FROM KNOWLEDGE TO BELIEF

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Daphne Koller
October 1994

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Joseph Y. Halpern
(Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Yoav Shoham

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

Vaughan Pratt

Approved for the University Committee on Graduate Studies:

Preface

When acting in the real world, an intelligent agent must make decisions under uncertainty. For example, a doctor may need to decide upon the treatment for a particular patient. The standard solution to this problem is based on decision theory. It requires the agent to assign degrees of belief or subjective probabilities to the relevant assertions. The degrees of belief assigned should be based on the information available to the agent. A doctor, for example, may have information about particular patients, statistical correlations between symptoms and diseases, physical laws, default rules, and more. This thesis describes one approach, called the random-worlds method, for inducing degrees of belief from very rich knowledge bases.

The random-worlds method is based on the principle of indifference: it treats as equally likely all the worlds that the agent considers possible. It deals with knowledge bases expressed in a language that augments first-order logic with statistical statements. By interpreting default rules as qualitative statistics, the approach integrates qualitative default reasoning with quantitative probabilistic reasoning. The thesis shows that a large number of desiderata that arise in *direct inference* (reasoning from statistical information to conclusions about individuals) and in default reasoning follow provably from the semantics of random worlds. Thus, random worlds naturally derives important patterns of reasoning such as specificity, inheritance, indifference to irrelevant information, and a default assumption of independence. Furthermore, the expressive power of random worlds and its intuitive semantics allow it to deal well with examples that are too complex for most other inductive reasoning systems.

The thesis also analyzes the problem of computing degrees of belief according to random worlds. This analysis uses techniques from *finite model theory* and *zero-one laws*. We show that, in general, the problem of computing degrees of belief is undecidable, even for knowledge bases with no statistical information. On the other hand, for knowledge bases that involve only unary predicates, there is a tight connection between the random-worlds method and the principle of maximum entropy. In fact, maximum entropy can be used as a computational tool for computing degrees of belief in many practical cases.

Acknowledgements

First and foremost, I would like to thank my advisor Joe Halpern, for his constant support and patience throughout the ups and downs of this research. Joe has been more than an advisor; he has been a mentor, a friend, and one of the best co-authors a person could have. From Joe I learned how to do research, how to write, and what to write about. For all this, and for much much more, thank you!

I have been fortunate to have more than one mentor. My first mentor in research was Danny Dolev, my M.Sc. advisor. He started me out on the right path. Since then, his advice and support have ensured that I stay on it. Thank you!

This dissertation is the culmination of three years of research, done in collaboration with Fahiem Bacchus, Adam Grove, and Joe Halpern. I would like to thank my co-authors for their enormous contribution to this work. Fahiem was our source for many of the AI issues addressed in this dissertation. He kept us grounded in the real problems this research is all about. Adam has contributed to every aspect of this research. His view on the issues, usually the opposite of my own, has enriched this research in many ways. Adam's support and friendship has helped me in more ways than I can count. Joe's curiosity constantly led us to investigate new problems, and his optimism assured that we eventually solve them. Thank you all!

I am also grateful to those people who read all or part of this work, and whose comments made it better: Ron Fagin, Nir Friedman, Vaughan Pratt, Yoav Shoham, and Moshe Vardi.

I would also like to thank my different sources of funding: a Rothschild Foundation doctoral fellowship, the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080, IBM Almaden Research Center, and research grants of Zohar Manna, Vaughan Pratt, and Yoav Shoham. They made this work possible.

I am grateful to many people whose friendship has made my years at Stanford so memorable. In particular, I would like to thank: Carolyn and Efi Arazi, Michal and Yossi Azar, Marcia Derr, Kathleen Fisher, Nir Friedman, Yossi Friedman, Adam Grove, Allegra and David Karger, Andrew Kosoresow, Hillary and Lawrence Levy, Merav and Alon Levy, Nina and Serge Plotkin, Scott Roy, Sam Srinivas, and Liz (Shoshi) Wolfe.

Above all I would like to thank my family. My parents, Diza and Dov Koller, taught me to strive to be the best that I could be. Their constant love and support is the foundation on which everything else is built. My brother, Dan Alexander Koller, was always proud and never critical. My husband, Dan Avida, made sure I remembered there is more to life than a Ph.D. His implicit (and very vocal) faith in my abilities kept me going many times. To all of you, with love, I dedicate this thesis.

Contents

List of Figures

Chapter 1

Introduction

1.1 Making Decisions

An agent acting in the real world must make autonomous decisions about its course of action. For example, a doctor (or a medical expert system) may need to decide on a treatment for a particular patient, say Eric (see Figure 1.2). These decisions are always made under *uncertainty* regarding the true state of world. The standard technique for making decisions under uncertainty is derived from *decision theory* [Sav54]. Essentially, it involves defining the outcome of each action in each state of the world. If we then assign *probabilities* to the different states of the world, and *utilities* to the various outcomes, we can choose the action that has the maximum *expected utility* among all possible actions (see Figure 1.1).

In this thesis, we investigate the problem of assigning probabilities to various events of interest. For example, the doctor's decision regarding a treatment for Eric should certainly depend on the probabilities for the different diseases that Eric might have. Thus, she may be interested in assigning a probability to the event "Eric has hepatitis". This probability should certainly be based, in some principled way, on the doctor's knowledge. In general, the doctor will have a very rich knowledge base, that might contain information of different types, including:

- statistical information, such as "80% of patients with jaundice have hepatitis",
- first-order laws, such as "all patients with hepatitis exhibit jaundice",

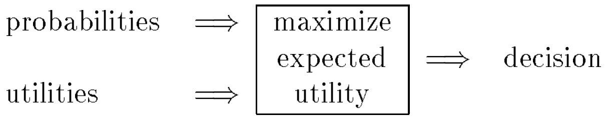

Figure 1.1: The decision-theoretic paradigm

Figure 1.2: A particularly simple decision-making situation

- default rules, such as “patients with hepatitis typically have a fever”,
- information about the particular patient at hand, such as “Eric has jaundice”.

However, even a very rich knowledge base will not often suffice to determine with certainty which event holds.

How do we use the knowledge base to determine the probabilities for the event of interest? The answer to this question is complicated by the fact that an event such as “Eric has hepatitis” is not a probabilistic one. In the true world, this assertion is either true or false (Eric either does or does not have hepatitis), so that the associated “probability” is either 0 or 1. This answer to the question is clearly not a useful one. It suggests that the only probability we can derive for this event is necessarily a *subjective probability* or a *degree of belief*. But if the probability is purely subjective, can we validate it? On the one hand, we cannot say that one subjective probability is correct while another one is not. On the other hand, it seems that certain techniques for deriving subjective probabilities are more appropriate than others. In particular, we would like to maintain a strong connection between the information in the knowledge base and the resulting degrees of belief. This thesis describes one particular method that allows the agent to use its knowledge base to assign degrees of belief in a principled manner; we call this method the *random-worlds method*.

1.2 Other approaches

There has been a great deal of work addressing aspects of this general problem. Two large bodies of work that are particularly relevant are the work on *direct inference*, going back to Reichenbach [Rei49], and the various approaches to *nonmonotonic reasoning.*

Direct inference deals with the problem of deriving degrees of belief from statistical information. The basic approach is to try to find a suitable *reference class* whose statistics we can use to determine the degree of belief. For example, reference class systems will ascribe a degree of belief to the assertion "Eric has hepatitis", based on the frequency of hepatitis among the class of individuals who are "just like Eric". Those cases, for which we do not have statistics for the appropriate class, are handled using the heuristic of *specificity* — choosing the most specific applicable reference class. As we show, this type of reasoning leads to intuitive answers in many cases. However, it runs into problems when attempting to deal with more complex examples. In particular, reference-class reasoning has difficulties combining different pieces of evidence in cases where the statistics from more than one reference class are applicable. We view this as resulting from a more general problem with this type of reasoning: it attempts to substitute a single local piece of information (the statistics for one reference class) for the entire knowledge base.

Direct inference addresses knowledge bases containing statistical information. Nonmonotonic reasoning, on the other hand, deals to a large extent with knowledge bases containing default rules. Nevertheless, some of the same reasoning patterns arise in both types of formalism. In particular, the issues of specificity and *inheritance* are relevant in both contexts. As we show, even in the narrow context of default reasoning, the goal of obtaining specificity and inheritance in a single formalism has proved to be elusive. Moreover, those few systems that have, to some extent, achieved this goal are typically very restricted in terms of expressive power.

As we shall show, none of the systems proposed for either reference-class reasoning or nonmonotonic reasoning can deal adequately with the large and complex knowledge bases we are interested in. Furthermore, none can handle rich knowledge bases that may contain first-order, default, and statistical information. Nevertheless, these approaches do provide useful yardsticks by which to measure the adequacy of the random-worlds approach. As we shall show, the random-worlds approach deals with the paradigmatic problems in both nonmonotonic and reference-class reasoning remarkably well.

1.3 The Random-Worlds Approach

We now provide a brief overview of the random-worlds approach. As we suggested above, we want to deal with rich knowledge bases that allow not only first-order information, but statistical information and default information. To do this, we use a variant of the language introduced by Bacchus [Bac90]. Bacchus's language augments first-order logic by allowing statements of $\|Hep(x)|Jaun(x)\|_x = 0.8$, which says that 80% of patients with jaundice have hepatitis. Notice that, in finite models, this statement has the (probably unintended) consequence that the number of patients with jaundice is a multiple of 5. To avoid this problem, we use approximate equality rather than equality, writing $||Hep(x)|Jaun(x)||_x \approx 0.8$, read "approximately 80% of patients with jaundice have hepatitis". Intuitively, this says that the proportion of jaundiced patients with hepatitis is arbitrarily close to 80%: i.e., within some tolerance τ of 0.8.

Not only does the use of approximate equality solve the problem of unintended consequences, it has another significant advantage: it lets us express default information. We interpret a statement such as "Birds typically fly" as expressing the statistical fact that "Almost all birds fly". Using approximate equality, we can represent this as $||Fly(x)|Bird(x)||_x \approx 1$. (This interpretation is closely related to various approaches applying probabilistic semantics to nonmonotonic logic; see Pearl [Pea89] for an overview and Section 7.2.3 for a discussion of the connection between the approaches.)

Having described the language in which our knowledge base is expressed, we now need to decide how to assign degrees of belief given a knowledge base. Perhaps the most widely used framework for assigning subjective probabilities is the Bayesian paradigm. There, one assumes a space of possibilities and a *prior* probability distribution over this space, and calculates *posterior* probabilities by conditioning on what is known (in our case, the knowledge base). To use this approach, we must specify the space of possibilities and the distribution over it. In Bayesian reasoning, relatively little is said about how this should be done. Indeed, the usual philosophy is that these decisions are subjective. The difficulty of making these decisions seems to have been the main reason for the historic unpopularity of the Bayesian approach in symbolic AI [MH69].

Our approach is different. We assume that the *KB* contains all the knowledge the agent has, and we allow a very expressive language so as to make this assumption reasonable. This assumption implies that any knowledge the agent has that could influence the prior distribution is already included in the *KB*. As a consequence, we give a single uniform construction of a space of possibilities and a distribution over it. Once we have this probability space, we can use the Bayesian approach: To compute the probability of an assertion φ given *KB*, we condition on *KB*, and then compute the probability of φ using the resulting posterior distribution.

So how do we choose the probability space? One very general strategy, dicussed by Halpern [Hal90], is to give semantics to degrees of belief in terms of a probability distribution over a set of *possible worlds*, or first-order models. This semantics clarifies the distinction between statistical assertions and degrees of belief. As we suggested above, a statistical assertion such as $||Hep(x)|Jaun(x)||_x \approx 0.8$ is true or false in a particular world, depending on how many jaundiced patients have hepatitis in that world. It is this fact that allows us to condition on a knowledge base containing statistical statements. On the other hand, a degree of belief is neither true nor false in a particular world — it has semantics only with respect to the entire set of possible worlds and a probability distribution over them. There is no necessary connection between the information in the agent's *KB* and the distribution over worlds that determines her degrees of belief. However, we clearly want to use the information in the knowledge base, particularly the statistical information, in assigning degrees of belief. As this thesis shows, the random-worlds method is a powerful technique for going from a statistical knowledge base to degrees of belief.

To define our probability space, we have to choose an appropriate set of possible worlds. Given some domain of individuals, we stipulate that the set of worlds is simply the set of all first-order models over this domain. That is, a possible world corresponds to a particular way of interpreting the symbols in the agent's vocabulary over the domain. In our context, we can assume that the "true world" has a finite domain, say of size N. In fact, without loss of generality, we assume that the domain is $\{1, \ldots, N\}$. This domain induces a (finite) set of possible worlds, as described.

Having defined the probability space (the set of possible worlds), we must construct a probability distribution over this set. For this, we give perhaps the simplest possible definition: we assume that all the possible worlds are equally likely (that is, each world has the same probability). This can be viewed as an application of the *principle of indifference*; since we assumed that all the agent knows is incorporated in its knowledge base, the agent has no *a priori* reason to prefer one world over the other. It is therefore reasonable to view all worlds as equally likely. Interestingly, the principle of indifference (sometimes also called the *principle of insufficient reason*) was originally promoted as part of the very definition of probability, when the field was originally formalized by Jacob Bernoulli and others; the principle was later popularized further and applied with considerable success by Laplace. (See [Hac75] for an historical discussion.) It later fell into disrepute as a general definition of probability, largely because of the existence of paradoxes that arise when the principle is applied to infinite and/or continuous probability spaces. This thesis makes no attempt to define the notion of probability. We claim, however, that the principle of indifference is a natural and effective way of assigning degrees of belief, particularly if we restrict attention to a finite collection of worlds with finite domains.

Combining our choice of possible worlds with the principle of indifference, we obtain our prior distribution. We can now induce a degree of belief in φ given KB by conditioning on KB to obtain a posterior distribution, and then computing the probability of φ according to this new distribution. It is easy to see that the degree of belief in φ given KB is the fraction of possible worlds satisfying KB that also satisfy φ. This process is demonstrated in Figures 1.3, 1.4, and 1.5. Figure 1.3 shows us a set of possible worlds over a vocabulary dealing with hepatitis, jaundice, and the individual Eric. Initially, all the possible worlds are equally likely. Figure 1.4 shows us the process of conditioning on the knowledge base KB — in this case, the assertions "80% of patients with jaundice have hepatitis" and "Eric has jaundice". The magnified world on the top right is outlawed by KB because the statistical assertion is violated. The one on the bottom right is outlawed because it does not support the fact that Eric has jaundice. Finally, Figure 1.5 shows us the procedure of computing the probability of φ in the posterior distribution, by counting the fraction of remaining worlds that satisfy φ — in this case, "Eric has hepatitis".

One problem with the approach as stated so far is that, in general, we do not know the domain size N. Typically, however, N is known to be large. We therefore approximate the degree of belief for the true but unknown N by computing the value of this degree of belief as N grows large. The result is our random-worlds method.

The key ideas in the approach are not new. Many of them can be found in the work of Johnson [Joh32] and Carnap [Car50, Car52], although these authors focus on knowledge bases that contain only first-order information, and restrict to unary predicates. Similar approaches have been used in the more recent works of Shastri [Sha89] and of Paris and Vencovska [PV89], in the context of a unary statistical language. Chuaqui's recent work [Chu91] is also relevant. His work, although technically quite different from ours, shares the idea of basing a general theory of probabilistic reasoning upon the notions of indifference and symmetry. The work of Chuaqui and the work of Carnap investigate very different issues from those we examine in this thesis. For example, Carnap, and others who later continued to develop his ideas, were very much interested in learning statistics, and even harder, learning universal laws. While we believe the question of learning is very important (see Section 8.1.3), we have largely concentrated on understanding (and generalizing) the process of going from statistical information and default rules to inferences about particular individuals. Many of the new results we describe reflect this very different emphasis.

1.4 Validation

Having defined the method, how do we judge its reasonableness? Fortunately, as we mentioned, there are two large bodies of work on related problems on which we can draw for guidance: reference-class reasoning and default reasoning. While, as we pointed out, none of the solutions suggested for these problems seems adequate, the years of research have resulted in some strong intuitions regarding what answers are intuitively reasonable for certain types of queries. As we have observed, these intuitions often lead to identical desiderata. In particular, most systems (of both types) espouse some form of preference for more specific information and the ability to ignore irrelevant information (irrelevance is often closely related to inheritance).

We show that the random-worlds approach also satisfies these desiderata. In fact, in the case of random worlds, these properties follow from a much more general theorem. We prove that, in those cases where there is some piece of statistical information that should "obviously" be used to determine a degree of belief, random worlds does use this information. The different desiderata, such as preference for more specific information and indifference to irrelevant information, including inheritance and even *exceptional-subclass inheritance*, follow as easy corollaries (see Section 2.2). We also show that random worlds provides reasonable answers in other contexts, not covered by the standard specificity and irrelevance heuristics. In particular, random worlds combines statistical information from different reference classes when appropriate.

Thus, the random-worlds method is indeed a powerful one, that can deal with rich knowledge bases and still produce the answers that people have identified as being the most appropriate ones.

1.5 Reader's Guide

The rest of the thesis is organized as follows. In Chapter 2, we outline some of the major themes and problems in the work on reference classes and on default reasoning. Since one of our major claims is that the random-worlds approach solves many of these problems, this will help set our work in context. In Chapter 3, we describe the random-worlds method in detail. We also show how the approach naturally embeds a very expressive nonmonotonic reasoning system. In Chapter 4, we state and prove a number of general theorems about the properties of the approach, and show how various desiderata follow from these theorems.

In the second part of the thesis, we investigate in depth the problem of computing degrees of belief. In Chapter 5, we show that the expressivity of random worlds leads to some negative computational consequences. In particular, computing degrees of belief, or even approximating them, is a highly undecidable problem. This is the case even for a language with no statistical statements. While this is unfortunate, it is not surprising, since the language used in random worlds contains the full expressive power of first-order logic. Therefore, as in first-order logic, we must search for classes of cases where these problems do not arise.

Chapters 6 and 7 address one such class: that of knowledge bases that contain only unary predicates and constants. As we explain later on, this is a very important and practical class of problems. Under this restriction, we show in Chapter 6 how to compute degrees of belief in a language with no statistical statements. In Chapter 7, we show a strong connection between random worlds and the *principle of maximum entropy* in the unary case. Thus, for a large class of interesting problems, a maximum-entropy computation can be used to calculate the degrees of belief.

Finally, in Chapter 8, we discuss some possible criticisms of the random-worlds method and its possible impact.

