. The two frame problems. In Frank M. Brown, editor, *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop*. Morgan Kaufman, April 1987.
- [Etherington, 1987] David W. Etherington. More on inheritance hierarchies with exceptions: Default theories and inferential distance. In *Proceedings of AAAI-87*, pages 352-357, 1987.
- [Fodor, 1987] Jerry Fodor. Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Pylyshyn, editor, *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, pages 139-149. Ablex Publishing Company, Norwood, New Jersey, 1987.
- [Ginsberg and Smith, 1987] Matthew L. Ginsberg and David E. Smith. Reasoning about action ii: The qualification problem. Technical report, Knowledge Systems Laboratory, Stanford University, May 1987.
- [Hartman and Tenenberg, 1987] Leo Hartman and Josh Tenenberg. Performance in practical problem solving. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 535-540, 1987.
- [Kyburg, 1970] Henry E. Kyburg, Jr. Conjunctivitis. In M. Swain, editor, *Induction, Acceptance and Rational Belief*. Reidel, 1970.
- [Kyburg, 1974] Henry E. Kyburg, Jr. *The Logical Foundations of Statistical Inference*. Reidel, 1974.
- [Kyburg, 1983a] Henry E. Kyburg, Jr. Rational belief. Technical Report Cognitive Science Technical Report 3, University of Rochester, 1983.
- [Kyburg, 1983b] Henry E. Kyburg, Jr. The reference class. *Philosophy of Science*, 50:374-397, 1983.
- [Lifschitz, 1987] Vladimir Lifschitz. Formal theories of action: Preliminary report. In Frank M. Brown, editor. *The Frame Problem in Artificial Intelligence: Proceedings of the 1987 Workshop*. Morgan Kaufman, April 1987.

- [Loui, 1987] Ronald P. Loui. *Theory and Computation of Uncertain Inference and Decision*. PhD thesis, University of Rochester Computer Science Department, September 1987.
- [McCarthy and Hayes, 1981] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In Bonnie Lynn Webber and Nils J. Nilsson, editors, *Readings in Artificial Intelligence*, pages 431-450. Tioga, 1981.
- [McCarthy, 1977] John McCarthy. Epistemological problems of artificial intelligence. In *Proceedings of IJCAI-77*, pages 1034-1044, Cambridge, MA, 1977.
- [McCarthy, 1980] John McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27-39, April 1980.
- [McCarthy, 1984] John McCarthy. Applications of circumscription to formalizing common sense knowledge. In *Proceedings of the AAAI Non-Monotonic Workshop*, pages 295-324, October 1984.
- [Nute, 1986] Donald Nute. Ldr: A logic for defeasible reasoning. Technical Report ACMC Research Report 01-0013, Advanced Computational Methods Center, University of Georgia, Athens, Georgia, 1986.
- [Pollock, 1984] John Pollock. Foundations for direct inference. *Theory and Decision*, 17, 1984.
- [Reichenbach, 1949] Hans Reichenbach. *The Theory of Probability*. University of California Press, 1949.
- [Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1,2):81-132, April 1980.
- [Shoham, 1988] Yoav Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*. The MIT Press, Cambridge, MA, 1988.
- [Touretzky, 1984] David S. Touretzky. Implicit ordering of defaults in inheritance systems. In *Proceedings of AAAI-84*, pages 322-325, 1984.
- [Weber, 1989] Jay C. Weber. *Principles and Algorithms for Causal Reasoning with Uncertain Knowledge*. PhD thesis, University of Rochester Computer Science Department, May 1989.