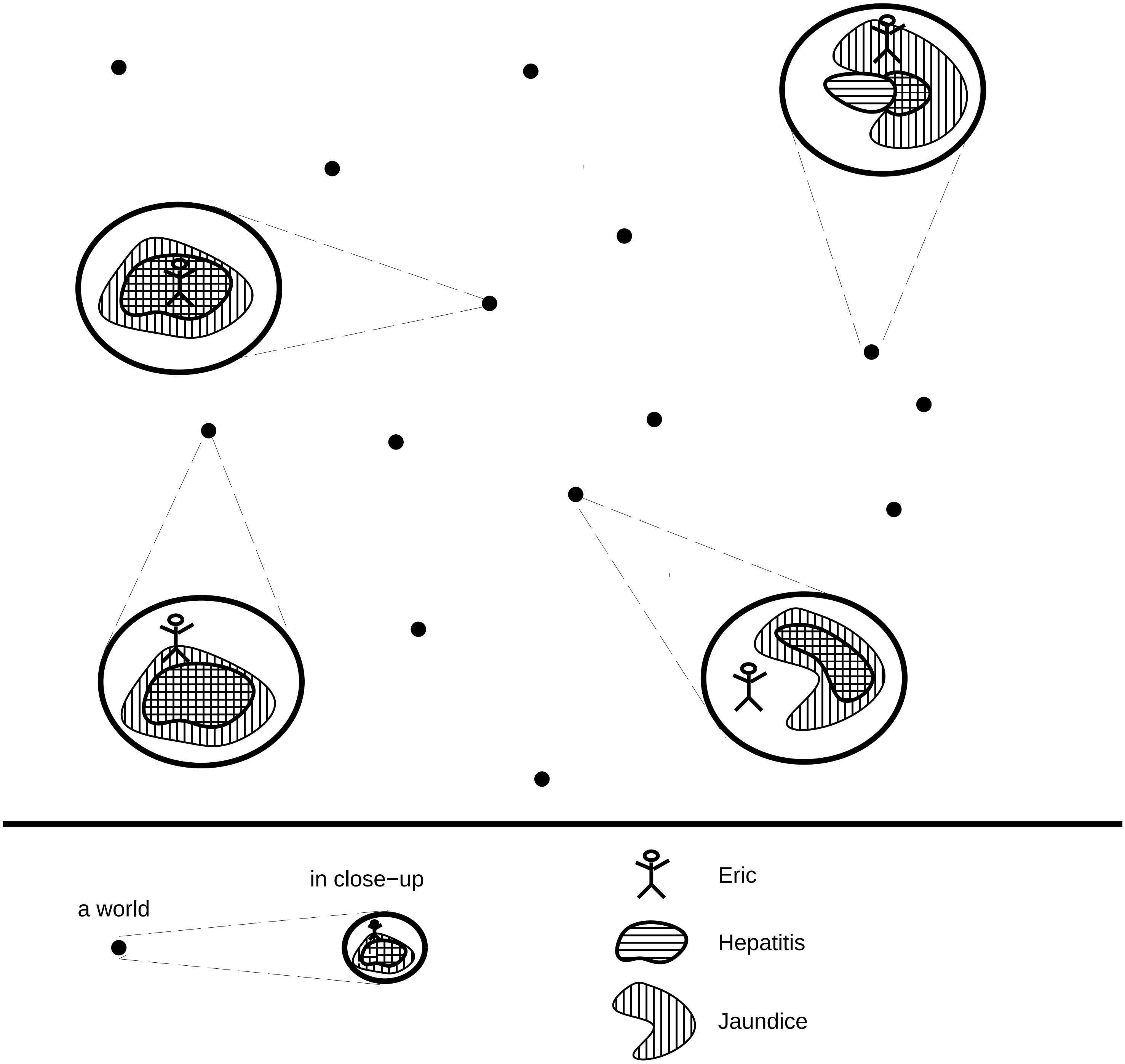

Figure 1.3: The set of possible worlds

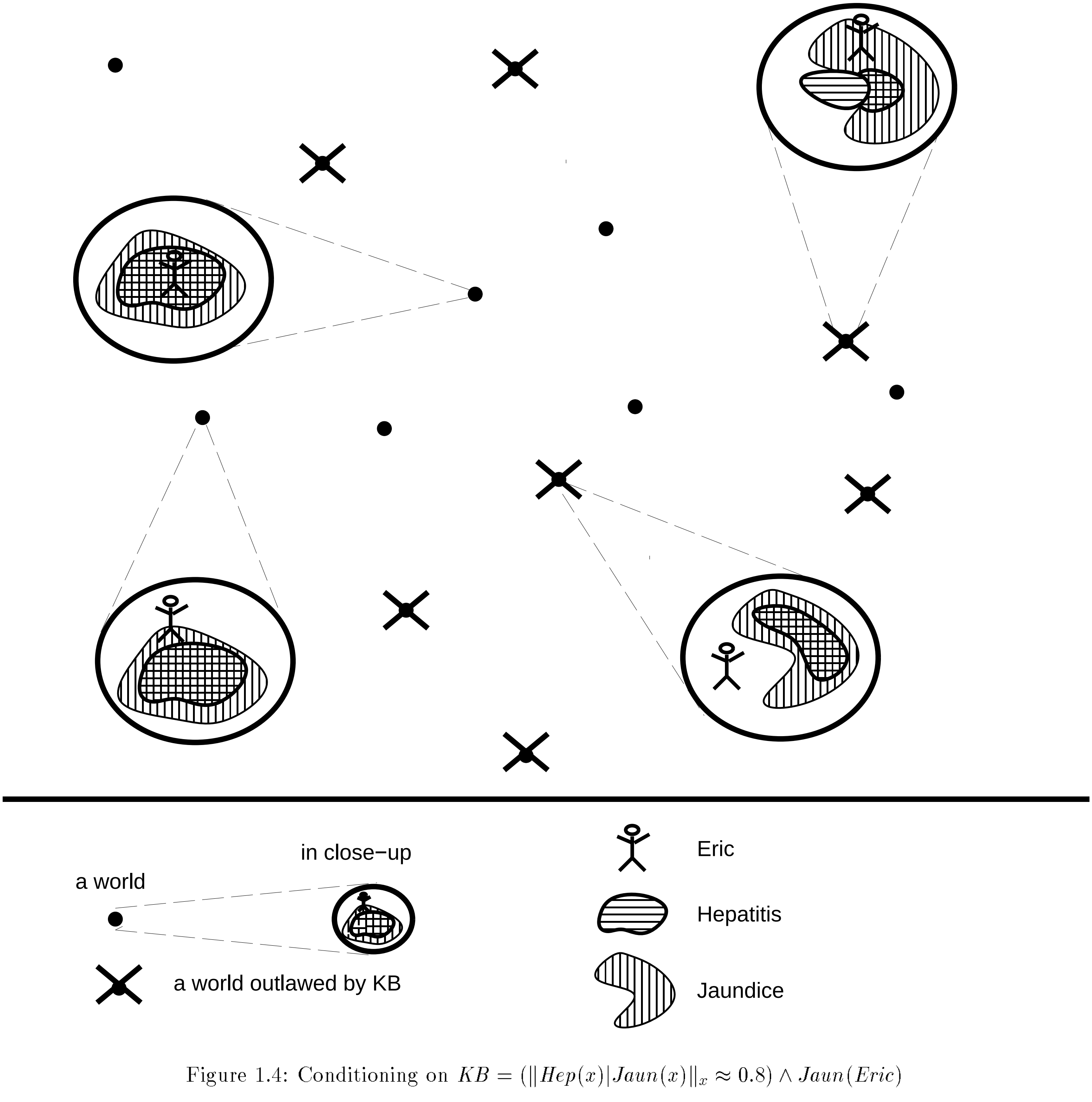

Figure 1.4: Conditioning on $KB = (\|Hep(x)|Jaun(x)\|_x \approx 0.8) \wedge Jaun(Eric)$

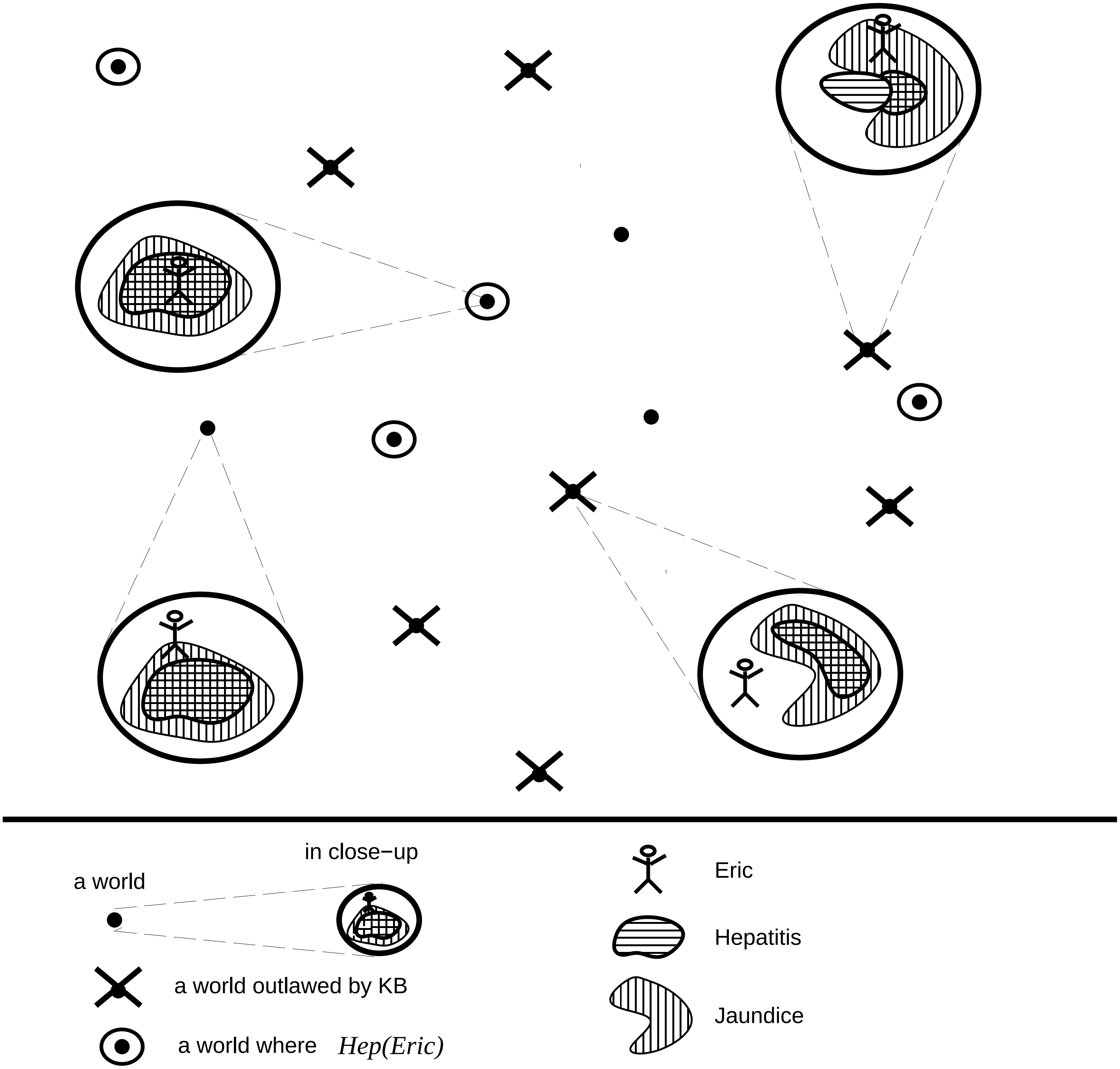

Figure 1.5: Computing the probability of $\varphi = \mathit{Hep}(\mathit{Eric})$

Chapter 2

Background work

2.1 Reference classes

Strictly speaking, the only necessary relationship between objective knowledge about frequencies and proportions on the one hand and degrees of belief on the other hand is the simple mathematical fact that they both obey the axioms of probability. But in practice we usually hope for a deeper connection: the latter should be based on the former in some "reasonable" way. Of course, the random-worlds approach that we are advocating is precisely a theory of how this connection can be made. But our approach is far from the first to attempt to connect statistical information and degrees of belief. Most of the earlier work is based on the idea of finding a suitable *reference class*. In this section, we review some of this work and show why we believe that this approach, while it has some intuitively reasonable properties, is inadequate as a general methodology. (See also [BGHK93b] for further discussion of this issue.) We go into some detail here, since the issues that arise provide some motivation for the results that we prove later regarding our approach.

2.1.1 The basic approach

The earliest sophisticated attempt at clarifying the connection between objective statistical knowledge and degrees of belief, and the basis for most subsequent proposals, is due to Reichenbach [Rei49]. Reichenbach describes the idea as follows:

> "If we are asked to find the probability holding for an individual future event, we must first incorporate the case in a suitable *reference class*. An individual thing or event may be incorporated in many reference classes.... We then proceed by considering the narrowest reference class for which suitable statistics can be compiled."

Although not stated explicitly in this quote, Reichenbach's approach was to equate the degree of belief in the individual event with the statistics from the chosen reference class. As an example, suppose that we want to determine a probability (i.e., a degree of belief) that Eric, a particular patient with jaundice, has the disease hepatitis. The particular individual Eric is a member of the class of all patients with jaundice. Hence, following Reichenbach, we can use the class of all such patients as a reference class, and assign a degree of belief equal to our statistics concerning the frequency of hepatitis among this class. If we know that this frequency is 80%, then we would assign a degree of belief of 0.8 to the assertion that Eric has hepatitis.

Reichenbach's approach consists of (1) the postulate that we use the statistics from a particular reference class to infer a degree of belief with the same numerical value, and (2) some guidance as to how to choose this reference class from a number of competing reference classes. We consider each point in turn.

In general, a *reference class* is simply a set of domain individuals[1] for which we have "suitable statistics" that contains the particular individual about whom we wish to reason. In our framework, we take a reference class to be the set of individuals satisfying a formula $\psi(x)$. The requirement that the particular individual c we wish to reason about belongs to the class is represented by the logical assertion $\psi(c)$. But what does the phrase "suitable statistics" mean? For purposes of illustration, we might suppose a suitable statistic is some (nontrivial) closed interval in which the proportion or frequency lies. (However, see the discussion in Section 2.1.3 regarding the problems with this interpretation.) More precisely, consider some query $\varphi(c)$, where φ is some logical assertion and c is a constant, denoting some individual in the domain. Then $\psi(x)$ is a reference class for this query if we know both $\psi(c)$ and $\|\varphi(x)|\psi(x)\|_x \in [\alpha, \beta]$, for some nontrivial interval $[\alpha, \beta]$. That is, we know that c has property ψ, and that among the class of individuals that possess property ψ, the proportion that also have property φ is between α and β. If we decide that this is the appropriate reference class, then Reichenbach's approach would allow us to conclude $\Pr(\varphi(c)) \in [\alpha, \beta]$, i.e., the (degree of belief) probability that c has property φ is between α and β. Note that the appropriate reference class for the query $\varphi(c)$ depends both on the formula $\varphi(x)$ and on the individual c.

Given a query $\varphi(c)$, there will often be many reference classes that are arguably appropriate for it. For example, say we know both $\psi_1(c)$ and $\psi_2(c)$, and we have two pieces of statistical information: $\|\varphi(x)|\psi_1(x)\|_x \in [\alpha_1, \beta_1]$ and $\|\varphi(x)|\psi_2(x)\|_x \in [\alpha_2, \beta_2]$. In this case both $\psi_1(x)$ and $\psi_2(x)$ are reference classes for $\varphi(c)$; depending on the values of the α's and β's, they could assign conflicting degrees of belief to $\varphi(c)$. The second part of Reichenbach's approach is intended to deal with the problem of how to choose a single reference class from a set of possible classes. Reichenbach recommended preferring the narrowest class, i.e., a preference for more specific information. In this example, if we know $\forall x\, (\psi_1(x) \Rightarrow \psi_2(x))$, i.e., if we know that the class $\psi_1(x)$ is a subset of the class $\psi_2(x)$, then Reichenbach's approach would allow us to conclude that $\Pr(\varphi(c)) \in [\alpha_1, \beta_1]$; that is, it instructs us to use the narrower reference class $\psi_1(x)$ in preference to the wider reference class $\psi_2(x)$.

[1] These "individuals" may be individual people, individual objects, or individual events (such as coin tosses). We use the term "individual" from here on, for definiteness. Furthermore, in our examples, we restrict our attention to reasoning about single individuals. In general, both reference-class reasoning and random worlds can be applied to queries such as "Did Eric infect Tom", which involve reasoning about a number of individuals simultaneously.

These two parts of Reichenbach's approach — using statistics taken from a class as a degree of belief about an individual and preferring statistics from more specific classes — are generally reasonable and intuitively compelling when applied to simple examples. Of course, even on the simplest examples Reichenbach's strategy cannot be said to be "correct" in any absolute sense. Nevertheless, it is impressive that people agree so widely on the reasonableness of the answers. As we show later, the random-worlds approach agrees with both aspects of Reichenbach's approach when applied to simple (and noncontroversial) examples. Unlike Reichenbach's approach, however, the random-worlds approach derives these intuitive answers from more basic principles. As a result it is able to deal well with more complex examples that defeat Reichenbach's approach.

Despite its successes, there are several serious problems with Reichenbach's approach. For one thing, defining what counts as a "suitable statistic" is not easy. For another, it is clear that the principle of preferring more specific information rarely suffices to deal with the cases that arise with a rich knowledge base. Nevertheless, much of the work on connecting statistical information and degrees of belief, including that of Kyburg [Kyb83, Kyb74] and of Pollock [Pol90], has built on Reichenbach's ideas of reference classes and the manner in which choices are made between reference classes. As a result, these later approaches all suffer from a similar set of difficulties. These problems are discussed in the remainder of this section.

2.1.2 Identifying reference classes

Recall that we took a reference class to be simply a set for which we have "suitable statistics". But the fact that any set of individuals can serve as a reference class leads to problems. Assume we know $\mathit{Jaun}(\mathit{Eric})$ and $\|\mathit{Hep}(x)|\mathit{Jaun}(x)\|_x \approx 0.8$. In this case $\mathit{Jaun}(x)$ is a legitimate reference class for the query $\mathit{Hep}(\mathit{Eric})$. Therefore, we would like to conclude that $\Pr(\mathit{Hep}(\mathit{Eric})) = 0.8$. But *Eric* is also a member of the narrower class $\{\textit{jaundiced patients without hepatitis}\} \cup \{\mathit{Eric}\}$ (i.e., the class defined by the formula $(\mathit{Jaun}(x) \wedge \neg\mathit{Hep}(x)) \vee x = \mathit{Eric})$, and the proportion of hepatitis patients in this class is approximately 0%. Thus, the conclusion that $\Pr(\mathit{Hep}(\mathit{Eric})) = 0.8$ is disallowed by the rule instructing us to use the most specific reference class. In fact, it seems that we can always find a narrower class that will give a different and intuitively incorrect answer. This example suggests that we cannot take an arbitrary set of individuals to be a reference class; it must satisfy additional criteria.

Kyburg and Pollock deal with this problem by by placing restrictions on the set of allowable reference classes that, although different, have the effect of disallowing disjunctive reference classes, including the problematic class described above. This approach suffers from from two problems. First, as Kyburg himself has observed [Kyb74], these restrictions do not eliminate the problem. The second problem is that restricting the set of allowable reference classes may prevent us from making full use of the information we have. For example, the genetically inherited disease Tay-sachs (represented by the predicate TS) appears only in babies of two distinct populations: Jews of east-European extraction (EEJ), and French-Canadians from a certain geographic area (FC). Within those populations, Tay-sachs occurs in 2% of the babies. The agent might represent this fact using the statement $\|\mathit{TS}(x)|\mathit{EEJ}(x) \vee \mathit{FC}(x)\|_x = 0.02$.

However, the agent would not be able to use this information in reasoning, since disjunctive reference classes are disallowed.

It is clear that if one takes the reference-class approach to generating degrees of belief, some restrictions on what constitutes a legitimate reference class are inevitable. Unfortunately, it seems that the current approaches to this problem are inadequate. The random-worlds approach does not depend on the notion of a reference class, and so is not forced to confront this issue.

2.1.3 Competing reference classes

Even if the problem of defining the set of "legitimate" reference classes can be resolved, the reference-class approach must still address the problem of choosing the "right" reference class out of the set of legitimate ones. The solution to this problem has typically been to posit a collection of rules indicating when one reference class should be preferred over another. The basic criterion is the one we already mentioned: choose the most specific reference class. But even in the cases to which this specificity rule applies, it is not always appropriate. Assume, for example, that we know that between 70% and 80% of birds chirp. We also know that between 0% and 99% of magpies chirp. If Tweety is a magpie, the specificity rule would tell us to use the more specific reference class, and conclude that $\Pr(\mathit{Chirps}(\mathit{Tweety})) \in [0, 0.99]$. Although the interval $[0, 0.99]$ is certainly not trivial, it is not very meaningful: had the 0.99 been a 1, it would have been trivial. We could then have ignored this reference class and used the far more detailed statistics of $[0.7, 0.8]$ derived from the class of birds.

The knowledge-base above might be appropriate for someone who knows little about magpies, and so feels less confidence in the statistics for magpies than he does for the class of birds as a whole. But since $[0.7, 0.8] \subseteq [0, 0.99]$, we know nothing that indicates that magpies are actually different from birds in general with respect to chirping. There is an alternative intuition that says that if the statistics for the less specific reference class (the class of birds) are more precise, and they do not contradict the statistics for the narrower class (magpies), then we should use them. That is, we should conclude that $\Pr(\mathit{Chirps}(\mathit{Tweety})) \in [0.7, 0.8]$. This intuition is captured and generalized in Kyburg's *strength rule*.

Unfortunately, neither the specificity rule nor its extension by Kyburg's strength rule are adequate in most cases. In typical examples, the agent has several incomparable classes relevant to the problem, so that neither rule applies. Reference-class systems such as Kyburg's and Pollock's simply give no useful answer in these cases. For example, suppose we know that Fred has high cholesterol and is a heavy smoker, and that 15% of people with high cholesterol get heart disease. If this is the only suitable reference class, then (according to all the systems) $\Pr(\mathit{Heart\text{-}disease}(\mathit{Fred})) = 0.15$. On the other hand, suppose we then acquire the additional information that 9% of heavy smokers develop heart disease (but still have no nontrivial statistical information about the intersection class of people with both attributes). In this case, neither class is the single right reference class, so approaches that rely on finding a single reference class generate a trivial degree of belief that Fred will contract heart disease in this case. For example, Kyburg's system will generate the interval $[0, 1]$ as the degree of belief.

Giving up completely in the face of conflicting evidence seems to us to be inappropriate. The entire enterprise of generating degrees of belief is geared to providing the agent with some guidance for its actions (in the form of degrees of belief) when deduction is insufficient to provide a definite answer. That is, the aim is to generate *plausible* inferences. The presence of conflicting information does not mean that the agent no longer needs guidance. When we have several competing reference classes, none of which dominates the others according to specificity or any other rule that has been proposed, then the degree of belief should most reasonably be some *combination* of the corresponding statistical values. As we show later, the random-worlds approach does indeed combine the values from conflicting reference classes in a reasonable way, giving well-motivated answers even when the reference-class approach would fail.

2.1.4 Other types of information

We have already pointed out the problems that arise with the reference-class approach if more than one reference class bears on a particular problem. A more subtle problem is encountered in cases where there is relevant information that is not in the form of a reference class. We have said that for $\psi(x)$ to be a reference class for a query about $\varphi(c)$ we must know $\psi(c)$ and have some statistical information about $||\varphi(x)|\psi(x)||_x$. However, it is not sufficient to consider only the query $\varphi(c)$. Suppose we also know $\varphi(c) \Leftrightarrow \sigma(c)$ for some other formula σ. Then we would want $\Pr(\varphi(c)) = \Pr(\sigma(c))$. But this implies that all of the reference classes for $\sigma(c)$ are relevant as well, because anything we can infer about $\Pr(\sigma(c))$ tells us something about $\Pr(\varphi(c))$. Both Pollock [Pol90] and Kyburg [Kyb83] deal with this considering all of the reference classes for any formula σ such that $\sigma(c) \Leftrightarrow \varphi(c)$ is known. However, they do not consider the case where it is known that $\sigma(c) \Rightarrow \varphi(c)$, which implies that $\Pr(\sigma(c)) \leq \Pr(\varphi(c))$, nor the case where it is known that $\varphi(c) \Rightarrow \sigma(c)$, which implies that $\Pr(\sigma(c)) \geq \Pr(\varphi(c))$. Thus, if we have a rich theory about $\varphi(c)$ and its implications, it can become very hard to locate all of the possible reference classes or even to define what qualifies as a possible reference class.

2.1.5 Discussion

A comparison between random worlds and reference-class approaches can be made in terms of the use of local versus global information. The reference-class approach is predicated on the assumption that there is always a local piece of information, i.e., the statistics over a single reference class, that captures all of the global information contained in the knowledge base. As is well known, local information cannot in general substitute for global information. So the difficulties encountered by the reference-class approach are not surprising. When generating degrees of belief from a rich knowledge base it will not always be possible to find a single reference class that captures all of the relevant information.

It is important to remember that although the notion of a reference class seems intuitive, it arises as part of one proposed *solution* strategy for the problem of computing degrees of belief. The notion of a reference classes is not part of the description of the problem, and there is no reason for it to necessarily be part of the solution. Indeed, as we have tried to argue, making it part of the solution leads to more problems than it solves.

Our approach, on the other hand, makes no attempt to locate a single local piece of information (a reference class). Thus, all of the problems described above that arise from trying locate the "right" reference class vanish. Rather, it uses a semantic construction that takes into account all of the information in the knowledge base in a uniform manner. As we shall see, the random-worlds approach generates answers that agree with the reference-class approach in those special cases where there is a single appropriate reference class. However, it continues to give reasonable answers in situations where no single local piece of information suffices. Furthermore, these answers are obtained directly from the simple semantics of random worlds, with no *ad hoc* rules and assumptions.

2.2 Default reasoning

One main claim of this thesis is that the random-worlds method of inference, coupled with our statistical interpretation of defaults, provides a well-motivated and successful system of default reasoning. Evaluating such a claim is hard because there are many, often rather vague, criteria for success that one can consider. In particular, not all criteria are appropriate for all default reasoning systems: Different applications (such as some of the ones outlined in [McC86]) require different interpretations for a default rule, and therefore need to satisfy varying desiderata. Nevertheless, there are certain properties that have gained acceptance as measures for the success of a new nonmonotonic reasoning system. While some of these properties are general ones (see Section 2.2.2), most research in the area has traditionally been driven by a small set of standard examples (more often than not involving a bird called "Tweety"). As we claim at the end of this section, this has made an "objective" validation of proposed systems difficult, to say the least. In this section, we survey some of the desired properties for default reasoning and the associated problems and issues. Of course, our survey cannot be comprehensive. The areas we consider are: the semantics of defaults, basic properties of default inference, inheritance and irrelevance, and expressive power.

2.2.1 Semantics of defaults

While it is possible to discuss the properties of an abstract default reasoning systems (see Section 2.2.2), the discussion of certain aspects of such systems requires us to identify a notion of *default rule*. In general, a default rule is an expression that has the form $A(x) \rightarrow B(x)$, whose intuitive interpretation is that if A holds for some individual x then typically (normally, usually, probably, ...) B holds for that individual.[2] While the syntax actually used differs significantly from case to case, most default reasoning systems have some construct of this

[2] We use $\rightarrow$ for a default implication, reserving $\Rightarrow$ for standard material implication.

type. For instance, in Reiter's *default logic* [Rei80] we would write

$$\frac{A(x)\ :\ B(x)}{B(x)}$$

while in a *circumscriptive* framework [McC80], we might use

$$\forall x\,(A(x) \wedge \neg Ab(x) \Rightarrow B(x))$$

while circumscribing $Ab(x)$. Theories based on first-order conditional logic [Del88] often do use the syntax $A(x) \rightarrow B(x)$. As we said in the introduction, in the random worlds framework, this default is captured using the statistical assertion $\|B(x)|A(x)\|_x \approx 1$.

While most systems of default inference have a notion of a default rule, not all of them address the issue of what the rule *means*. In particular, while all systems describe how a default rule should be used, some do not ascribe semantics (or ascribe only unintuitive semantics) to such rules. Without a good, intuitive semantics for defaults it becomes very difficult to judge the reasonableness of a collection of default statements. For example, as we mentioned above, one standard reading of $\varphi \rightarrow \psi$ is "φ's are typically ψ's". Under this reading, the pair of defaults $A \rightarrow B$ and $A \rightarrow \neg B$ should be inconsistent. In approaches such as Reiter's default logic, $A \rightarrow B$ and $A \rightarrow \neg B$ can be simultaneously adopted; they are not "contradictory" because there is no relevant notion of contradiction.

There are several ways to address this issue. The one we use is to find a single logic, with semantics, that covers both first-order information and default information. Such an approach enables us, among other things, to verify the consistency of a collection of defaults and to see whether a default follows logically from a collection of defaults. Of existing theories, those based on conditional and/or modal logic come closest to doing this.

2.2.2 Properties of default inference

As we said, default reasoning systems have typically been measured by testing them on a number of important examples. Recently, a few tools have been developed that improve upon this approach. Gabbay [Gab84] (and later Makinson [Mak89] and Kraus, Lehmann, and Magidor [KLM90]) introduced the idea of investigating the input/output relation of a default reasoning system, with respect to certain general properties that such an inference relation might possess. Makinson [Mak] gives a detailed survey of this work.

The idea is simple. Fix a theory of default reasoning and let *KB* be some knowledge base appropriate to this theory. Suppose φ is a default conclusion reached from *KB* according to the particular default approach being considered. In this case, we write $KB \mathrel{|\!\sim} \varphi$. The relation $\mathrel{|\!\sim}$ clearly depends on the default theory being considered. It is necessary to assume in this context that *KB* and φ are both expressed in the same logical language, that has a notion of valid implication. Thus, for example, if we are considering default logic or ϵ-semantics, we must assume that the defaults are fixed (and incorporated in the notion of $\mathrel{|\!\sim}$) and that both *KB* and φ are first-order or propositional formulas. Similarly, in the case of circumscription, the circumscriptive policy must also be fixed and incorporated into $\mid\!\sim$. (See also the discussion at the beginning of Section 2.2.3.)

With this machinery we can state a few desirable properties of default theories in a way that is independent of the (very diverse) details of such theories. There are five properties of $\mid\!\sim$ that have been viewed as being particularly desirable [KLM90]:

- *Right Weakening.* If $\varphi \Rightarrow \psi$ is logically valid and $KB \mid\!\sim \varphi$, then $KB \mid\!\sim \psi$.
- *Reflexivity.* $KB \mid\!\sim KB$.
- *Left Logical Equivalence.* If $KB \Leftrightarrow KB'$ is logically valid, then $KB \mid\!\sim \varphi$ if and only if $KB' \mid\!\sim \varphi$.
- *Cut.* If $KB \mid\!\sim \theta$ and $KB \wedge \theta \mid\!\sim \varphi$ then $KB \mid\!\sim \varphi$.
- *Cautious Monotonicity.* If $KB \mid\!\sim \theta$ and $KB \mid\!\sim \varphi$ then $KB \wedge \theta \mid\!\sim \varphi$.

While it is beyond the scope of this thesis to defend these criteria (see [KLM90]), we do want to stress *Cut* and *Cautious Monotonicity*, since they will be useful in our later results. They tell us that we can safely add to KB any conclusion θ that we can derive from KB, where "safely" is interpreted to mean that the set of conclusions derivable (via $\mid\!\sim$) from $KB \wedge \theta$ is precisely the same as that derivable from KB alone.

As shown in [KLM90], numerous other conditions can be derived from these properties. For example, we can prove:

- *And.* If $KB \mid\!\sim \varphi$ and $KB \mid\!\sim \psi$ then $KB \mid\!\sim \varphi \wedge \psi$.

Other plausible properties do not follow from these basic five. For example, the following property captures reasoning by cases:

- *Or.* If $KB \mid\!\sim \varphi$ and $KB' \mid\!\sim \varphi$, then $KB \vee KB' \mid\!\sim \varphi$.

Perhaps the most interesting property that does not follow from the basic five properties is what has been called *Rational Monotonicity* [KLM90]. Note that the property of (full) monotonicity, which we do not want, says that $KB \mid\!\sim \varphi$ implies $KB \wedge \theta \mid\!\sim \varphi$, no matter what θ is. It seems reasonable that default reasoning should satisfy the same property in those cases where where θ is "irrelevant" to the connection between KB and φ. While it is difficult to characterize "irrelevance", one situation where we may believe that θ should not affect the conclusions we can derive from KB is if θ is not implausible, given KB, i.e., if it is not the case that $KB \mid\!\sim \neg\theta$. The following property asserts that monotonicity holds when adding such a formula θ to our knowledge base:

- *Rational Monotonicity.* If $KB \mid\!\sim \varphi$ and it is not the case that $KB \mid\!\sim \neg\theta$, then $KB \wedge \theta \mid\!\sim \varphi$.

Several people, notably Lehmann and Magidor [LM92], have argued strongly for the desirability of this principle. One advantage of *Rational Monotonicity* is that it covers some fairly noncontroversial patterns of reasoning involving property inheritance. We explore this further in the next section.

The set of properties we have discussed provides a simple, but useful, system for classifying default theories. There are certainly applications in which some of the properties are inappropriate; Reiter's default logic is still popular even though it does not satisfy *Cautious Monotonicity*, *Or*, or *Rational Monotonicity*. (We briefly discuss one of the consequent disadvantages of default logic in the next section.) Nevertheless, many people would argue that these properties constitute a reasonable, if incomplete, set of desiderata for mainstream default theories.

2.2.3 Specificity and inheritance

As we have pointed out, systems of default reasoning have particular mechanisms for expressing default rules. A collection of such rules (perhaps in conjunction with other information) forms a default theory (or default knowledge base). For example, a particular default theory KB_{def} might contain the default "A's are typically B's"; we denote this by writing $[A(x) \rightarrow B(x)] \in KB_{def}$. A default theory KB_{def} is used by a default reasoning system in order to reason from various premises to default conclusions. For example, a theory KB_{def} containing the above default might infer $B(c)$ from $A(c)$. We write $\mid\!\sim_{def}$ to indicate the input/output relationship generated by a default reasoning system that uses KB_{def}. Thus, $A(c) \mid\!\sim_{def} B(c)$ indicates that some particular default reasoning system is able to conclude $B(c)$ from the premise $A(c)$ using the default theory KB_{def}.

Clearly, the presence of a default rule in its theory does not necessarily mean that a default reasoning system has the ability to apply that rule to a particular individual. Nevertheless, unless something special is known about that individual, the following seems to be an obvious requirement for any default reasoning system:

- *Direct Inference for Defaults.* If $[A(x) \rightarrow B(x)] \in KB_{def}$ and KB_{def} contains no assertions mentioning c, then $A(c) \mid\!\sim_{def} B(c)$.

This requirement has been previously discussed by Poole [Poo91], who called it the property of *Conditioning*. We have chosen a different name that relates the property more directly to earlier notions arising in direct inference.

We view *Direct Inference for Defaults* as stating a (very weak) condition for how a default theory should behave on simple problems involving the *inheritance* of properties from one class to another class or individual. Consider the following standard example, in which our default knowledge base KB_{fly} is

$$\begin{array}{l} Bird(x) \rightarrow Fly(x) \wedge \\ Penguin(x) \rightarrow \neg Fly(x) \wedge, \\ \forall x\,(Penguin(x) \Rightarrow Bird(x)). \end{array}$$

Should Tweety inherit the property of flying from the class of birds, or the property of not flying from the class of penguins? For any system satisfying *Direct Inference for Defaults* we must have $\mathit{Penguin}(\mathit{Tweety}) \mid\!\sim_{\mathit{fly}} \neg \mathit{Fly}(\mathit{Tweety})$. The reason is that, since we have a default whose precondition exactly matches the information we have about Tweety, this default *automatically* takes precedence over any other default. In particular, it takes precedence over defaults for more general classes, such as birds. This *specificity* property — the automatic preference for the most specific default — is of course the direct analogue of the preference for more specific subsets that we saw in the context of reference-class reasoning. It is perhaps the least controversial desideratum in default reasoning. As we have just seen, it is a direct consequence of *Direct Inference for Defaults*.

In approaches such as default reasoning or circumscription, the most obvious encoding of these defaults does not satisfy specificity, and hence does not satisfy *Direct Inference for Defaults*. Default logic and circumscription are sufficiently powerful for us to be able to arrange specificity. For example, in default logic, this can be done by means of *non-normal defaults* [RC81]. There is a cost to doing this, however: adding a default rule can require that all older default rules are reexamined, and possibly changed, to enforce the desired precedences. (Although see one possible solution in [Eth88].)

Direct Inference for Defaults is a weak principle, since in most interesting cases of inheritance there is no default that fits the case at hand perfectly. Suppose we learn that Tweety is a yellow penguin. Should we still conclude that Tweety does not fly? That is, should we conclude $\mathit{Penguin}(\mathit{Tweety}) \wedge \mathit{Yellow}(\mathit{Tweety}) \mid\!\sim_{\mathit{fly}} \neg \mathit{Fly}(\mathit{Tweety})$? Most people would say we should, because we have been given no reason to suspect that yellowness is relevant to flight. In other words, in the absence of more specific information about yellow penguins we should use the most specific superclass that we do have knowledge for, namely penguins. That is, a default reasoning system should allow the *inheritance* of the default for flight from the class of penguins to the class of yellow penguins, thus retaining the conclusion $\neg \mathit{Fly}(\mathit{Tweety})$ in the face of the extra information $\mathit{Yellow}(\mathit{Tweety})$. The *inheritance* property, i.e., the ability to solve such examples correctly, is a second, stronger criterion for successful default reasoning.

In some sense, we can view *Rational Monotonicity* as providing a partial solution to this problem [LM92]. If a nonmonotonic reasoning system is "reasonable" and satisfies *Rational Monotonicity* in addition to *Direct Inference for Defaults* then it does get inheritance in a large number of examples. In particular, as we have already observed, given $\mathit{KB}_{\mathit{fly}}$, we get $\mathit{Penguin}(\mathit{Tweety}) \mid\!\sim_{\mathit{fly}} \neg \mathit{Fly}(\mathit{Tweety})$ by *Direct Inference for Defaults*. Since $\mathit{KB}_{\mathit{fly}}$ gives us no reason to believe that yellow penguins are unusual, any "reasonable" default reasoning system would have $\mathit{Penguin}(\mathit{Tweety}) \not\mid\!\sim_{\mathit{fly}} \neg \mathit{Yellow}(\mathit{Tweety})$. From these two statements, *Rational Monotonicity* allows us to conclude $\mathit{Penguin}(\mathit{Tweety}) \wedge \mathit{Yellow}(\mathit{Tweety}) \mid\!\sim_{\mathit{fly}} \neg \mathit{Fly}(\mathit{Tweety})$, as desired.

However, *Rational Monotonicity* is still insufficient for inheritance reasoning in general. Suppose we add the default $\mathit{Bird}(x) \rightarrow \mathit{Warm\text{-}blooded}(x)$ to $\mathit{KB}_{\mathit{fly}}$. We would surely expect Tweety to be warm-blooded. However, *Rational Monotonicity* does not apply. To see why, observe that $\mathit{Bird}(\mathit{Tweety}) \mid\!\sim_{\mathit{fly}} \mathit{Warm\text{-}blooded}(\mathit{Tweety})$, while we want to conclude that $\mathit{Bird}(\mathit{Tweety}) \wedge \mathit{Penguin}(\mathit{Tweety}) \mid\!\sim_{\mathit{fly}} \mathit{Warm\text{-}blooded}(\mathit{Tweety})$. (Using *Left Logical Equivalence* we could then conclude from this latter statement that $\mathit{Penguin}(\mathit{Tweety}) \mathrel{|\!\!\sim}_{\mathit{fly}} \mathit{Warm\text{-}blooded}(\mathit{Tweety})$, as desired.) Clearly, we can use *Rational Monotonicity* to go from the first statement to the second, provided we could show that $\mathit{Bird}(\mathit{Tweety}) \not\mathrel{|\!\!\sim}_{\mathit{fly}} \neg \mathit{Penguin}(Tweety)$. However, many default reasoning systems do not support this statement. In fact, since penguins are exceptional birds that do not fly, it is not unreasonable to conclude by default that $\mathit{Bird}(Tweety) \mathrel{|\!\!\sim}_{\mathit{fly}} \neg \mathit{Penguin}(Tweety)$. Thus, *Rational Monotonicity* fails to allow us to conclude that *Tweety* the penguin is warm-blooded.

It seems undesirable that if a subclass is exceptional in any one respect, then inheritance of all other properties is blocked. However, it is possible to argue that this blocking of *inheritance to exceptional subclasses* is reasonable. Since penguins are known to be exceptional birds perhaps we should be cautious and not allow them to inherit *any* of the normal properties of birds. However, there are many examples which demonstrate that the complete blocking of inheritance to exceptional subclasses yields an inappropriately weak theory of default reasoning. For example, suppose we add to $\mathit{KB}_{\mathit{fly}}$ the default $\mathit{Yellow}(x) \rightarrow \mathit{Easy\text{-}to\text{-}see}(x)$. This differs from standard exceptional-subclass inheritance in that yellow penguins are not known to be exceptional members of the class of yellow things. That is, while penguins are known to be different from birds (and so perhaps the normal properties of birds should not be inherited), there is no reason to suppose that yellow penguins are different from other yellow objects. Nevertheless, *Rational Monotonicity* does not suffice even in this less controversial case. Indeed, there are well-known systems that satisfy *Rational Monotonicity* but cannot conclude that Tweety, the yellow penguin, is easy to see [LM92, Pea90]. This problem has been called the *drowning problem* [Ash93, BCD$^+$93].

Theories of default reasoning have had considerable difficulty in capturing an ability to inherit from superclasses that can deal properly with all of these different cases. In particular, the problem of inheritance to exceptional subclasses has been the most difficult. While some recent propositional theories have been more successful at dealing with exceptional subclass inheritance [GMP90, Gef92, GP92], they encounter other difficulties, discussed in the next section.

2.2.4 Expressivity

In the effort to discover basic techniques and principles for default reasoning, people have often looked at weak languages based on propositional logic. For instance, ϵ-semantics and variants [GP90, GMP90], modal approaches such as autoepistemic logic [Moo85], and conditional logics [Bou91], are usually considered in a propositional framework. Others, such as Reiter's default logic and Delgrande's conditional logic [Del88], use a first-order language, but with a syntax that tends to decouple the issues of first-order reasoning and default reasoning; we discuss this below. Of the better-known systems, circumscription seems to have the ability, at least in principle, of making the richest use of first-order logic.

It seems uncontroversial that, ultimately, a system of default reasoning should be built around a powerful language. Sophisticated knowledge representation systems almost invariably use languages with at least the expressive power of first-order logic. It is hard or impractical to encode the knowledge we have about almost any interesting domain without the expressive power provided by non-unary predicates and first-order quantifiers. It seems uncontroversial that, ultimately, a system of default reasoning should be built around a powerful language. We would also like to reason logically as well as by default within the same system, and to allow perhaps even richer languages. One of the advantages of our language is its ability to express first-order, statistical, and default information in one framework.

It is not easy to integrate first-order logic and defaults completely. One problem concerns "open" defaults, that are intended to apply to all individuals. For instance, suppose we wish to make a general statement that birds typically fly, and be able to use this when reasoning about different birds. Let us see how some existing systems do this.

In propositional approaches, the usual strategy is to claim that there are different types of knowledge (see, for example, [GP92] and the references therein). General defaults, such as $\mathit{Bird} \rightarrow \mathit{Fly}$, are in one class. When we reason about an individual, such as Tweety, its properties are described by knowledge in a different class, the *context*. For Tweety, the context might be $\mathit{Bird} \wedge \mathit{Yellow}$. In a sense, the symbol *Bird* stands for a general property when used in a default and talks about Tweety (say) when it appears in the context. First-order approaches have more expressive power in this regard. For example, Reiter's default logic uses defaults with free variables, e.g., $\mathit{Bird}(x) \rightarrow \mathit{Fly}(x)$. That Tweety flies can then be written $\mathit{Fly}(\mathit{Tweety})$, which seems much more natural. The default itself is treated essentially as a schema, implying all substitution instances (such as $\mathit{Bird}(\mathit{Tweety}) \Rightarrow \mathit{Fly}(\mathit{Tweety})$).

One example shows the problems with both these ideas. Suppose we know that:

> Elephants typically like zookeepers.
> Fred is a zookeeper, but elephants typically do not like Fred.

Using this information we can apply specificity to determine reasonable answers to such questions as "Does Clyde the Elephant like Fred?" (No) or "Does Clyde like Eric the Zookeeper?" (Yes). But the propositional strategy of classifying knowledge seems to fail here. Is "Elephants typically do not like Fred" a general default, or an item of contextual knowledge? Since it talks about elephants in general and also about one particular zookeeper, it does not fit either category well. In a rich first-order language, there is no clear-cut distinction between specific facts and general knowledge. So, while creating such a distinction can be used to avoid explicit first-order syntax, one loses generality.

Next, consider the first-order, substitutional, approach. It is easy to see that this does not work at all. One substitution instance of

$$\mathit{Elephant}(x) \wedge \mathit{Zookeeper}(y) \rightarrow \mathit{Likes}(x,y)$$

is

$$\mathit{Elephant}(x) \wedge \mathit{Zookeeper}(\mathit{Fred}) \rightarrow \mathit{Likes}(x,\mathit{Fred}),$$

which will contradict the second default. Of course, we could explicitly exclude Fred:

$$\mathit{Elephant}(x) \wedge \mathit{Zookeeper}(y) \wedge y \neq Fred \rightarrow \mathit{Likes}(x,y).$$

However, explicit exclusion is similar to the process of explicitly disabling less specific defaults, mentioned in the previous section. Both are examples of hand-coding the answers, and are therefore highly impractical for large knowledge bases.

The zookeeper example is similar to an example given by Lehmann and Magidor [LM90]. The solution they suggest to this problem fails to provide an explicit interpretation for open defaults. Rather, the meaning of an open default is implicitly determined by a set of rules provided for manipulating such defaults. These rules can cope with the zookeeper example, but the key step in the application of these rules is the use of *Rational Monotonicity.* More precisely, the above knowledge base, with the assumption that $\mathit{Likes}(\mathit{Clyde}, y)$, entails by default that $y \neq \mathit{Fred}$. However, it does not entail that $y \neq \mathit{Eric}$ for some zookeeper Eric. Therefore, *Rational Monotonicity* allows us to assume that $y = \mathit{Eric}$ and conclude that $\mathit{Likes}(\mathit{Clyde}, \mathit{Eric})$. We cannot use the same reasoning to conclude $\mathit{Likes}(\mathit{Clyde}, \mathit{Fred})$. But, we have seen the problem with *Rational Monotonicity* in Section 2.2.3 — it is easily blocked by "irrelevant" exceptionality. If Eric is known to be exceptional in some way (even one unrelated to zookeeping), then Lehmann and Magidor's approach will not be able conclude that he is liked by Clyde. This is surely undesirable.

Thus, it seems to be very hard to interpret generic (open) defaults properly. This is perhaps the best-known issue regarding the expressive power of various approaches to default logic. There are, of course, others; we close by mentioning one.

Morreau [Mor93] has discussed the usefulness of being able to refer to "the class of individuals satisfying a certain default". For example, the assertion:

> Typically, people who normally go to bed late normally rise late.

refers to "the class of people who normally go to bed late". The structure of this assertion is essentially:

$$(\mathit{Day}(y) \to \mathit{To\text{-}bed\text{-}late}(x, y)) \to (\mathit{Day}(y') \to \mathit{Rises\text{-}late}(x, y')).$$

This is a default whose precondition and conclusion are descriptions of people whose behaviors are themselves defined using defaults. Default logic, for example, cannot even express such defaults. Many theories of conditional logics (which can express generalizations like the example) do not give the correct answers [Del88, Bou91]. While this problem does not appear to have been investigated in a circumscriptive framework, it seems unlikely that any straightforward encoding of this default in this framework would behave appropriately. (Although, again, circumscription can be forced to give perhaps any answer with sufficient hand-coding.) We also note that the example has many variants. For instance, there is clearly a difference between the above default and the one "Typically, people who go to bed late rise late (i.e., the next morning)"; formally, the latter statement could be written:

$$(\mathit{Day}(y) \land \mathit{To\text{-}bed\text{-}late}(x, y)) \to (\mathit{Next\text{-}day}(y', y) \Rightarrow \mathit{Rises\text{-}late}(x, y')),$$

There are also other variations. We would like to express and reason correctly with them all. The real issue here is that we need to define various properties of individuals, and while many of these properties can be expressed in first-order logic, others need to refer to defaults explicitly. This argues, yet again, that it is a mistake to have a different language for defaults than the one used for other knowledge.

2.2.5 The lottery paradox

The *lottery paradox* [Kyb61] addresses the issue of how different default conclusions interact. It provides a challenging test of the intuitions and semantics of any default reasoning system. There are a number of variants of this paradox; we consider three here.

First, imagine that a large number N of people buy tickets to a lottery in which there is only one winner. For a particular person c, it seems sensible to conclude by default that c does not win the lottery. But we can argue this way for any individual, which seems to contradict the fact that someone definitely will win. Of course some theories, such as those based on propositional languages, do not have enough expressive power to even state this version of this problem. Among theories that can state it, there would seem to be several options. Clearly, one solution is to deny that default conclusions are closed under arbitrary conjunction, i.e., to give up on the *And Rule*. But aside from explicitly probabilistic theories, we are not aware of work taking this approach (although the existence of multiple extensions in theories such as Reiter's is certainly related). Without logical closure, there is a danger of being too dependent on merely syntactic features of a problem. Another solution is to prevent a theory from reasoning about all N individuals at once. Finally, one can simply deny that $\neg \mathit{Winner}(c)$ follows by default. Circumscription, for instance, does this: The standard representation of the problem would result in multiple extensions, such that for each individual c, there is one extension where c is the winner. While this seems reasonable, circumscription only allows us to conclude things that hold in all extensions; thus, we would not be able to conclude $\neg \mathit{Winner}(c)$. The problem with these "solutions" is that the lottery problem seems like an extremely reasonable application of default reasoning: if you buy a lottery ticket you *should* continue your life under the assumption that you will not win.

The lottery paradox is also a suitable setting to discuss an issue raised by Lifschitz's list of benchmark problems [Lif89]. Suppose we have a default, for instance $\mathit{Ticket}(x) \rightarrow \neg \mathit{Winner}(x)$, and no other knowledge. Should $\forall x\ (\mathit{Ticket}(x) \Rightarrow \neg \mathit{Winner}(x))$ be a default conclusion? Likewise, if we know $\mathit{Winner}(c)$ but consider it possible that the lottery has more than one winner, should we nevertheless conclude that $\forall x\ ((\mathit{Ticket}(x) \wedge x \neq c) \Rightarrow \neg \mathit{Winner}(x))$? In circumscription, although not in many other theories, we get both universal conclusions (as Lifschitz argues for). The desire for these universal conclusions is certainly controversial; in fact it seems that we often *expect* default rules to have some exceptions. However, as Lifschitz observes, there is now a technical difficulty: How can we conclude from the default $\mathit{Ticket}(x) \rightarrow \neg \mathit{Winner}(x)$ that, by default, each individual c is not a winner, and yet not make the universal conclusion that, by default, no one wins? Because of its treatment of open defaults, Reiter's default logic is able to obtain precisely this conclusion. As we shall see, the random-worlds approach obtains this conclusion as well.

Poole [Poo91] has considered a variant of the lottery paradox that avoids entirely the issue of named individuals. In his version, there is a formula describing the types of birds we are likely to encounter, such as:

$$\forall x(\mathit{Bird}(x) \Leftrightarrow (\mathit{Emu}(x) \vee \mathit{Penguin}(x) \vee \ldots \vee \mathit{Canary}(x))).$$

We then add to the knowledge base defaults such as birds typically fly, but penguins typically do not fly, and we similarly assert that every other species of bird is exceptional in some way. Now suppose all we know is that $\mathit{Bird}(\mathit{Tweety})$. Can we conclude that Tweety flies? If we conclude that he, then a similar argument would also allow us to conclude that he is a typical bird in all other respects. But this would contradict the fact he must be exceptional in some respect. If we do not conclude that Tweety flies, then the default "Birds typically fly" has been effectively ignored. Poole uses such examples to give an exhaustive analysis of how various systems might react to the Lottery Paradox. He shows that in any theory, some desideratum, such as closure under conjunction or "conditioning" (which is essentially *Direct inference for defaults*), must be sacrificed. Perhaps the most interesting "way out" he discusses is the possibility of declaring that certain combinations of defaults are inadmissible or inconsistent. Is it really reasonable to say that the class of birds is the union of subclasses all of which are exceptional? In many theories, such as Reiter's default logic, there is nothing to prevent one from asserting this. But in a theory which gives reasonable semantics to defaults, we may be able to determine and justify the incompatibility of certain sets of defaults. This, indeed, is how our approach avoids Poole's version of the lottery paradox.

2.2.6 Discussion

In this section, we have presented a limited list of desiderata that seem appropriate for a default reasoning system. While this list may be limited, it is interesting to point out that there does not seem to be a single default reasoning system that fulfills all these desiderata in a satisfactory way.

Unfortunately, to the best of our knowledge, there is (as yet) no general framework for evaluating default reasoning systems. In particular, the evaluation still tends to be on the level of "Does this theory solve these particular examples correctly?" (see, for example, the list of benchmark problems in [Lif89]). While such examples are often important in identifying interesting aspects of the problem and defining our intuitions in these cases, they are clearly not a substitute for a comprehensive framework. Had there been such a framework, perhaps the drowning problem from Section 2.2.3 would not have remained undiscovered for so long. While we do not attempt to provide such a general framework in this thesis, in Chapter 4 we prove a number of general theorems concerning the random-worlds approach. These theorems provide a precise formulation of properties such as *Direct Inference for Defaults*, and show that they hold for random worlds. The fact that we also get properties such as specificity and exceptional subclass inheritance follows immediately from these theorems. Thus, our proof that the random-worlds approach deals well with the paradigm examples in default reasoning is part of a comprehensive theorem, rather than via a case-by-case analysis.

Chapter 3

The formalism

3.1 The language

We are interested in a formal logical language that allows us to express both statistical information and first-order information. We therefore define a statistical language $\mathcal{L}^{\approx}$, which is a variant of a language designed by Bacchus [Bac90]. For the remainder of the thesis, let Φ be a finite first-order vocabulary, consisting of predicate and constant symbols, and let $\mathcal{X}$ be a set of variables.[1]

Our statistical language augments standard first-order logic with a form of statistical quantifier. For a formula $\psi(x)$, the term $||\psi(x)||_x$ is a *proportion expression*. It will be interpreted as a rational number between 0 and 1, that represents the proportion of domain elements satisfying $\psi(x)$. We actually allow an arbitrary set of variables in the subscript and in the formula ψ. Thus, for example, $||\mathit{Child}(x,y)||_x$ describes, for a fixed y, the proportion of domain elements that are children of y; $||\mathit{Child}(x,y)||_y$ describes, for a fixed x, the proportion of domain elements whose child is x; and $||\mathit{Child}(x,y)||_{x,y}$ describes the proportion of pairs of domain elements that are in the child relation.[2]

We also allow proportion expressions of the form $||\psi(x)|\theta(x)||_x$, which we call *conditional proportion expressions*. Such an expression is intended to denote the proportion of domain elements satisfying ψ from among those elements satisfying θ. Finally, any rational number is also considered to be a proportion expression, and the set of proportion expressions is closed under addition and multiplication.

One important difference between our syntax and that of [Bac90] is the use of *approximate equality* to compare proportion expressions. As we argued in the introduction, exact comparisons are sometimes inappropriate. Consider a statement such as "80% of patients with

[1] For simplicity, we assume that Φ does not contain function symbols, since these can be defined in terms of predicates.

[2] Strictly speaking, these proportion expression should be written with sets of variables in the subscript, as in $||\mathit{Child}(x,y)||_{\{x,y\}}$. However, when the interpretation is clear, we often abuse notation and drop the set delimiters.

jaundice have hepatitis". If this statement appears in a knowledge base, it is almost certainly there as a summary of a large pool of data. It is clear that we do not mean that *exactly* 80% of all patients with jaundice have hepatitis. Among other things, this would imply that the number of jaundiced patients is a multiple of five, which is surely not an intended implication. We therefore use the approach described in [GHK92, KH92], and compare proportion expressions using (instead of $=$ and $\leq$) one of an infinite family of connectives $\approx_i$ and $\preceq_i$, for $i = 1, 2, 3 \ldots$ ("i-approximately equal" or "i-approximately less than or equal").[3] For example, we can express the statement "80% of jaundiced patients have hepatitis" by the *proportion formula* $||Hep(x)|Jaun(x)||_x \approx_1 0.8$. The intuition behind the semantics of approximate equality is that each comparison should be interpreted using some small tolerance factor to account for measurement error, sample variations, and so on. The appropriate tolerance will differ for various pieces of information, so our logic allows different subscripts on the "approximately equals" connectives. A formula such as $||Fly(x)|Bird(x)||_x \approx_1 1 \wedge ||Fly(x)|Bat(x)||_x \approx_2 1$ says that both $||Fly(x)|Bird(x)||_x$ and $||Fly(x)|Bat(x)||_x$ are approximately 1, but the notion of "approximately" may be different in each case.

We can now give a recursive definition of the language $\mathcal{L}^{\approx}$.

Definition 3.1.1: The set of *terms* in $\mathcal{L}^{\approx}$ is $\mathcal{X} \cup \mathcal{C}$ where $\mathcal{C}$ is the set of constant symbols in Φ. The set of *proportion expressions* is the least set that

(a) contains the rational numbers,

(b) contains *proportion terms* of the form $||\psi||_X$ and $||\psi|\theta||_X$, for formulas $\psi, \theta \in \mathcal{L}^{\approx}$ and a finite set of variables $X \subseteq \mathcal{X}$, and

(c) is closed under addition and multiplication.

The set of formulas in $\mathcal{L}^{\approx}$ is the least set that

(a) contains *atomic formulas* of the form $R(t_1, \ldots, t_r)$, where R is a predicate symbol in $\Phi \cup \{=\}$ of arity r and $t_1, \ldots, t_r$ are terms,

(b) contains *proportion formulas* of the form $\zeta \approx_i \zeta'$ and $\zeta \preceq_i \zeta'$, where ζ and ζ' are proportion expressions and i is a natural number, and

(c) is closed under conjunction, negation, and first-order quantification. ∎

Notice that this definition allows arbitrary nesting of quantifiers and proportion expressions. In Section 3.3 we demonstrate the expressive power of the language. As observed in [Bac90], the subscript x in a proportion expressions binds the variable x in the expression; indeed, we can view $||\cdot||_x$ as a new type of quantification.

We now need to define the semantics of the logic. As we shall see below, most of the definitions are fairly straightforward. The two features that cause problems are approximate

[3] In [BGHK92] the use of approximate equality was suppressed in order to highlight other issues.

comparisons and conditional proportion expressions. We interpret the approximate connective $\zeta \approx_i \zeta'$ to mean that ζ is very close to ζ'. More precisely, it is within some very small, but unknown, tolerance factor. We formalize this using a *tolerance vector* $\vec{\tau} = \langle \tau_1, \tau_2, \ldots \rangle$, $\tau_i > 0$. Intuitively $\zeta \approx_i \zeta'$ if the values of ζ and ζ' are within τ_i of each other.

A difficulty arises when interpreting conditional proportion expressions because we need to deal with the problem of conditioning on an event of measure 0. That is, we need to define semantics for $\|\psi|\theta\|_X$ even when there are no assignments to the variables in X that would satisfy θ. When standard equality is used rather than approximate equality, this problem is easily overcome. Following [Hal90], we can eliminate conditional proportion expressions altogether by viewing a statement such as $\|\psi|\theta\|_X = \alpha$ as an abbreviation for $||\psi \wedge \theta||_X = \alpha||\theta||_X$. This approach agrees with the standard interpretation of conditionals if $||\theta||_X \neq 0$. If $||\theta||_X = 0$, it enforces the convention that formulas such as $\|\psi|\theta\|_X = \alpha$ or $\|\psi|\theta\|_X \leq \alpha$ are true for any α. We used the same approach in [GHK92], where we allowed approximate equality. Unfortunately, as the following example shows, this interpretation of conditional proportions can interact in an undesirable way with the semantics for approximate comparisons. In particular, this approach does not preserve the standard semantics of conditional equality if $||\theta||_X$ is *approximately* 0.

Example 3.1.2: Consider the knowledge base:[4]

$$KB = (||Penguin(x)||_x \approx_1 0) \wedge (||Fly(x)|Penguin(x)||_x \approx_2 0).$$

We expect this to mean that the proportion of penguins is very small (arbitrarily close to 0 in large domains), but also that the proportion of fliers among penguins is also very small. However, if we interpret conditional proportions as discussed above, we obtain the knowledge base

$$KB' = (||Penguin(x)||_x \approx_1 0) \wedge (||Fly(x) \wedge Penguin(x)||_x \approx_2 0 \cdot ||Penguin(x)||_x).$$

The knowledge base KB' is equivalent to

$$(||Penguin(x)||_x \approx_1 0) \wedge (||Fly(x) \wedge Penguin(x)||_x \approx_2 0).$$

This simply says that that the proportion of penguins and the proportion of flying penguins are both small, but says nothing about the proportion of fliers among penguins. In fact, the world where all penguins fly is consistent with KB'. Clearly, the process of multiplying out across an approximate connective does not preserve the intended interpretation of the formulas. ∎

Because of this problem, we cannot treat conditional proportions as abbreviations and instead have added them as primitive expressions in the language. Of course, we now have to give them a semantics that avoids the problem illustrated by Example 3.1.2. We would like to maintain the conventions used when we had equality in the language. Namely, in worlds where

[4] We remark that, here and in our examples below, the actual choice of subscript for $\approx$ is unimportant. However, we use different subscripts for different approximate comparisons unless the tolerances for the different measurements are known to be the same.

$||\theta(x)||_x \neq 0$, we want $\|\varphi(x)|\theta(x)\|_x$ to denote the fraction of elements satisfying $\theta(x)$ that also satisfy $\varphi(x)$. In worlds where $||\theta(x)||_x = 0$, we want formulas of the form $\|\varphi(x)|\theta(x)\|_x \approx_i \alpha$ or $\|\varphi(x)|\theta(x)\|_x \preceq_i \alpha$ to be true. There are a number of ways of accomplishing this. The way we take is perhaps not the simplest, but it introduces machinery that will be helpful later.

We give semantics to the language $\mathcal{L}^{\approx}$ by providing a translation from formulas in $\mathcal{L}^{\approx}$ to formulas in a language $\mathcal{L}^{=}$ whose semantics is more easily described. The language $\mathcal{L}^{=}$ is essentially the language of [Hal90], that uses true equality rather than approximate equality. More precisely, the definition of $\mathcal{L}^{=}$ is identical to the definition of $\mathcal{L}^{\approx}$ given in Definition 3.1.1, except that:

- we use $=$ and $\leq$ instead of $\approx_i$ and $\preceq_i$,
- we allow the set of proportion expressions to include arbitrary real numbers (not just rational numbers),
- we do not allow conditional proportion expressions,
- we assume that $\mathcal{L}^{=}$ has a special family of variables ε_i, interpreted over the reals.

As we shall see, the variable ε_i is used to interpret the approximate equality connectives $\approx_i$ and $\preceq_i$. We view an expression in $\mathcal{L}^{=}$ that uses conditional proportion expressions as an abbreviation for the expression obtained by multiplying out.

The semantics for $\mathcal{L}^{=}$ is quite straightforward, and follows the lines of [Hal90]. Recall that we give semantics to $\mathcal{L}^{=}$ in terms of *worlds*, or finite first-order models. For any natural number N, let $\mathcal{W}_N$ consist of all worlds with domain $\{1, \ldots, N\}$, and let $\mathcal{W}^*$ denote $\cup_N \mathcal{W}_N$.

Now, consider some world $W \in \mathcal{W}^*$ over the domain D, some valuation $V : \mathcal{X} \to \{1, \ldots, N\}$ for the variables in $\mathcal{X}$, and some tolerance vector $\vec{\tau}$. We simultaneously assign to each proportion expression ζ a real number $[\zeta]_{(W,V,\vec{\tau})}$ and to each formula ξ a truth value with respect to $(W, V, \vec{\tau})$. Most of the clauses of the definition are completely standard, so we omit them here. In particular, variables are interpreted using V, the tolerance variables ε_i are interpreted using the tolerances τ_i, the predicates and constants are interpreted using W, the Boolean connectives and the first-order quantifiers are defined in the standard fashion, and when interpreting proportion expressions, the real numbers, addition, multiplication, and $\leq$ are given their standard meaning. It remains to interpret proportion terms. Recall that we eliminate conditional proportion terms by multiplying out, so that we need to deal only with unconditional proportion terms. If ζ is the proportion expression $||\psi||_{x_{i_1},\ldots,x_{i_k}}$ (for $i_1 < i_2 < \ldots < i_k$), then

$$[\zeta]_{(W,V,\vec{\tau})} = \frac{1}{|D|^k}\Big|\Big\{(d_1, \ldots, d_k) \in D^k \ : \ (W, V[x_{i_1}/d_1, \ldots, x_{i_k}/d_k], \vec{\tau}) \models \psi\Big\}\Big|.$$

Thus, if $W \in \mathcal{W}_N$, the proportion expression $||\psi||_{x_{i_1},\ldots,x_{i_k}}$ denotes the fraction of the N^k k-tuples of elements in D that satisfy ψ. For example, $[||\mathit{Child}(x, y)||_x]_{(W,V,\vec{\tau})}$ is the fraction of domain elements d that are children of $V(y)$.

We now show how a formula $\chi \in \mathcal{L}^{\approx}$ can be associated with a formula $\chi^* \in \mathcal{L}^{=}$. We proceed as follows:

- every proportion formula $\zeta \preceq_i \zeta'$ in χ is (recursively) replaced by $\zeta \Leftrightarrow \zeta' \leq \varepsilon_i$,
- every proportion formula $\zeta \approx_i \zeta'$ in χ is (recursively) replaced by the conjunction $(\zeta \Leftrightarrow \zeta' \leq \varepsilon_i) \wedge (\zeta' \Leftrightarrow \zeta \leq \varepsilon_i)$,
- finally, conditional proportion expressions are eliminated as in Halpern's semantics, by multiplying out.

This translation allows us to embed $\mathcal{L}^{\approx}$ into $\mathcal{L}^{=}$. Thus, for the remainder of the thesis, we regard $\mathcal{L}^{\approx}$ is a sublanguage of $\mathcal{L}^{=}$. We can now easily define the semantics of formulas in $\mathcal{L}^{\approx}$: For $\chi \in \mathcal{L}^{\approx}$, we say that $(W, V, \vec{\tau}) \models \chi$ iff $(W, V, \vec{\tau}) \models \chi^*$. It is sometimes useful in our future results to incorporate particular values for the tolerances into the formula χ^*. Thus, let $\chi[\vec{\tau}]$ represent the formula that results from χ^* if each variable ε_i is replaced with its value according to $\vec{\tau}$ — τ_i.[5]

Typically we are interested in closed sentences, that is, formulas with no free variables. In that case, it is not hard to show that the valuation plays no role. Thus, if χ is closed, we write $(W, \vec{\tau}) \models \chi$ rather than $(W, V, \vec{\tau}) \models \chi$.

3.2 Degrees of belief

As we explained in the introduction, we give semantics to degrees of belief by considering all worlds of size N to be equally likely, conditioning on KB, and then checking the probability of φ over the resulting probability distribution. In the previous section, we defined what it means for a sentence χ to be satisfied in a world of size N using a tolerance vector $\vec{\tau}$. Given N and $\vec{\tau}$, we define $\#worlds^{\vec{\tau}}_N(\chi)$ to be the number of worlds in $\mathcal{W}_N$ such that $(W, \vec{\tau}) \models \chi$. Since we are taking all worlds to be equally likely, the degree of belief in φ given KB with respect to $\mathcal{W}_N$ and $\vec{\tau}$ is

$$\Pr{}^{\vec{\tau}}_N(\varphi|KB) = \frac{\#worlds^{\vec{\tau}}_N(\varphi \wedge KB)}{\#worlds^{\vec{\tau}}_N(KB)}.$$

If $\#worlds^{\vec{\tau}}_N(KB) = 0$, this degree of belief is not well-defined.

Strictly speaking, we should write $\mathcal{W}^{\Phi}_N$ rather than $\mathcal{W}_N$, since the set of worlds under consideration clearly depends on the vocabulary. Similarly, the number of worlds in $\mathcal{W}_N$ depends on the vocabulary. Thus, both $\#worlds^{\vec{\tau}}_N(\varphi)$ and $\#worlds^{\vec{\tau}}_N(\varphi \wedge KB)$ depend on the choice of Φ. Fortunately, this dependence "cancels out" when defining the probability $\Pr^{\vec{\tau}}_N(\varphi|KB)$:

Proposition 3.2.1: *Let Φ, Φ' be finite vocabularies, and let φ, θ be sentences in both $\mathcal{L}(\Phi)$ and $\mathcal{L}(\Phi')$. Then*

$$\frac{\#worlds^{\Phi}_N(\varphi \wedge KB)}{\#worlds^{\Phi}_N(KB)} = \frac{\#worlds^{\Phi'}_N(\varphi \wedge KB)}{\#worlds^{\Phi'}_N(KB)},$$

where $\#worlds^{\Phi,\vec{\tau}}(\chi)$ denotes the number of worlds W in $\mathcal{W}^{\Phi}_N$ such that $(W, \vec{\tau}) \models \chi$.

[5] Note that some of the tolerances τ_i may be irrational; it is for this reason that we allowed irrational numbers in proportion expressions in $\mathcal{L}^{=}$.

Proof: We first prove the claim for the case $\Phi' = \Phi \cup \{R\}$ for some symbol $R \notin \Phi$. Let $\xi \in \mathcal{L}(\Phi)$ be an arbitrary sentence. A world over Φ' determines the denotations of the symbols in Φ, and the denotation of R. Let r be the number of possible denotations of R over a domain of size N. Since ξ does not mention R, it is easy to see that each model of ξ over Φ corresponds to r models of ξ over Φ', one for each possible denotation of R. Therefore, $\#\mathit{worlds}_N^{\Phi'}(\xi) = r \cdot \#\mathit{worlds}_N^{\Phi}(\xi)$. From this, the claim follows immediately.

Now, given arbitrary Φ and Φ', a straightforward induction on the cardinality of $\Phi' \Leftrightarrow \Phi$

$$\frac{\#\mathit{worlds}_N^{\Phi}(\varphi \wedge \mathit{KB})}{\#\mathit{worlds}_N^{\Phi}(\mathit{KB})} = \frac{\#\mathit{worlds}_N^{\Phi \cup \Phi'}(\varphi \wedge \mathit{KB})}{\#\mathit{worlds}_N^{\Phi \cup \Phi'}(\mathit{KB})}.$$

A similar argument for $\frac{\#\mathit{worlds}_N^{\Phi'}(\varphi \wedge \mathit{KB})}{\#\mathit{worlds}_N^{\Phi'}(\mathit{KB})}$ proves the result. ∎

Since for most of our discussion we are interested the case of a fixed finite vocabulary, we will eliminate the dependence on Φ in $\#\mathit{worlds}_N^{\vec{\tau}}(\chi)$.

Typically, we know neither N nor $\vec{\tau}$ exactly. All we know is that N is "large" and that $\vec{\tau}$ is "small". Thus, we would like to take our *degree of belief* in φ given KB to be $\lim_{\vec{\tau} \to \vec{0}} \lim_{N \to \infty} \Pr_N^{\vec{\tau}}(\varphi | \mathit{KB})$. Notice that the order of the two limits over $\vec{\tau}$ and N is important. If the limit $\lim_{\vec{\tau} \to \vec{0}}$ appeared last, then we would gain nothing by using approximate equality, since the result would be equivalent to treating approximate equality as exact equality.

This definition, however, is not sufficient; the limit may not exist. We observed above that $\Pr_N^{\vec{\tau}}(\varphi | \mathit{KB})$ is not always well-defined. In particular, it may be the case that for certain values of $\vec{\tau}$, $\Pr_N^{\vec{\tau}}(\varphi | \mathit{KB})$ is not well-defined for arbitrarily large N. In order to deal with this problem of well-definedness, we define KB to be *eventually consistent* if for all sufficiently small $\vec{\tau}$ and sufficiently large N, $\#\mathit{worlds}_N^{\vec{\tau}}(\mathit{KB}) > 0$. Among other things, eventual consistency implies that the KB is satisfiable in finite domains of arbitrarily large size. For example, a KB stating that "there are exactly 7 domain elements" is not eventually consistent. For most of this thesis, we assume that all knowledge bases are eventually consistent.

Even if KB is eventually consistent, the limit may not exist. For example, it may be the case that $\Pr_N^{\vec{\tau}}(\varphi | \mathit{KB})$ oscillates between $\alpha + \tau_i$ and $\alpha \Leftrightarrow \tau_i$ for some i as N gets large. In this case, for any particular $\vec{\tau}$, the limit as N grows will not exist. However, it seems as if the limit as $\vec{\tau}$ grows small "should", in this case, be α, since the oscillations about α go to 0. We avoid such problems by considering the *lim sup* and *lim inf*, rather than the limit. For any set $S \subset I\!R$, the infimum of S, $\inf S$, is the greatest lower bound of S. The *lim inf* of a sequence is the limit of the infimums; that is,

$$\liminf_{N \to \infty} a_N = \lim_{N \to \infty} \inf\{a_i : i > N\}.$$

The lim inf exists for any sequence bounded from below, even if the limit does not. The *lim sup* is defined analogously, where $\sup S$ denotes the least upper bound of S. If $\lim_{N \to \infty} a_N$ does exist, then $\lim_{N \to \infty} a_N = \liminf_{N \to \infty} a_N = \limsup_{N \to \infty} a_N$. Since, for any $\vec{\tau}$, the sequence $\Pr_N^{\vec{\tau}}(\varphi | \mathit{KB})$ is always bounded from above and below, the lim sup and lim inf always exist. Thus, we do not have to worry about the problem of nonexistence for particular values of $\vec{\tau}$. We can now present the final form of our definition.

Definition 3.2.2: If

$$\lim_{\vec{\tau}\to\vec{0}} \liminf_{N\to\infty} \Pr_N^{\vec{\tau}}(\varphi|KB) \quad \text{and} \quad \lim_{\vec{\tau}\to\vec{0}} \limsup_{N\to\infty} \Pr_N^{\vec{\tau}}(\varphi|KB)$$

both exist and are equal, then the *degree of belief in φ given KB*, written $\Pr_\infty(\varphi|KB)$, is defined as the common limit; otherwise $\Pr_\infty(\varphi|KB)$ does not exist.

We point out that, even using this definition, there are many cases where the degree of belief does not exist. However, as some of our examples show, in many situations the nonexistence of a degree of belief can be understood intuitively, and is sometimes related to the existence of multiple extensions of a default theory (see Sections 3.3 and 4.3).

We remark that Shastri [Sha89] used a somewhat similar approach to defining degrees of belief. His language does not allow us to talk about statistical information directly, but does allow us to talk about the number of domain individuals that satisfy a given predicate. He then gives a definition of degree of belief similar to ours. Since he has no notion of approximate equality in his language, and presumes a fixed domain size, he does not have to deal with limits as we do.

3.3 Statistical interpretation for defaults

As we mentioned in the introduction, there are many similarities between direct inference from statistical information and default reasoning. In order to capitalize on this observation, and use random worlds as a default reasoning system, we need to interpret defaults as statistical statements. However, finding the appropriate statistical interpretation is not straightforward. For example, as is well known, if we interpret "Birds typically fly" as "Most (i.e., more than 50% of) birds fly", then we get a default system that fails to satisfy some of the most basic desiderata, such as the *And* rule, discussed in Section 2.2.2. Using a higher fixed threshold in a straightforward way does not help. More successfully, Adams [Ada75], and later Geffner and Pearl [GP90], suggested an interpretation of defaults based on "almost all". In their framework, this is done using *extreme probabilities* — conditional probabilities that are arbitrarily close to 1: i.e., within $1 \Leftrightarrow \epsilon$ for some ϵ, and considering the limit as $\epsilon \to 0$. The basic system derived from this idea is called *ϵ-semantics.* Later, stronger systems (ones able to make more inferences) based on the same probabilistic idea were introduced (see Pearl [Pea89] for a survey).

The intuition behind ϵ-semantics and its extensions is statistical. However, since the language used in these approaches is propositional, this intuition cannot be expressed directly. Indeed, these approaches typically make no distinction between the statistical nature of the default and the degree of belief nature of the default conclusion. Nevertheless, there is a sense in which we can view our approach as the generalization of one of the extensions of ϵ-semantics, namely the maximum-entropy approach of Goldszmidt, Morris, and Pearl [GMP90], to the first-order setting. This issue is discussed in more detail in Section 7.2.3, where it is shown that this maximum-entropy approach can be embedded in our framework.

Of course, the fact that our syntax is so rich allows us to express a great deal of information that simply cannot be expressed in any propositional approach. We observed earlier that a propositional approach that distinguishes between default knowledge and contextual knowledge has difficulty in dealing with the elephant-zookeeper example (see Section 2.2.4). This example is easily dealt with in our framework.

Example 3.3.1: The following knowledge base, $\mathit{KB}_{\mathit{likes}}$, is a formalization of the elephant-zookeeper example. Recall, this problem concerns the defaults that "Elephants typically like zookeepers", but "Elephants typically do not like Fred". As discussed earlier, simply expressing this knowledge sensibly can be a considerable challenge. We have no problems, however:

$$\begin{array}{l} \|\mathit{Likes}(x,y)|\mathit{Elephant}(x) \land \mathit{Zookeeper}(y)\|_{x,y} \approx_1 1 \land \\ \|\mathit{Likes}(x,\mathit{Fred})|\mathit{Elephant}(x)\|_x \approx_2 0 \land \\ \mathit{Zookeeper}(\mathit{Fred}). \quad \blacksquare \end{array}$$

Furthermore, our interpretation of defaults allows us to deal well with interactions between first-order quantifiers and defaults.

Example 3.3.2: We may know that people who have at least one tall parent tend to be tall. This default can easily be expressed in our language:

$$\|\mathit{Tall}(x)|\exists y\,(\mathit{Child}(x,y) \land \mathit{Tall}(y))\|_x \approx_i 1. \quad \blacksquare$$

We can even define defaults over classes themselves defined using default rules (as described by Morreau [Mor93]).

Example 3.3.3: In Section 2.2.4, we describe the problems associated with defining the nested default "Typically, people who normally go to bed late normally rise late." To express this default in our formalism, we simply use nested proportion statements: The individuals who normally rise late are the ones who rise late most days. Thus, they are the x's satisfying $\|\mathit{Rises\text{-}late}(x,y)|\mathit{Day}(y)\|_y \approx_1 1$; similarly, the individuals who go to bed late are the x's satisfying $\|\mathit{To\text{-}bed\text{-}late}(x,y')|\mathit{Day}(y')\|_{y'} \approx_2 1$. Interpreting "typically" as "almost all", we can capture the default by saying most x's that go to bed late also rise late. That is, using the knowledge base $\mathit{KB}_{\mathit{late}}$:

$$\|\ (\|\mathit{Rises\text{-}late}(x,y)|\mathit{Day}(y)\|_y \approx_1 1)\ |\ (\|\mathit{To\text{-}bed\text{-}late}(x,y')|\mathit{Day}(y')\|_{y'} \approx_2 1)\ \|_x \approx_3 1.$$

On the other hand, the related default that "Typically, people who go to bed late rise late (i.e., the next morning)" can be expressed as:

$$\|\ \forall y'\,(\mathit{Next\text{-}day}(y',y) \Rightarrow \mathit{Rises\text{-}late}(x,y'))\ |\mathit{Day}(y) \land \mathit{To\text{-}bed\text{-}late}(x,y)\|_{x,y} \approx_1 1,$$

which is clearly different from the first default. $\blacksquare$

Chapter 4

Properties of random worlds

We now show that the random-worlds method validates several desirable reasoning patterns, including essentially all of those discussed in Sections 2.1 and 2.2. As we have seen, this success contrasts with many other theories of reference-class and default reasoning. Furthermore, all properties we validate are derived theorems, rather than being deliberately built into the reasoning process. We also note that all the results in this section hold for our language in its full generality: the formulas can contain arbitrary non-unary predicates, and have nested quantifiers and proportion statements. Finally, we note that the theorems we state are not the most general ones possible. It is quite easy to construct examples for which the conditions of the theorems do not hold, but random worlds still gives the intuitively plausible answer. Thus, we can clearly find other theorems that deal with additional cases. The main difficulty in doing this is in finding conditions that are easy to state and check, and yet cover an interestingly large class of examples. The problems in doing do are a direct consequence of the richness of our language. There are many interesting properties that hold in most cases, and which we would like to formally state and prove. Unfortunately, these properties are not universally true: we can use the expressive power of our language to construct counter-examples to them. These examples are ones that would never come up in a real knowledge base, but it is very difficult to find syntactic conditions that disallow them. This is why we have concentrated on the simple, but nevertheless quite powerful, theorems we state below.

4.1 Random worlds and default reasoning

In this subsection, we focus on formulas which are assigned degree of belief 1. Given any knowledge base KB (which can, in particular, include defaults using the statistical interpretation of Section 3.3), we say that φ *is a default conclusion from* KB, and write $KB \mathrel{|\!\sim}_{rw} \varphi$, if $\Pr_\infty(\varphi|KB) = 1$. As we now show, the relation $\mathrel{|\!\sim}_{rw}$ satisfies all the basic properties of default inference discussed in Section 2.2.2. We start by proving two somewhat more general results.

Proposition 4.1.1: *If* $\models KB \Leftrightarrow KB'$, *then* $\Pr_\infty(\varphi|KB) = \Pr_\infty(\varphi|KB')$ *for all formulas* φ.

Proof: By assumption, precisely the same set of worlds satisfy KB and KB'. Therefore, for all N and $\vec{\tau}$, $\Pr_N^{\vec{\tau}}(\varphi|KB)$ and $\Pr_N^{\vec{\tau}}(\varphi|KB')$ are equal. Therefore, the limits are also equal. ▮

Proposition 4.1.2: *If* $KB \mathrel{|\!\!\sim}_{rw} \theta$, *then* $\Pr_\infty(\varphi|KB) = \Pr_\infty(\varphi|KB \wedge \theta)$ *for any* φ.

Proof: Fix N and $\vec{\tau}$. Then, by standard properties of conditional probability, we get

$$\Pr_N^{\vec{\tau}}(\varphi|KB) = \Pr_N^{\vec{\tau}}(\varphi|KB \wedge \theta) \cdot \Pr_N^{\vec{\tau}}(\theta|KB) \; + \Pr_N^{\vec{\tau}}(\varphi|KB \wedge \neg\theta) \cdot \Pr_N^{\vec{\tau}}(\neg\theta|KB).$$

By assumption, $\Pr_N^{\vec{\tau}}(\theta|KB)$ tends to 1 when we take limits, so the first term tends to $\Pr_N^{\vec{\tau}}(\varphi|KB \wedge \theta)$. On the other hand, $\Pr_N^{\vec{\tau}}(\neg\theta|KB)$ has limit 0. Because $\Pr_N^{\vec{\tau}}(\varphi|KB \wedge \neg\theta)$ is bounded, we conclude that the second summand also tends to 0. The result follows. ▮

Theorem 4.1.3: *The relation* $\mathrel{|\!\!\sim}_{rw}$ *satisfies the properties of And, Cautious Monotonicity, Cut, Left Logical Equivalence, Or, Reflexivity, and Right Weakening.*

Proof:

And: As we mentioned earlier, this follows from the other properties proved below.

Cautious Monotonicity and Cut: These follow immediately from Proposition 4.1.2.

Left Logical Equivalence: Follows immediately from Proposition 4.1.1.

Or: Assume $\Pr_\infty(\varphi|KB) = \Pr_\infty(\varphi|KB') = 1$, so that $\Pr_\infty(\neg\varphi|KB) = \Pr_\infty(\neg\varphi|KB') = 0$. Fix N and $\vec{\tau}$. Then

$$\begin{aligned}
\Pr_N^{\vec{\tau}}(\neg\varphi|KB \vee KB') &= \Pr_N^{\vec{\tau}}(\neg\varphi \wedge (KB \vee KB')|KB \vee KB') \\
&\leq \Pr_N^{\vec{\tau}}(\neg\varphi \wedge KB|KB \vee KB') \; + \Pr_N^{\vec{\tau}}(\neg\varphi \wedge KB'|KB \vee KB') \\
&\leq \Pr_N^{\vec{\tau}}(\neg\varphi|KB) + \Pr_N^{\vec{\tau}}(\neg\varphi|KB').
\end{aligned}$$

Taking limits, we conclude that $(KB \vee KB') \mathrel{|\!\!\sim}_{rw} \varphi$.

Reflexivity: Because we restrict attention to KB's that are eventually consistent, $\Pr_\infty(KB|KB)$ is well-defined. But then $\Pr_\infty(KB|KB)$ is clearly equal to 1.

Right Weakening: Suppose $\Pr_\infty(\varphi|KB) = 1$. If $\models \varphi \Rightarrow \varphi'$, then the set of worlds satisfying φ' is a superset of the set of worlds satisfying φ. Therefore, for any N and $\vec{\tau}$, $\Pr_N^{\vec{\tau}}(\varphi'|KB) \geq \Pr_N^{\vec{\tau}}(\varphi|KB)$. Taking limits, we obtain that

$$1 \geq \Pr_\infty(\varphi'|KB) \geq \Pr_\infty(\varphi|KB) = 1,$$

and so necessarily $\Pr_\infty(\varphi'|KB) = 1$. ▮

Besides demonstrating that $\mid\!\sim_{rw}$ satisfies the minimal standards of reasonableness for a default inference relation, these properties, particularly the stronger form of *Cut* and *Cautious Monotonicity* proved in Proposition 4.1.2, will prove quite useful in computing degrees of belief, especially when combined with some other properties we prove below (see also Section 8.1.4). In particular, many of our later results show how random-worlds behaves for knowledge bases and queries that have certain restricted forms. Sometimes a *KB* that does not satisfy these requirements can be changed into one that does, simply by extending *KB* with some of its default conclusions. We then appeal to Proposition 4.1.2 to justify using the new knowledge base instead of the old one. The other rules can also be useful in certain cases, as shown in the following analysis of Poole's "broken-arm" example [Poo89].

Example 4.1.4: Suppose we have predicates *LeftUsable*, *LeftBroken*, *RightUsable*, *RightBroken*, indicating, respectively, that the left hand is usable, the left hand is broken, the right hand is usable, and the right hand is broken. Let KB'_{arm} consist of the statements

- $||\mathit{LeftUsable}(x)||_x \approx 1$, $||\mathit{LeftUsable}(x)|\mathit{LeftBroken}(x)||_x \approx 0$ (left hands are typically usable, but not if they are broken),
- $||\mathit{RightUsable}(x)||_x \approx 1$, $||\mathit{RightUsable}(x)|\mathit{RightBroken}(x)||_x \approx 0$ (right hands are typically usable, but not if they are broken).

Now, consider $\mathit{KB}_{arm} = (\mathit{KB}'_{arm} \wedge (\mathit{LeftBroken}(\mathit{Eric}) \vee \mathit{RightBroken}(\mathit{Eric})))$; that is, we know that Eric has a broken arm. Poole observes that if we use Reiter's default logic, there is precisely one extension of KB_{arm}, and in that extension, both arms are usable. However, it is easy to see that

$$\mathit{KB}'_{arm} \wedge \mathit{LeftBroken}(\mathit{Eric}) \mid\!\sim_{rw} \neg\mathit{LeftUsable}(\mathit{Eric}) \vee \neg\mathit{RightUsable}(\mathit{Eric})$$

(this is a simple conclusion of Theorem 4.2.1); the same conclusion is obtained from $\mathit{KB}'_{arm} \wedge \mathit{RightBroken}(\mathit{Eric})$. By the *Or* rule, it follows that

$$\mathit{KB}_{arm} \mid\!\sim_{rw} \neg\mathit{LeftUsable}(\mathit{Eric}) \vee \neg\mathit{RightUsable}(\mathit{Eric}).$$

Using similar reasoning, we can also show that

$$\mathit{KB}_{arm} \mid\!\sim_{rw} \mathit{LeftUsable}(\mathit{Eric}) \vee \mathit{RightUsable}(\mathit{Eric}).$$

By applying the *And* rule, we conclude by default from KB_{arm} that exactly one of Eric's hands is usable (although we do not know which one). ∎

The final property mentioned in Section 2.2.2 is *Rational Monotonicity*. Recall that *Rational Monotonicity* asserts that if $\mathit{KB} \mid\!\sim_{rw} \varphi$ and $\mathit{KB} \not\mid\!\sim_{rw} \neg\theta$ then $(\mathit{KB} \wedge \theta) \mid\!\sim_{rw} \varphi$. We cannot prove that random worlds satisfies *Rational Monotonicity* in full generality. The problem lies with the clause $\mathit{KB} \not\mid\!\sim_{rw} \neg\theta$, which is an abbreviation for $\Pr_\infty(\neg\theta|\mathit{KB}) \neq 1$. Now there are two reasons that we might have $\Pr_\infty(\neg\theta|\mathit{KB}) \neq 1$. The first is that $\Pr_\infty(\neg\theta|\mathit{KB}) = \alpha < 1$; the second is that the limit does not exist. In the former case, which is what typically arises in the examples of interest to us, *Rational Monotonicity* does hold, as we show below. In the latter case it may not, since if $\Pr_\infty(\neg\theta|KB)$ does not have a limiting value, then $\Pr_\infty(\varphi|KB \wedge \theta)$ may not have a limit either (although if it does have a limit, its value must be 1). We point out that the same problem is encountered by other formalisms, for example by ϵ-semantics. Thus, we get the following restricted form of *Rational Monotonicity*:

Theorem 4.1.5: *Assume that $KB \mathrel{|\!\sim}_{rw} \varphi$ and $KB \mathrel{|\!\not\sim}_{rw} \neg\theta$. Then $KB \wedge \theta \mathrel{|\!\sim}_{rw} \varphi$ provided that $\Pr_\infty(\varphi|KB \wedge \theta)$ exists. Moreover, a sufficient condition for $\Pr_\infty(\varphi|KB \wedge \theta)$ to exist is that $\Pr_\infty(\theta|KB)$ exists.*

4.2 Specificity and inheritance in random worlds

One way of using random worlds is to form conclusions about particular individuals, using general statistical knowledge. This is, of course, the type of reasoning reference-class theories were designed to deal with. Recall, these theories aim to discover a single *local* piece of data — the statistics for a single reference class — that captures all the relevant information. This idea is also useful in default reasoning, where we sometimes want to find a single appropriate default. Random worlds rejects this idea as a general approach, but supports it as a valuable heuristic in special cases.

In this section, we give two theorems covering some of these cases where random worlds agrees with the basic philosophy of reference classes. Both results concern *specificity* — the idea of using the "smallest" relevant reference class for which we have statistics. However, both results also allow some indifference to irrelevance information. In particular, the second theorem also covers certain forms of *inheritance* (as described in Section 2.2.3). The results cover almost all of the noncontroversial applications of specificity and inheritance that we are aware of, and do not suffer from any of the commonly found problems such as the disjunctive reference class problem (see Section 2.1.2). Because our theorems are derived properties rather than postulates, consistency is assured and there are no *ad hoc* syntactic restrictions on the choice of possible reference classes.

Our first, and simpler, result is *basic direct inference*, where we have a single reference class that is precisely the "right one". That is, assume that the assertion $\psi(c)$ represents everything the knowledge base tells us about the constant c. In this case, we can view the class defined by $\psi(x)$ as the class of all individuals who are "just like c". If we have adequate statistics for the class $\psi(x)$, then we should clearly use this information. For example, assume that all we know about Eric is that he exhibits jaundice, and let ψ represent the class of patients with jaundice. If we know that 80% of patients with jaundice exhibit hepatitis, then basic direct inference would dictate a degree of belief of 0.8 in Eric having hepatitis. We would, in fact, like this to hold regardless of any other information we might have in the knowledge base. For example, we may know the proportion of hepatitis among patients in general, or that patients with jaundice and fever typically have hepatitis. But if all we know about Eric is that he has jaundice, we would still like to use the statistics for this class, regardless of this additional information.

Our result essentially asserts the following: "If we are interested in obtaining a degree of belief in $\varphi(c)$, and the KB is of the form $\psi(c) \wedge \|\varphi(x)|\varphi(x)\|_x \approx \alpha \wedge KB'$, then conclude that $\Pr_\infty(\varphi(c)|KB) = \alpha$." Clearly, in order for the result to hold we must make certain assumptions. The first is that $\psi(c)$ represents all the information we have about c. This intuition is formalized by assuming that KB' does not mention c. However, we need to make two other assumptions: that c also does not appear in either $\varphi(x)$ or $\psi(x)$. To understand why c cannot appear in $\varphi(x)$, suppose that $\varphi(x)$ is $Q(x) \vee x = c$, $\psi(x)$ is *true*, and the KB is $\|\varphi(x)|\mathit{true}\|_{\approx_1} 0.5$. If the result held in this case, we would be able to conclude that $\Pr_\infty(\varphi(c)|KB) = 0.5$. But since $\varphi(c)$ holds vacuously, we actually obtain that $\Pr_\infty(\varphi(c)|KB) = 1$. To see why the constant c cannot appear in $\psi(x)$, suppose that $\psi(x)$ is $P(x) \vee (x \neq c \wedge \neg P(x))$, $\varphi(x)$ is $\neg P(x)$, and the KB is $\psi(c) \wedge \|\neg P(x)|\psi(x)\|_x \approx_2 0.5$. Again, if the result held, we would be able to conclude that $\Pr_\infty(\neg P(c)|KB) = 0.5$. But $\psi(c)$ is equivalent to $P(c)$, so in fact $\Pr_\infty(\neg P(c)|KB) = 0$.

As we now, these assumption suffice to guarantee the desired result. In fact, the theorem generalizes the basic principle to properties and classes dealing with more than one individual at a time (as is shown in some of the examples following the theorem). In the following, let $\vec{x} = \{x_1, \ldots, x_k\}$ and $\vec{c} = \{c_1, \ldots, c_k\}$ be sets of distinct variables and distinct constants, respectively.

Theorem 4.2.1: *Let KB be a knowledge base of the form $\psi(\vec{c}) \wedge KB'$, and assume that for all sufficiently small tolerance vectors $\vec{\tau}$,*

$$KB[\vec{\tau}] \models \|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \in [\alpha, \beta].$$

If no constant in $\vec{c}$ appears in KB', in $\varphi(\vec{x})$, or in $\psi(\vec{x})$, then $\Pr_\infty(\varphi(\vec{c})|KB) \in [\alpha, \beta]$.

Proof: First, fix any sufficiently small tolerance vector $\vec{\tau}$, and consider a domain size N for which $KB[\vec{\tau}]$ is satisfiable. The proof strategy is to partition the size N worlds that satisfy $KB[\vec{\tau}]$ into disjoint clusters and then prove that, within each cluster, the probability of $\varphi(\vec{c})$ given $KB[\vec{\tau}]$ is in the range $[\alpha, \beta]$. From this, we can show that the (unpartitioned) probability is in this range also.

The size N worlds satisfying $KB[\vec{\tau}]$ are partitioned so that two worlds are in the same cluster if and only if they agree on the denotation of all symbols in the vocabulary Φ except for the constants in $\vec{c}$. Now consider one such cluster, and let $A \subseteq \{1, \ldots, N\}^k$ be the denotation of $\psi(\vec{x})$ inside the cluster. That is, if W is a world in the cluster, then

$$A = \{(d_1, \ldots, d_k) \in \{1, \ldots, N\}^k \ : \ (W, \vec{\tau}) \models \psi(d_1, \ldots, d_k)\}.$$

Note that, since $\psi(\vec{x})$ does not mention any of the constants in $\vec{c}$, and the denotation of everything else is fixed throughout the cluster, the set A is independent of the world W chosen in its definition. Similarly, let $B \subseteq A$ be the denotation of $\varphi(\vec{x}) \wedge \psi(\vec{x})$ in the cluster. Since the worlds in the cluster all satisfy $KB[\vec{\tau}]$, and $KB[\vec{\tau}] \models \|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \in [\alpha, \beta]$, we know that $|B|/|A| \in [\alpha, \beta]$. Since none of the constants in $\vec{c}$ are mentioned in KB except for the statement $\psi(\vec{c})$, each k-tuple in A is a legal denotation for $\vec{c}$. There is precisely one world in the cluster for each such denotation, and all worlds in the cluster are of this form. Among those worlds, only those corresponding to tuples in B satisfy $\varphi(\vec{c})$. Therefore, the fraction of worlds in the cluster satisfying $\varphi(\vec{c})$ is $|B|/|A| \in [\alpha, \beta]$.

The probability $\Pr_N^{\vec{\tau}}(\varphi(\vec{c})|KB)$ is a weighted average of the probabilities within the individual clusters, so it also has to be in the range $[\alpha, \beta]$.

It follows that $\liminf_{N\to\infty} \Pr_N^{\vec{\tau}}(\varphi(\vec{c})|KB)$ and $\limsup_{N\to\infty} \Pr_N^{\vec{\tau}}(\varphi(\vec{c})|KB)$ are also in the range $[\alpha, \beta]$. Since this holds for every sufficiently small $\vec{\tau}$, we conclude that if both limits

$$\lim_{\vec{\tau}\to\vec{0}} \liminf_{N\to\infty} \Pr_N^{\vec{\tau}}(\varphi(\vec{c})|KB) \quad \text{and} \quad \lim_{\vec{\tau}\to\vec{0}} \limsup_{N\to\infty} \Pr_N^{\vec{\tau}}(\varphi|KB)$$

exist and are equal, then $\Pr_\infty(\varphi(\vec{c})|KB)$ has to be in the range $[\alpha, \beta]$, as desired. ▮

This theorem refers to any statistical information about $\|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}}$ that can be inferred from the knowledge base. An important special case is when the knowledge base explicitly contains the relevant information.

Corollary 4.2.2: *Let KB' be the conjunction*

$$\psi(\vec{c}) \wedge (\alpha \preceq_i \|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \preceq_j \beta)\,.$$

Let KB be a knowledge base of the form $KB' \wedge KB''$ such that no constant in $\vec{c}$ appears in KB'', in $\varphi(\vec{x})$, or in $\psi(\vec{x})$. Then

$$\Pr_\infty(\varphi(\vec{c})|KB) \in [\alpha, \beta].$$

Proof: Let $\epsilon > 0$, and let $\vec{\tau}$ be sufficiently small that $\tau_i, \tau_j < \epsilon$. For this $\vec{\tau}$, the formula $(\alpha \preceq_i \|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \preceq_j \beta)$ implies $\|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \in [\alpha \Leftrightarrow \epsilon, \beta + \epsilon]$. Therefore, by Theorem 4.2.1, $\Pr_\infty(\varphi(\vec{c})|KB) \in [\alpha \Leftrightarrow \epsilon, \beta + \epsilon]$. But since this holds for any $\epsilon > 0$, it is necessarily the case that $\Pr_\infty(\varphi(\vec{c})|KB) \in [\alpha, \beta]$. ▮

It is interesteing to note one way in which our result diverges from the reference-class paradigm. Suppose we consider a query $\varphi(c)$, and that our knowledge base KB is as in the hypothesis of Corollary 4.2.2. While we can indeed conclude that $\Pr_\infty(\varphi(\vec{c})|KB) \in [\alpha, \beta]$, the exact value of this degree of belief depends on the other information in the knowledge base. Thus, while random worlds uses the local information $\alpha \preceq_i \|\varphi(x)|\psi(x)\|_x \preceq_j \beta$, it does not ignore the rest of the knowledge base. On the other hand, if the interval $[\alpha, \beta]$ is sufficiently small (and, in particular, when $\alpha = \beta$), then we may not care exactly where in the interval the degree of belief lies. In this case, we can ignore all the information in KB', and use the single piece of local information for computing the degree of belief. This potential ability to ignore large parts of the knowledge base may lead to important computational benefits (see also Section 8.1.4).

We now present a number of examples that demonstrate the behavior of the direct inference result.

Example 4.2.3: Consider a knowledge base describing the hepatitis example discussed earlier. In the notation of Corollary 4.2.2:

$$KB'_{hep} = \mathit{Jaun}(\mathit{Eric}) \land \|\mathit{Hep}(x)|\mathit{Jaun}(x)\|_x \approx_1 0.8,$$

and

$$\begin{aligned} KB_{hep} \;&=\; KB'_{hep} \land \|\mathit{Hep}(x)\|_{\preceq_2} 0.05 \;\land \\ &\quad\; \|\mathit{Hep}(x)|\mathit{Jaun}(x) \land \mathit{Fever}(x)\|_x \approx_2 1. \end{aligned}$$

Then $\Pr_\infty(\mathit{Hep}(\mathit{Eric})|KB_{hep}) = 0.8$ as desired; information about other reference classes (whether more general or more specific) ignored. Other kinds of information are also ignored, for example, information about other individuals. Thus,

$$\Pr_\infty(\mathit{Hep}(\mathit{Eric})|KB_{hep} \land \mathit{Hep}(\mathit{Tom})) = 0.8. \quad \blacksquare$$

Although it is nothing but an immediate application of Theorem 4.2.1, it is worth remarking that the principle of *Direct Inference for Defaults* (Section 2.2.3) is satisfied by random-worlds:

Corollary 4.2.4: *Suppose KB implies* $\|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \approx_i 1$*, and no constant in* $\vec{c}$ *appears in KB,* φ*, or* ψ*. Then* $\Pr_\infty(\varphi(\vec{c})|KB \land \psi(\vec{c})) = 1$*.*

As discussed in Section 2.2.3, this shows that simple forms of inheritance reasoning are possible.

Example 4.2.5: The knowledge base KB_{fly} from Section 2.2.3 is, under our interpretation of defaults:

$$\begin{aligned} &\|\mathit{Fly}(x)|\mathit{Bird}(x)\|_x \approx_1 1 \land \\ &\|\mathit{Fly}(x)|\mathit{Penguin}(x)\|_x \approx_2 0 \land \\ &\forall x\, (\mathit{Penguin}(x) \Rightarrow \mathit{Bird}(x)). \end{aligned}$$

Then $\Pr_\infty(\mathit{Fly}(\mathit{Tweety})|KB_{\mathit{fly}} \land \mathit{Penguin}(\mathit{Tweety})) = 0$. That is, we conclude that Tweety the penguin does not fly, even though he is also a bird and birds generally do fly. $\blacksquare$

Given this preference for the most specific reference class, one might wonder why random worlds does not encounter the problem of *disjunctive reference classes* (see Section 2.1.2). The following example, based on the example from Section 2.1.2, provides one answer.

Example 4.2.6: Recall the knowledge base KB'_{hep} from the hepatitis example above, and consider the reference class $\psi(x) =_{\mathrm{def}} \mathit{Jaun}(x) \land (\neg\mathit{Hep}(x) \lor x = \mathit{Eric})$. Clearly, as the domain size grows large, $\|\mathit{Hep}(x)|\psi(x)\|_x$ grows arbitrarily close to 0.[1] Therefore, for any fixed $\epsilon > 0$,

$$\Pr_\infty\left(\|\mathit{Hep}(x)|\psi(x)\|_x \in [0,\epsilon] \,\middle|\, KB'_{hep}\right) = 1.$$

[1] This actually relies on the fact that, with high probability, the proportion (as the domain size grows) of jaundiced patients with hepatitis is nonzero. We do not prove this fact here; see [PV89] and the related discussion in Chapter 7.

From Theorem 4.1.3, by *Cautious Monotonicity* and *Cut* we can assume (for the purposes of making further inferences) that KB'_{hep} actually contains this assertion. Moreover, $\psi(Eric)$ is $Jaun(Eric) \wedge (\neg Hep(Eric) \vee Eric = Eric)$, which is equivalent to $Jaun(Eric)$: i.e., precisely the information about Eric in KB'_{hep}. Therefore $\psi(x)$ would appear to be a more specific reference class for $Hep(Eric)$ than $Jaun(x)$, and with very different statistics. But in Example 4.2.3 we showed that $\Pr_\infty(Hep(Eric)|KB'_{hep}) = 0.8$. So random worlds avoids using the spurious disjunctive reference class $\psi(x)$. This is explained by noting that the reference class $\psi(x)$ class violates the conditions of Theorem 4.2.1 because it explicitly mentions the constant *Eric*. It is worth pointing out that, in this example, the real problem is not disjunction at all. Indeed, disjunctive classes that are not improperly defined (by referring to the constants) are not ignored; see Example 4.2.17. ▮

As we have said, we are not limited to unary predicates, nor to examining only one individual at a time.

Example 4.2.7: In Example 3.3.1, we showed how to formalize the elephant-zookeeper example discussed in Section 2.2.4. As we now show, the natural representation of KB_{likes} indeed yields the answers we expect. Suppose we know that $Elephant(Clyde)$ and $Zookeeper(Eric)$. We consider two queries. First, assume we are interested in finding out whether Clyde likes Eric. In this case, our reference class $\psi(x,y)$ is $Elephant(x) \wedge Zookeeper(y)$. Based on the first statement in KB_{likes}, Corollary 4.2.4 allows us to conclude that $\Pr_\infty(Likes(Clyde, Eric)|KB_{likes}) = 1$. But does Clyde like Fred? In this case, our reference class is $\psi(x) = Elephant(x)$, and by Corollary 4.2.4, we use the second default to conclude $\Pr_\infty(Likes(Clyde, Fred)|KB_{likes}) = 0$. Note that we cannot use the same reasoning for Fred as we did for Eric in order to conclude that Clyde likes Fred. If we try to apply the reference class $Elephant(x) \wedge Zookeeper(y)$ to the pair $(Clyde, Fred)$, the conditions of the corollary are violated, because the constant Fred is used elsewhere in the knowledge base. ▮

The same principles continue to hold for more complex sentences; for example, we can mix first-order logic and statistical knowledge arbitrarily and we can nest defaults.

Example 4.2.8: In Example 3.3.2, we showed how to express the default: "People who have at least one tall parent are typically tall." If we have this default, and also know that $\exists y\,(Child(Alice, y) \wedge Tall(y))$ (Alice has a tall parent), Corollary 4.2.4 tells us that we could conclude by default that $Tall(Alice)$. ▮

Example 4.2.9: In Example 3.3.3, we showed how the default "Typically, people who normally go to bed late normally rise late" can be expressed in our language using the knowledge base KB_{late}. Let KB'_{late} be

$$\begin{array}{l} KB_{late} \wedge \\ \|To\text{-}bed\text{-}late(Alice, y')|Day(y')\|_{y'} \approx_2 1 \wedge \\ Day(Tomorrow). \end{array}$$

By Corollary 4.2.4, Alice typically rises late. That is,

$$\Pr_\infty(\|\mathit{Rises\text{-}late}(x,y)|\mathit{Day}(y)\|_y \approx_1 1 \;|\mathit{KB}'_{late}) = 1.$$

By *Cautious Monotonicity* and *Cut*, we can add this conclusion (which is itself a default) to KB'_{late}. By Corollary 4.2.4 again, we then conclude that Alice can be expected to rise late on any particular day. So, for instance:

$$\Pr_\infty(\mathit{Rises\text{-}late}(\mathit{Alice}, \mathit{Tomorrow})|\mathit{KB}'_{late}) = 1. \quad \blacksquare$$

In all the examples presented so far in this section, we have statistics for precisely the right reference class to match our knowledge about the individual(s) in question; Theorem 4.2.1 and its corollaries require this. Unfortunately, in many cases our statistical information is not detailed enough for Theorem 4.2.1 to apply. Consider the knowledge base KB_{hep} from the hepatitis example. Here we have statistics for the occurrence of hepatitis among the class of patients who are just like Eric, so we can use those to induce a degree of belief in $\mathit{Hep}(\mathit{Eric})$. But now consider the knowledge base $\mathit{KB}_{hep} \wedge \mathit{Tall}(\mathit{Eric})$. Since we do not have statistics for the frequency of tall patients with hepatitis, the results we have seen so far do not apply. We would like to be able to ignore $\mathit{Tall}(\mathit{Eric})$. But what entitles us to ignore $\mathit{Tall}(\mathit{Eric})$ and not $\mathit{Jaun}(\mathit{Eric})$? To solve this problem in complete generality requires a better theory of irrelevance than we currently have. Nevertheless, our next theorem covers many cases, including many of the more uncontroversial examples found in the default reasoning literature.

The theorem we present deals with a knowledge base KB that defines a "minimal" reference class ψ_0 with respect to the query $\varphi(c)$. More precisely, assume that KB gives statistical information regarding $\|\varphi(x)|\psi_i(x)\|_x$ for a number of different reference classes $\psi_i(x)$. However, among these classes, there is one class $\psi_0(x)$ that is *minimal* — all other reference classes are strictly larger or entirely disjoint from it (see Figure 4.1 for an illustration). If we also know $\psi_0(c)$, we can use the statistics for $\|\varphi(x)|\psi_0(x)\|_x$ to induce a degree of belief in $\varphi(c)$. This type of reasoning is best explained using an example:

Example 4.2.10: Assume we have a knowledge base $\mathit{KB}_{taxonomy}$ containing information about birds and animals; in particular, $\mathit{KB}_{taxonomy}$ contains a taxonomic hierarchy of this domain. Moreover, $\mathit{KB}_{taxonomy}$ contains the following information about the swimming ability of various types of animals:

$$\begin{array}{rcll}
\|\mathit{Swims}(x)|\mathit{Penguin}(x)\|_x & \approx_1 & 0.9 & \wedge \\
\|\mathit{Swims}(x)|\mathit{Sparrow}(x)\|_x & \approx_2 & 0.01 & \wedge \\
\|\mathit{Swims}(x)|\mathit{Bird}(x)\|_x & \approx_3 & 0.05 & \wedge \\
\|\mathit{Swims}(x)|\mathit{Animal}(x)\|_x & \approx_4 & 0.3 & \wedge \\
\|\mathit{Swims}(x)|\mathit{Fish}(x)\|_x & \approx_5 & 1. &
\end{array}$$

If we also know that Opus is a penguin, then in order to determine whether Opus swims, the best reference class is surely the class of penguins. The remaining reference classes are either larger (in the case of birds or animals), or disjoint (in the case of sparrows and fish). This is the case even if we know that Opus is a black penguin with a large nose. That is, Opus *inherits*

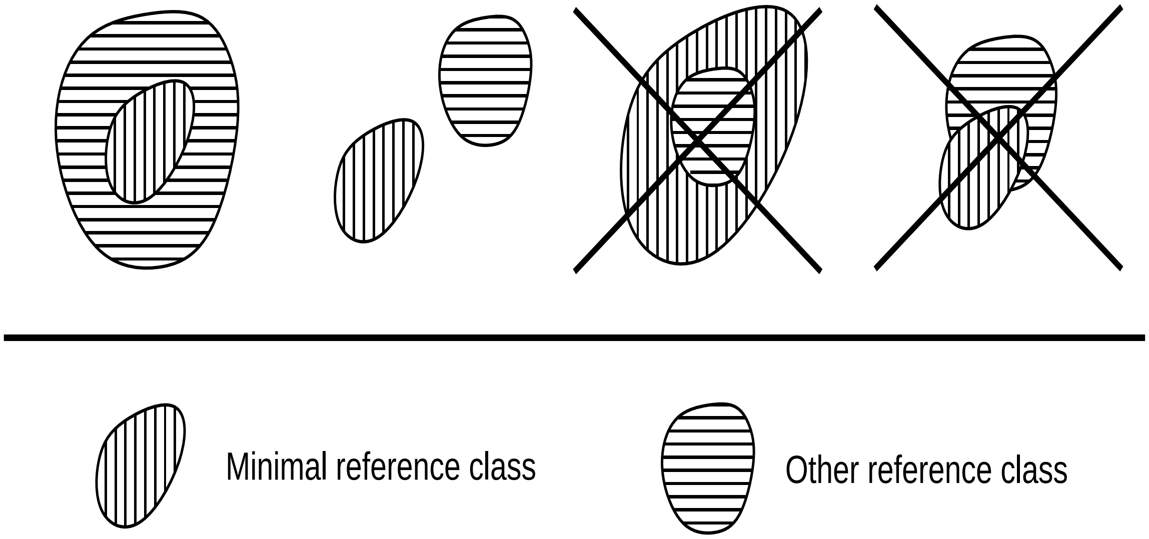

Figure 4.1: Possible relations between a minimal reference class and another reference class

the statistics for the minimal class ψ_0 — penguins — even though the class of individuals just like Opus is smaller than ψ_0. ∎

That random-worlds validates this intuition is formalized in the following theorem.

Theorem 4.2.11: *Let c be a constant and let KB be a knowledge base satisfying the following conditions:*

(a) $KB \models \psi_0(c)$,

(b) for any expression of the form $\|\varphi(x)|\psi(x)\|_x$ in KB, it is the case that either $KB \models \psi_0 \Rightarrow \psi$ or that $KB \models \psi_0 \Rightarrow \neg\psi$,

(c) the (predicate and constant) symbols in $\varphi(x)$ appear in KB only on the left-hand side of the conditional in the proportion expressions described in condition (b),

(d) the constant c does not appear in the formula $\varphi(x)$.

Assume that for all sufficiently small tolerance vectors $\vec{\tau}$:

$$KB[\vec{\tau}] \models \|\varphi(x)|\psi_0(x)\|_x \in [\alpha, \beta].$$

Then $\Pr_\infty(\varphi(c)|KB) \in [\alpha, \beta]$.

The proof of this theorem is based on the same clustering argument used in the proof of Theorem 4.2.1, but with a different definition of the cluster. See Appendix A for the complete proof.

Again, the following analogue to Corollary 4.2.2 is immediate:

Corollary 4.2.12: *Let KB' be the conjunction*

$$\psi_0(c) \wedge (\alpha \preceq_i \|\varphi(x)|\psi_0(x)\|_x \preceq_j \beta).$$

Let KB be a knowledge base of the form $KB' \wedge KB''$ that satisfies conditions (b), (c), and (d) of Theorem 4.2.11. Then

$$\Pr_\infty(\varphi(c)|KB) \in [\alpha,\beta].$$

This theorem and corollary have many useful applications.

Example 4.2.13: Consider the knowledge bases KB'_{hep} and KB_{hep} concerning jaundice and hepatitis from Example 4.2.3. In that example, we supposed that the only information about Eric contained in the knowledge base was that Eric has jaundice. It is clearly more realistic to assume that Eric's hospital records contain more information than just this fact. This theorem allows us to ignore this information in a large number of cases.

For example,

$$\Pr_\infty(\mathit{Hep}(\mathit{Eric})|KB'_{hep} \wedge \mathit{Fever}(\mathit{Eric}) \wedge \mathit{Tall}(\mathit{Eric})) = 0.8,$$

as desired. On the other hand,

$$\Pr_\infty(\mathit{Hep}(\mathit{Eric})|KB_{hep} \wedge \mathit{Fever}(\mathit{Eric}) \wedge \mathit{Tall}(\mathit{Eric})) = 1.$$

(Recall that KB_{hep} includes $\|\mathit{Hep}(x)|\mathit{Jaun}(x) \wedge \mathit{Fever}(x)\|_x \approx_2 1$, while KB'_{hep} does not.) This shows why it is important that we identify the most specific reference class for the query φ. The most specific reference class for $\mathit{Hep}(\mathit{Eric})$ with respect to $KB'_{hep} \wedge \mathit{Fever}(\mathit{Eric}) \wedge \mathit{Tall}(\mathit{Eric})$ is $\|\mathit{Hep}(x)|\mathit{Jaun}(x)\|_x \approx_1 0.8$, while with respect to $KB_{hep} \wedge \mathit{Fever}(\mathit{Eric}) \wedge \mathit{Tall}(\mathit{Eric})$ it is $\|\mathit{Hep}(x)|\mathit{Jaun}(x) \wedge \mathit{Fever}(x)\|_x \approx_2 1$. In the latter case, the less-specific reference classes *Jaun* and *true* are ignored. ▮

As discussed in Section 2.2.3, various inheritance properties are considered desirable in default reasoning as well. To begin with, we note that Theorem 4.2.11 covers the simpler cases (which can also be seen as applications of rational monotonicity):

Example 4.2.14: In simple cases, Theorem 4.2.11 shows that random worlds is able to apply defaults in the presence of "obviously irrelevant" additional information. For example, using the knowledge base KB_{fly} (see Example 4.2.5):

$$\Pr_\infty(\mathit{Fly}(\mathit{Tweety})|KB'_{\mathit{fly}} \wedge \mathit{Penguin}(\mathit{Tweety}) \wedge \mathit{Yellow}(Tweety)) = 0.$$

That is, Tweety the yellow penguin continues not to fly. ▮

Theorem 4.2.11 also validates the more difficult reasoning patterns that have caused problems for many default reasoning theories. In particular, we validate exceptional-subclass inheritance, in which a class that is exceptional in one respect can nevertheless inherit other unrelated properties:

Example 4.2.15: If we consider the property of warm-bloodedness as well as flight, we get:

$$\Pr_\infty\left(\mathit{Warm\text{-}blooded}(\mathit{Tweety}) \;\middle|\; \begin{array}{l} \mathit{KB}_{\mathit{fly}} \wedge \mathit{Penguin}(\mathit{Tweety}) \wedge \\ \|\mathit{Warm\text{-}blooded}(x)|\mathit{Bird}(x)\|_x \approx_3 1 \end{array} \right) = 1.$$

Knowing that Tweety does not fly because he is a penguin does not prevent us from assuming that he is like typical birds in other respects. ∎

The drowning-problem variant of the exceptional-subclass inheritance problem is also covered by the theorem.

Example 4.2.16: Suppose we know, as in Section 2.2.3, that yellow things tend to be highly visible. Then:

$$\Pr_\infty\left(\mathit{Easy\text{-}to\text{-}see}(\mathit{Tweety}) \;\middle|\; \begin{array}{l} \mathit{KB}_{\mathit{fly}} \wedge \mathit{Penguin}(\mathit{Tweety}) \wedge \mathit{Yellow}(\mathit{Tweety}) \wedge \\ \|\mathit{Easy\text{-}to\text{-}see}(x)|\mathit{Yellow}(x)\|_x \approx_3 1 \end{array} \right) = 1.$$

Here, all that matters is that Tweety is a yellow object. The fact that he is a bird, and an exceptional bird at that, is rightly ignored. ∎

Notice that, unlike Theorem 4.2.1, the conditions of Theorem 4.2.11 do not extend to inferring degrees of belief in $\varphi(\vec{c})$, where $\vec{c}$ is a tuple of constants. Roughly speaking, the reason is the ability of the language to create connections between the different constants in the tuple. For example, let $\psi_0(x_1, x_2)$ be *true*, and let KB' be $\|\mathit{Hep}(x) \wedge \neg\mathit{Hep}(y)\|_{x,y} \approx_1 0.3$. By Theorem 4.2.1, $\Pr_\infty(\mathit{Hep}(\mathit{Tom}) \wedge \neg\mathit{Hep}(\mathit{Eric})|\mathit{KB}') = 0.3$. But, of course, $\Pr_\infty(\mathit{Hep}(\mathit{Tom}) \wedge \neg\mathit{Hep}(\mathit{Eric})|\mathit{KB}' \wedge \mathit{Tom} = \mathit{Eric}) = 0$. The additional information regarding *Tom* and *Eric* cannot be ignored. This example might suggest that this is a minor problem related only to the use of equality, but more complex examples that do not mention equality can also be constructed.

As a final example in this section, we revisit the issue of disjunctive reference classes. As we saw in Example 4.2.6, random worlds does not suffer from the "disjunctive reference class" problem. As we observed in Section 2.1.2, some systems often avoid this problem by simply outlawing disjunctive reference classes. However, as the following example demonstrates, such classes can often be useful. The example demonstrates that random worlds does, in fact, treat disjunctive reference classes appropriately.

Example 4.2.17: Recall that in Section 2.1.2 we motivated the importance of disjunctive reference classes using, as an example, the disease Tay-sachs. The corresponding statistical information was represented, in our framework, as the knowledge base KB:

$$\|\mathit{TS}(x)|\mathit{EEJ}(x) \vee \mathit{FC}(x)\|_x \approx_1 0.02.$$

Given a baby *Eric* of eastern-European extraction, Theorem 4.2.11 shows us that

$$\Pr_\infty(\mathit{TS}(\mathit{Eric})|\mathit{KB} \wedge \mathit{EEJ}(\mathit{Eric})) = 0.02.$$

That is, random worlds is able to use the information derived from the disjunctive reference class, and apply it to an individual known to be in the class; indeed, it also deals with the case where we have additional information determining to which of the two populations this specific individual belongs. Thus, disjunctive reference classes are treated as is any other "legitimate" reference class (one that does not mention the constants in the query). ∎

The type of specificity and inheritance reasoning convered by our theorems are special cases of general inheritance reasoning [THT87]. While these theorems show that random worlds does validate a substantial part of the noncontroversial aspects of such reasoning, proving a general theorem asserting this claim is surprisingly subtle (partly because of the existence of numerous divergent semantics and intuitions for inheritance reasoning [THT87]). We are currently working on proving such a general claim. We do point out, however, that random worlds does not validate the generalization of inheritance reasoning to the statistical context. As shown in Example 4.3.3, we do not get, nor do we want, simple inheritance in all contexts involving statistical information.

4.3 Competing reference classes

In previous sections we have always been careful to consider examples in which there was an obviously best reference class. In practice, we will not always be this fortunate. Reference-class theories usually cannot give useful answers when there are competing candidates for the "best" class. However, random worlds does not have this problem, because the degrees of belief it defines can be combinations of the values for competing classes. In this section we examine, in very general terms, three types of competing information. The first concerns conflicts between specificity and accuracy, the second between information that is too specific and information that is too general, and the last concerns competing classes where the specificity principle is entirely irrelevant.

We discussed the conflict between specificity and accuracy in Section 2.1.3. This problem was noticed by Kyburg who, to some extent, successfully addresses this issue with his strength rule. In Section 2.1.3, we argued that, in order to assign a degree of belief to *Chirps*(*Tweety*), we should be able to use the tighter interval $[0.7, 0.8]$ even though it is associated with a less specific reference class. As we observed, Kyburg's strength rule attempts to capture this intuition. As the following result shows, the random worlds method also captures this intuition, at least when the reference classes form a chain.[2]

Theorem 4.3.1: *Suppose KB has the form*

$$\bigwedge_{i=1}^{m} (\alpha_i \preceq_{\ell_i} \|\varphi(x)|\psi_i(x)\|_x \preceq_{r_i} \beta_i) \;\wedge\; \psi_1(c) \;\wedge\; KB',$$

[2] Kyburg's rule also applies to cases where the reference classes do not form a chain. In these cases, the intuitions of the strength rule and those of random worlds diverge. We do not explore this issue further here.

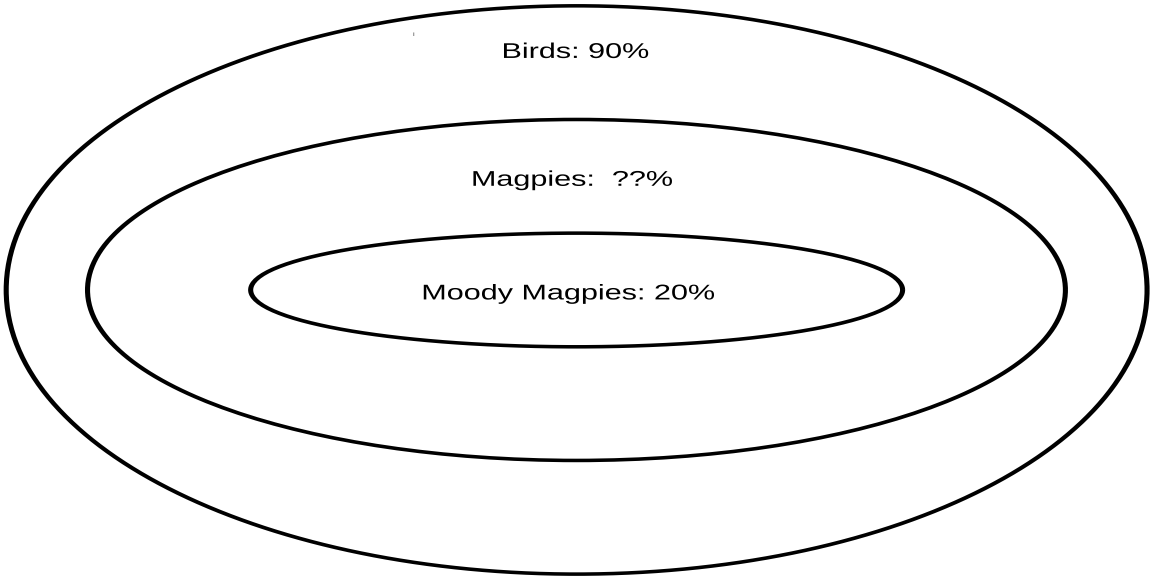

Figure 4.2: Competing reference classes: moody magpies vs. birds

and, for all i, $KB \models \forall x\ (\psi_i(x) \Rightarrow \psi_{i+1}(x)) \land \neg(||\psi_1(x)||_x \approx_1 0)$. Assume also that no symbol appearing $\varphi(x)$ appears in KB' or in any $\psi_i(c)$. Further suppose that, for some j, $[\alpha_j, \beta_j]$ is the tightest interval. That is, for all $i \neq j$, $\alpha_i < \alpha_j < \beta_j < \beta_i$. Then

$$\Pr_\infty(\varphi(c)|KB) \in [\alpha_j, \beta_j].$$

Example 4.3.2: The example described in Section 2.1.3 is essentially captured by the following knowledge base KB_{chirps}:

$$\begin{array}{l} 0.7 \preceq_1 ||Chirps(x)|Bird(x)||_x \preceq_2 0.8 \land \\ 0 \preceq_3 ||Chirps(x)|Magpie(x)||_x \preceq_4 0.99 \land \\ \forall x\,(Magpie(x) \Rightarrow Bird(x)) \land \\ Magpie(Tweety). \end{array}$$

It follows from Theorem 4.3.1 that $\Pr_\infty(Chirps(Tweety)|KB_{chirps}) \in [0.7, 0.8]$. ▮

We now consider a different situation where competing reference classes come up: when one reference class is too specific, and the other too general.

Example 4.3.3: We illustrate the problem with a example based on one of Goodwin [Goo92]. Consider KB_{magpie} (also represented in Figure 4.2):

$$\begin{array}{l} ||Chirps(x)|Bird(x)||_x \approx_1 0.9 \land \\ ||Chirps(x)|Magpie(x) \land Moody(x)||_x \approx_2 0.2 \land \\ \forall x\,(Magpie(x) \Rightarrow Bird(x)) \land \\ Magpie(Tweety). \end{array}$$

Reference class theories would typically ignore the information about moody magpies: since Tweety is not known to be moody, the class of moody magpies is not even a legitimate reference class. Using such approaches, the degree of belief would be 0.9. Goodwin argues that this is not completely reasonable: why should we ignore the information about moody magpies? Tweety could be moody (the knowledge base leaves the question open). In fact, it it is consistent with KB_{magpie} that magpies are generally moody. But ignoring the second statistic in effect amounts to assuming that magpies generally are not moody. It is hard to see that this is a reasonable assumption. The random-worlds approach supports Goodwin's intuition, and the degree of belief that Tweety flies, given KB_{magpie}, can be shown to have a value which is less than 0.9. As a general rule, if we do not have exactly the right reference class (as for Theorem 4.2.1), then random worlds combines information from more specific classes as well as from more general classes. ∎

The third and most important type of conflict is when we have different candidate reference classes which are not related by specificity. As we argued in Section 2.1.3, this case is likely to come up very often in practice. While the complete characterization of the behavior of random worlds in such cases is somewhat complex, the following theorem illustrates what happens when the competing references classes are essentially disjoint. We capture "essentially disjoint" here by assuming that the overlap between these classes consists of precisely one member: the individual c addressed in our query We can generalize the following theorem to the case where we simply assume that the overlap between competing reference classes ψ and ψ' is small relative to the sizes of the two classes; that is, where $\|\psi(x) \wedge \psi'(x) | \psi(x)\|_x \approx 0$ and $\|\psi(x) \wedge \psi'(x) | \psi'(x)\|_x \approx 0$. For simplicity, we omit details here.

It turns out that, under this assumption, random worlds provides an independent derivation for a well-known technique for combining evidence: Dempster's rule of combination [Sha76]. Dempster's rule addresses the issue of combining *independent* pieces of evidence. Consider a query $P(c)$, and assume we have competing reference classes that are all appropriate for this query. In this case, the different pieces of evidence are the proportions of the property P of in the different competing reference classes. More precisely, we can view the fact that the proportion $\|P(x) | \psi(x)\|_x$ is α as giving evidence of weight α in favor of $P(c)$. The fact that the intersection between the different classes is "small" means that almost disjoint samples were used to compute these pieces of evidence; thus, they can be viewed as independent. Under this interpretation, Dempster's rule tells us how to combine the different pieces of evidence to obtain an appropriate degree of belief in $P(c)$. The function used in Dempter's rule is $\delta : (0,1)^m \rightarrow (0,1)$, defined as follows:

$$\delta(\alpha_1, \ldots, \alpha_m) = \frac{\prod_{i=1}^m \alpha_i}{\prod_{i=1}^m \alpha_i + \prod_{i=1}^m (1 \Leftrightarrow \alpha_i)}.$$

As the following theorem shows, this is also the answer obtained by random worlds.

Theorem 4.3.4: *Let P be a unary predicate, and consider a knowledge base KB of the following* *form:*

$$\bigwedge_{i=1}^{m} (\|P(x)|\psi_i(x)\|_x \approx_i \alpha_i \wedge \psi_i(c)) \ \wedge \ \bigwedge_{\substack{i,j=1\\ i\neq j}}^{m} \exists! x \ (\psi_i(x) \wedge \psi_j(x)) \ ,$$

where $0 < \alpha_j < 1$*, for* $j = 1, \ldots, m$*. Then, if neither* P *nor* c *appear anywhere in the formulas* $\psi_i(x)$*, then*

$$\Pr_\infty(P(c)|KB) = \delta(\alpha_1, \ldots, \alpha_m).$$

We illustrate this theorem on what is, perhaps, the most famous example of conflicting information — the *Nixon Diamond* [RC81]. Suppose we are interested in assigning a degree of belief to the assertion "Nixon is a pacifist". Assume that we know that Nixon is both a Quaker and a Republican, and we have statistical information for the proportion of pacifists within both classes. This is an example where we have two incomparable reference classes for the same query. More formally, assume that KB_{Nixon} is

$$\begin{array}{l} \|\mathit{Pacifist}(x)|\mathit{Quaker}(x)\|_x \approx_1 \alpha \wedge \\ \|\mathit{Pacifist}(x)|\mathit{Republican}(x)\|_x \approx_2 \beta \wedge \\ \mathit{Quaker}(\mathit{Nixon}) \wedge \mathit{Republican}(\mathit{Nixon}) \wedge \\ \exists! x \ (\mathit{Quaker}(x) \wedge \mathit{Republican}(x)) \ , \end{array}$$

and that φ is *Pacifist*(*Nixon*). The degree of belief $\Pr_\infty(\varphi|KB_{Nixon})$ takes different values, depending on the values α and β for the two reference classes. If $\{\alpha, \beta\} \neq \{0, 1\}$, then this limit always exists and its value is $\Pr_\infty(\varphi|KB_{Nixon}) = \frac{\alpha\beta}{\alpha\beta + (1-\alpha)(1-\beta)}$. If, for example, $\beta = 0.5$, so that the information for Republicans is neutral, we get that $\Pr_\infty(\varphi|KB_{Nixon}) = \alpha$: the data for Quakers is used to determine the degree of belief. If the evidence given by the two reference classes is conflicting — $\alpha > 0.5 > \beta$ — then $\Pr_\infty(\varphi|KB_{Nixon}) \in [\alpha, \beta]$: some intermediate value is chosen. If, on the other hand, the two reference classes provide evidence in the same direction, then the limiting probability is greater than both α and β. For example, if $\alpha = \beta = 0.8$, then the value of the limit would be around 0.94. This has a reasonable explanation: if we have two independent bodies of evidence, both supporting φ quite strongly, when we combine them we should get even more support for φ.

Now, assume that our information is not entirely quantitative. For example, we may know that "Quakers are typically pacifists". In our framework, this corresponds to assigning $\alpha = 1$. If our information for Republicans is not a default — $\beta > 0$ — then the limiting probability $\Pr_\infty(\varphi|KB_{Nixon})$ is 1. As expected, a default (i.e., an "extreme" value) dominates. But what happens in the case where we have conflicting defaults for the two reference classes? It turns out that, in this case, the limiting probability does not exist. This is because the limit is *non-robust*: its value depends on the way in which the tolerances $\vec{\tau}$ tend to 0. More precisely, if $\tau_1 \ll \tau_2$, so that the "almost all" in the statistical interpretation of the first conjunct is much closer to "all" than the "almost none" in the second conjunct is closer to "none", then the limit is 1. We can view the magnitude of the tolerance as representing the strength of the default. Thus, in this case, the first conjunct represents a default with higher priority than the second conjunct. Symmetrically, if $\tau_1 \gg \tau_2$, then the limit is 0. On the other hand, if $\tau_1 = \tau_2$, then the limit is 1/2.

The nonexistence of this limit is not simply a technical artifact of our approach. The fact that we obtain different limiting degrees of belief depending on how $\vec{\tau}$ goes to 0 is closely related to the existence of *multiple extensions* in many other theories of default reasoning (for instance, in default logic [Rei80].) Both non-robustness and the existence of more than one extension suggest a certain incompleteness of our knowledge. It is well-known that, in the presence of conflicting defaults, we often need more information about the strength of the different defaults in order to resolve the conflict. Our approach has the advantage of pinpointing the type of information that would suffice to reach a decision. Note that our formalism does give us an explicit way to state in this example that the two extensions are equally likely, by asserting that the defaults that generate them have equal strength; namely, we can use $\approx_1$ to capture both default statements, rather than using $\approx_1$ and $\approx_2$. In this case, we get the answer 1/2, as expected. However, it is not always appropriate to conclude that defaults have equal strength. We can easily extend our language to allow the user to prioritize defaults, by defining the relative sizes of the components τ_i of the tolerance vector.

As we mentioned, Theorem 4.3.4 only tells us how to combine statistics from competing reference classes in this very special case where the intersection of the different refernce classes is small. We are pessimistic about the chances of getting one general result that covers "most" cases of interest. Rather, further research will probably result in a collection of special case theorems. That further results are possible is illustrated by Shastri [Sha89]. He states a result which is, in many ways, quite similar to ours. However, his result addresses a different special case. He essentially assumes that, in addition to the statistics for P within each reference class, the statistics for P in the general population are also known. While this is also a restrictive assumption, it is entirely orthogonal to the assumption needed for our theorem. We may certainly hope to find other results of this general form.

4.4 Independence

As we have seen so far, random worlds captures a large number of the natural reasoning heuristics that have been proposed in the literature. Another heuristic is a default assumption that all properties are probabilistically independent unless we know otherwise. Random-worlds captures this principle as well, in many cases. It is, in general, very hard to give simple syntactic tests for when a knowledge base forces two properties to be dependent. The following theorem concerns one very simple scenario where we can be sure that no relationship is forced.

Consider two disjoint vocabularies Φ and Φ', and two respective knowledge-base and query pairs: $KB, \varphi \in \mathcal{L}(\Phi)$, and $KB', \varphi' \in \mathcal{L}(\Phi')$. We can prove that

$$\Pr_\infty(\varphi \wedge \varphi' | KB \wedge KB') = \Pr_\infty(\varphi | KB) \cdot \Pr_\infty(\varphi' | KB').$$

That is, if we have no way of forcing a connection between the symbols in the two vocabularies, the two queries will be independent: the probability of their conjunction is the product of their probabilities. We now prove a slightly more general case, where the two queries are both allowed to refer to some constant c.

Theorem 4.4.1: *Let Φ_1 and Φ_2 be two vocabularies disjoint except for the constant c. Consider $KB_1, \varphi_1 \in \mathcal{L}(\Phi_1)$ and $KB_2, \varphi_2 \in \mathcal{L}(\Phi_2)$. Then*

$$\Pr_\infty(\varphi_1 \wedge \varphi_2 | KB_1 \wedge KB_2) = \Pr_\infty(\varphi_1 | KB_1) \cdot \Pr_\infty(\varphi_2 | KB_2).$$

Although very simple, this theorem allows us to deal with such examples as the following:

Example 4.4.2: Consider the knowledge base KB_{hep}, and a knowledge base stating that 40% of hospital patients are over 60 years old:

$$KB_{>60} =_{\rm def} \|\mathit{Over60}(x)|\mathit{Patient}(x)\|_x \approx_5 0.4$$

Then

$$\begin{aligned}&\Pr_\infty(\mathit{Hep}(\mathit{Eric}) \wedge \mathit{Over60}(\mathit{Eric}) | KB_{hep} \wedge KB_{>60}) = \\ &\quad \Pr_\infty(\mathit{Hep}(\mathit{Eric}) | KB_{hep}) \cdot \Pr_\infty(\mathit{Over60}(\mathit{Eric}) | KB_{>60}) = 0.8 \cdot 0.4 = 0.32. \ \blacksquare\end{aligned}$$

In the case of a unary vocabulary (i.e., one containing only unary predicates and constants), Theorem 4.4.1 follows from results we state in Chapter 7, where we describe a deep connection between the random-worlds method and maximum entropy for unary vocabularies. It is a well-known fact that using maximum entropy often leads to probabilistic independence. The result above proves that, with random-worlds, this phenomenon appears in the non-unary case as well.

We remark that the connection between maximum entropy and independence is often overstated. For example, neither maximum entropy nor random worlds lead to probabilistic independence in examples like the following:

Example 4.4.3: Consider the knowledge base KB, describing a domain of animals:

$$\|\mathit{Black}(x)|\mathit{Bird}(x)\|_x \approx_1 0.2 \ \wedge \ \|\mathit{Bird}(x)\|_x \approx_2 0.1.$$

It is perfectly consistent for *Bird* and *Black* to be probabilistically independent. If this were the case, we would expect the proportion of black animals to be the same as that of black birds. In this case, our degree of belief in *Black*(*Clyde*), for some arbitrary animal Clyde, would also be 0.2. However, this is not the case. Since all the predicates here are unary, using the maximum entropy techniques discussed in Chapter 7, we can show that $\Pr_\infty(\mathit{Black}(\mathit{Clyde})|KB) = 0.47$. ∎

4.5 The lottery paradox and unique names

In Section 2.2.5 we discussed the *lottery paradox* and the challenge it poses to theories of default reasoning. How does random-worlds perform?

To describe the original problem in our framework, let $\mathit{Ticket}(x)$ hold if x purchased a lottery ticket. Consider the knowledge base consisting of

$$\mathit{KB} = \exists! x\ \mathit{Winner}(x) \land \forall x\, (\mathit{Winner}(x) \Rightarrow \mathit{Ticket}(x)).$$

That is, there is a unique winner, and in order to win one must purchase a lottery ticket. If we also know the size of the lottery, say N, we can add to our knowledge base the assertion $\exists^N x\ \mathit{Ticket}(x)$ stating that there are precisely N ticket holders. (This assertion can easily be expressed in first-order logic using equality.) We also assume for simplicity that each individual buys at most one lottery ticket. Then our degree of belief that the individual denoted by a particular constant c wins the lottery is

$$\Pr_\infty(\mathit{Winner}(c) | \mathit{KB} \land \exists^N x\ \mathit{Ticket}(x) \land Ticket(c)) = \frac{1}{N}.$$

Our degree of belief that *someone* wins will obviously be 1. We would argue that these are reasonable answers. It is true that we do not get the default conclusion that c does not win (i.e., degree of belief 0). But since our probabilistic framework can and does express the conclusion that c is very unlikely to win, this is not a serious problem (unlike in systems which either reach a default conclusion or not, with no possibilities in between).

If we do not know the exact number of ticket holders, but have only the qualitative information that this number is “large”, then our degree of belief that c wins the lottery is simply $\Pr_\infty(\mathit{Winner}(c)|\mathit{KB} \land \mathit{Ticket}(c)) = 0$, although, as before, $\Pr_\infty(\exists x\ \mathit{Winner}(x)|\mathit{KB}) = 1$. In this case we do conclude by default that any particular individual will not win, although we still have degree of belief 1 that someone does win. This shows that the tension Lifschitz sees between concluding a fact for any particular individual and yet not concluding the universal does in fact have a solution in a probabilistic setting such as ours.

Finally, we consider where random worlds fits into Poole's analysis of the lottery paradox. Recall, his argument concentrated on examples in which a class (such as $\mathit{Bird}(x)$) is known to be equal to the union of a number of subclasses ($\mathit{Penguin}(x), \mathit{Emu}(x), \ldots$), each of which is exceptional in at least one respect. However, using our statistical interpretation of defaults, “exceptional” implies “makes up a negligible fraction of the population”. So under our interpretation, Poole's example is inconsistent: we cannot partition the set of birds into a finite number of subclasses, each of which makes up a negligible fraction of the whole set. We view the inconsistency in this case as a feature: it alerts the user that this collection of facts cannot all be true of the world (given our interpretation of defaults), just as would the inconsistency of the default “Birds typically fly” with “Birds typically do not fly” or “No bird flies”.

Our treatment of Poole's example clearly depends on our interpretation of defaults. For instance, we could interpret a default statement such as “Birds typically fly” as $\|\mathit{Fly}(x)|\mathit{Bird}(x)\|_x \succeq \alpha$ for some appropriately chosen α which is less than 1. In this case, “exceptional” subclases (such as penguins which are nonflying birds) *can* include a nonvanishing fraction of the domain. While allowing an interpretation of default not based on “almost all” does make Poole's KB consistent, it entails giving up many of the attractive properties of the ≈ 1 representation (such as having default conclusions assigned a degree of belief 1, and the associated properties described in Section 4.1). An alternative solution would be to use a reasoning system such as the one presented in [KH92]. Such a system could interpret defaults as "almost all" whenever such an interpretation is consistent (and get the benefits associated with this interpretation), yet allow for such inconsistencies when they occur. In case of inconsistency, the system automatically generates the set of "possible" α's (those that prevent inconsistency) for the different rules used.

We conclude this section be remarking on another property of the random-worlds method. Applications of default reasoning are often simplified by using the *unique names* assumption, which says that any two constants should (but perhaps only by default) denote different objects. In random worlds, there is a strong *automatic* bias towards unique names. If c_1 and c_2 are not mentioned anywhere in KB, then $\Pr_\infty(c_1 = c_2 | KB) = 0$ (see Lemma D.4.1 for a formal proof of this fact). Of course, when we know something about c_1 and c_2 it is possible to find examples for which this result fails; for instance $\Pr_\infty(c_1 = c_2 | (c_1 = c_2) \vee (c_2 = c_3) \vee (c_1 = c_3)) = \frac{1}{3}$. It is hard to give a general theorem saying precisely when the bias towards unique names overrides other considerations. However, we note that both of the "benchmark" examples that Lifschitz has given concerning unique names [Lif89] are correctly handled by random-worlds. For instance, Lifschitz's problem C1 is:

1. Different names normally denote different people.
2. The names "Ray" and "Reiter" denote the same person.
3. The names "Drew" and "McDermott" denote the same person.

The desired conclusion here is:

- The names "Ray" and "Drew" denote different people.

Random worlds gives us this conclusion. That is,

$$\Pr_\infty(\mathit{Ray} \neq \mathit{Drew} | \mathit{Ray} = \mathit{Reiter} \wedge \mathit{Drew} = \mathit{McDermott}) = 1.$$

Furthermore, we do not have to state the unique names default explicitly.

Chapter 5

Non-Unary Knowledge Bases

5.1 Introduction

In previous chapters, we defined the random-worlds method and investigated its properties. Among other things, our results allow us to compute the degree of belief in certain cases. More precisely, for KB's and φ's having certain properties, the theorems in the previous chapter can sometimes be used to compute $\Pr_\infty(\varphi|KB)$. However, these theorems do not provide a general technique for computing degrees of belief. For most of the remainder of this thesis, we will investigate the general problem of computing degrees of belief.

In this chapter, we investigate this problem for the case where φ and KB are both first-order sentences. While this is a severe restriction on the language, we will see that the results for this case provide a lot of insight on the general problem. In particular, in this chapter we demonstrate a number of serious problems that arise when attempting to compute asymptotic conditional probabilities for first-order sentences. The same problems will certainly arise in the more general case of the statistical language.

In the first-order case, our work is closely related to the work on 0-1 laws for first-order logic. In fact, precisely the same definition of asymptotic probability is used in both frameworks, except that in the context of 0-1 laws, there is no conditioning on a knowledge base of prior information. The original 0-1 law, proved independently by Glebskiĭ et al. [GKLT69] and Fagin [Fag76], states that the asymptotic probability of any first-order sentence φ with no constant or function symbols is either 0 or 1. Intuitively, such a sentence is true in almost all finite structures, or in almost none.

The random-worlds method for the first-order case differs from the original work on 0-1 laws in two respects. The first is relatively minor: we need to allow the use of constant symbols in φ, as they are necessary when discussing individuals (such as patients). Although this is a minor change, it is worth observing that it has a significant impact: It is easy to see that once we allow constant symbols, the asymptotic probability of a sentence φ is no longer either 0 or 1; for example, the asymptotic probability of $P(c)$ is $\frac{1}{2}$. The more significant difference, however, is that we are interested in the asymptotic *conditional* probability of φ, given the knowledge base KB. That is, we want the probability of φ over the class of finite structures defined by KB.

Some work has already been done on aspects of this question. Fagin [Fag76] and Liogon'kiĭ [Lio69] independently showed that asymptotic conditional probabilities do not necessarily converge to any limit. Subsequently, 0-1 laws were proved for special classes of first-order structures (such as graphs, tournaments, partial orders, etc.; see the overview paper [Com88] for details and further references). In many cases, the classes considered could be defined in terms of first-order constraints. Thus, these results can be viewed as special cases of the problem that we are interested in: computing asymptotic *conditional* probabilities relative to structures satisfying the constraints of a knowledge base. Lynch [Lyn80] showed that asymptotic probabilities exist for first-order sentences involving unary functions, although there is no 0-1 law. (Recall that the original 0-1 result is specifically for first-order logic *without* function symbols.) This can also be viewed as a special case of an asymptotic conditional probability for first-order logic without functions, since we can replace the unary functions by binary predicates, and condition on the fact that they are functions.

The most comprehensive work on this problem is the work of Liogon'kiĭ [Lio69]. In addition to pointing out that asymptotic conditional probabilities do not exist in general, he shows that it is undecidable whether such a probability exists. He then investigates the special case of conditioning on formulas involving unary predicates only (but no equality). In this case, he proves that the asymptotic conditional probability does exist and can be effectively computed, even if the left side of the conditional has predicates of arbitrary arity and equality. In this chapter, we examine the case of conditioning on a non-unary knowledge base. The other case, where the knowledge base is assumed to be unary, is investigated in depth in the next chapter. As we explain in Chapter 7, the unary case is very important for our application.

We extend the results of [Lio69] for the non-unary case in a number of ways. We first show, in Section 5.3, that under any standard weakening of the concept of limit, asymptotic conditional probabilities still do not exist. We define three independent questions related to the asymptotic conditional probability: deciding whether it is well-defined (i.e., is there an infinite sequence of probabilities $\Pr_N(\varphi|KB)$ to take the limit over); deciding whether it exists, given that it is well-defined; and computing or approximating it, given that it exists. We show in Section 5.4 that all three problems are undecidable, and precisely characterize the degree of their undecidability. These results are based on the enormous expressivity of even a single binary predicate. They therefore continue to hold for many quite restrictive sublanguages of first-order logic. We then present one "positive" result: In perhaps the most restrictive sublanguage that is still of any interest, if there is a fixed, finite vocabulary, and the quantifier depths of φ and KB are bounded, there is a linear time algorithm that computes the asymptotic conditional probability of φ given KB. Moreover, for each fixed vocabulary and fixed bound on quantifier depth, we can construct a finite set of algorithms, one of which is guaranteed to be one that solves the problem. However, it follows from our undecidability results that we cannot tell which algorithm is the correct one. So even this result holds no real promise.

5.2 Technical Preliminaries

Recall that in Chapter 3, we defined $\Pr_N^{\vec{\tau}}(\varphi|KB)$ relative to a particular tolerance vector $\vec{\tau}$, and $\Pr_\infty(\varphi|KB)$ using a limit as $\vec{\tau}$ goes to 0. In this chapter and in the next, we restrict attention to a first-order language with no statistical statements. Let $\mathcal{L}(\Phi)$ denote the first-order fragment of our language: the set of first-order sentences over $\Phi \cup \{=\}$. Let $\mathcal{L}^-(\Phi)$ denote the set of first-order sentences over Φ, i.e., without equality. For sentences in $\mathcal{L}(\Phi)$, the tolerance vector clearly plays no role. Therefore, for the purposes of this chapter and Chapter 6, we eliminate the tolerance vector from consideration. Thus, $\Pr_N(\varphi|KB)$ is defined to be

$$\frac{\#\mathit{worlds}_N^\Phi(\varphi \wedge KB)}{\#\mathit{worlds}_N^\Phi(KB)}.$$

As we observed, this probability is not well-defined if $\#\mathit{worlds}_N^\Phi(KB) = 0$. In the previous chapters, we chose to circumvent this problem by assuming eventual consistency of the knowledge base. Liogon'kiĭ, on the other hand, simply takes $\Pr_N(\varphi|KB) = 1/2$ for those N where $\Pr_N(\varphi|KB)$ is not well-defined. In this chapter, we are interested in investigating the problems that arise in the random-worlds framework. We therefore take a somewhat more refined approach.

It might seem reasonable to say that the asymptotic probability is not well-defined if $\#\mathit{worlds}_N^\Phi(KB) = 0$ for infinitely many N. However, suppose that KB is a sentence that is satisfiable only when N is even and, for even N, $\varphi \wedge KB$ holds in one third of the models of KB. In this case, we might want to say that there is an asymptotic conditional probability of 1/3, even though $\#\mathit{worlds}_N^\Phi(KB) = 0$ for infinitely many N. Thus, we actually consider two notions: the persistent limit, denoted $\Diamond\Box\Pr_\infty(\varphi|KB)$, and the intermittent limit, denoted $\Box\Diamond\Pr_\infty(\varphi|KB)$ (the prefixes stand for the temporal logic representation of the persistence and intermittence properties [MP92]). In either case, we say that the limiting probability is either not well-defined, does not exist, or is some number between 0 or 1. The only difference between the two notions lies in when the limiting probability is taken to be well-defined. This difference is made precise in the following definition.

Definition 5.2.1: Let $\mathcal{N}(KB)$ denote the set $\{N : \#\mathit{worlds}_N^\Phi(KB) \neq 0\}$. The asymptotic conditional probability $\Diamond\Box\Pr_\infty(\varphi|KB)$ is *well-defined* if $\mathcal{N}(KB)$ contains all but finitely many N's; $\Box\Diamond\Pr_\infty(\varphi|KB)$ is *well-defined* if $\mathcal{N}(KB)$ is infinite.

If the asymptotic probability $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) is well-defined, then we take $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) to denote

$$\lim_{N\to\infty, N\in\mathcal{N}(KB)} \Pr_N(\varphi|KB). \quad \blacksquare$$

Recall that for a first-order language, we have no approximate equality statements. In this case, we can ignore tolerance vectors and the outer limit.

Note that for any formula φ, the issue of whether $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$ or $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB)$ is well-defined is completely determined by KB. Therefore, when investigating the question of how to decide whether such a probability is well-defined it is often useful to ignore φ. We therefore say that $\Box\Diamond\mathrm{Pr}_\infty(*|KB)$ (resp., $\Diamond\Box\mathrm{Pr}_\infty(*|KB)$) is *well-defined* if $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$ (resp., $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB)$) is well-defined for every formula φ (which is true iff $\Diamond\Box\mathrm{Pr}_\infty(\mathit{true}|KB)$ (resp., $\Box\Diamond\mathrm{Pr}_\infty(\mathit{true}|KB)$) is well-defined).

Remark 5.2.2:

(a) If $\Diamond\Box\mathrm{Pr}_\infty(*|KB)$ is well-defined, then so is $\Box\Diamond\mathrm{Pr}_\infty(*|KB)$. The converse is not necessarily true.

(b) For any formula φ, if both $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB)$ and $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$ are well-defined, then they are equal.[1] ▮

It follows from our later results that the two notions of limiting probability coincide if we restrict to unary predicates or to languages without equality.

5.3 Nonexistence results

In this section, we show that the limiting probability $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$ (and hence $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB)$) does not always exist. In fact, for most reasonable concepts of limit (including, for example, the Cesàro limit), there are sentences for which the sequence $\mathrm{Pr}_N(\varphi|KB)$ does not converge.

5.3.1 Nonexistence for conventional limits

As we mentioned above, the fact that asymptotic conditional probabilities do not always exist is well known.

Theorem 5.3.1: [Lio69, Fag76] *Let Φ be a vocabulary containing at least one non-unary predicate symbol. There exist sentences $\varphi, KB \in \mathcal{L}(\Phi)$ such that neither $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$ nor $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB)$ exists, although both are well-defined.*

Proof: Fagin's proof of this theorem is quite straightforward (see also Figure 5.1). Let R be a binary predicate in Φ (although, clearly, any non-unary predicate will do). Using R and equality, it is not hard to construct sentences φ_{even} and φ_{odd} such that:

- φ_{even} and φ_{odd} both force R to be a symmetric antireflexive binary relation that divides the domain elements into pairs, where i, j is a pair precisely when $R(i, j)$. Both φ_{even} and φ_{odd} force each element to be paired up with at most one other element.

[1] When we say that two limits are equal, we mean that one is well-defined iff the other is, one exists iff the other does, and if they (both) exist then they are equal.

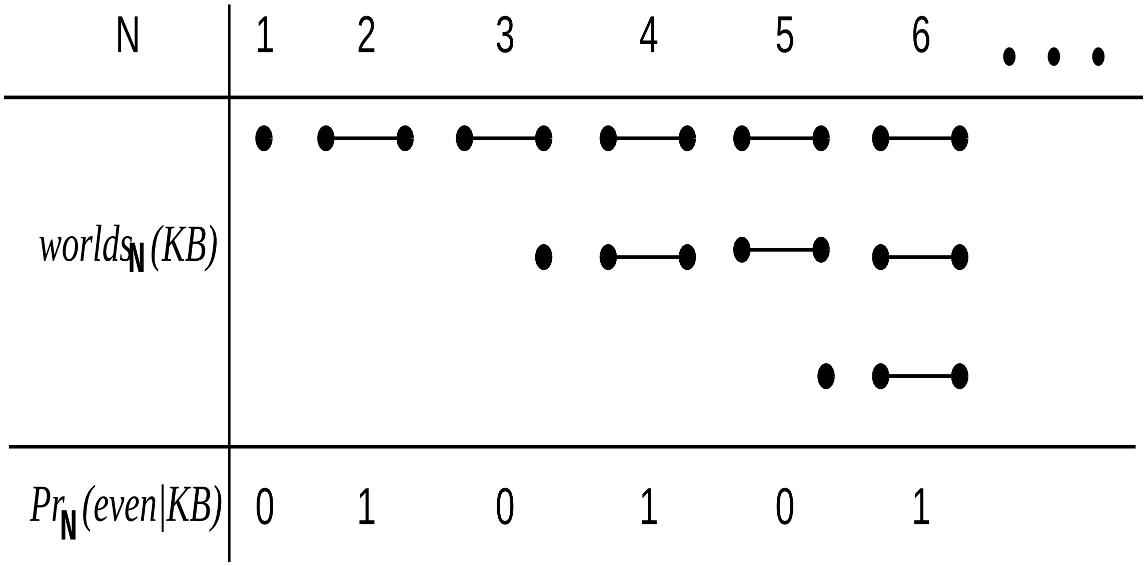

Figure 5.1: Construction demonstrating non-existence of limits

- φ_{even} forces the pairing to be complete; that is, each element is paired up with precisely one domain element. It is clear that φ_{even} is satisfiable if and only if the domain size is even.

- φ_{odd} forces the pairing to be almost-complete; that is, all elements but one are perfectly paired. It is clear that φ_{odd} is satisfiable if and only if the domain size is odd.

We then take φ to be φ_{odd} and KB to be $\varphi_{even} \vee \varphi_{odd}$. Clearly, $\Pr_N(\varphi|KB)$ alternates between 0 and 1 as N increases, and does not approach an asymptotic limit. ▮

Although this shows that the asymptotic limit does not exist in general, a good argument can be made that in this case there is a reasonable degree of belief that one can hold. In the absence of any information about domain size, $\frac{1}{2}$ seems the natural answer. Perhaps if we modified our definition of asymptotic probability slightly, we could increase the applicability of our techniques.

There is indeed a reasonable modification that will let us assign a degree of belief of $\frac{1}{2}$ in this case: we can use the Cesàro limit instead of the conventional limit.[2] The Cesàro limit of a sequence $s_1, s_2, \ldots$ is the conventional limit of the sequence s_1, $(s_1+s_2)/2$, $(s_1+s_2+s_3)/3, \ldots$, whose kth element is the average of the first k elements of the original sequence. It is well known that if the conventional limit exists, then so does the Cesàro limit, and they are equal. However, there are times when the Cesàro limit exists and the conventional limit does not. For

[2] We remark that Cesàro limits have been used before in the context of 0-1 laws; see Compton's overview [Com88] for details and further references.

example, for a sequence of the form $1, 0, 1, 0, \ldots$ (which, of course, is precisely the sequence that arises in the proof of Theorem 5.3.1), the conventional limit does not exist, but the Cesàro limit does, and is $\frac{1}{2}$. So does the Cesàro limit always exist? In the next section, we show that, unfortunately, this is not the case. In fact, no other reasonable notion of limit can solve the nonexistence problem.

5.3.2 Weaker limits

Fagin's non-existence example in Theorem 5.3.1 was based on a sequence $\Pr_N(\varphi|KB)$ that consistently alternated between 0 and 1. We have shown that using the Cesàro limit in place of the conventional limit when computing the limit of this sequence gives us the plausible answer of $\frac{1}{2}$. This may lead us to hope that by replacing the conventional limit in our definition of asymptotic conditional probability, we can circumvent the nonexistence problem. Unfortunately, this is not the case. It is relatively easy to construct examples that show that even Cesàro limits of the conditional probabilities $\Pr_N(\varphi|KB)$ do not necessarily converge. In this section, we will prove a far more general theorem. Essentially, the theorem shows that no reasonable notion of limit will ensure convergence in all cases. We begin by describing the general framework that allows us to formalize the notion of "reasonable notion of limit".

The Cesàro limit is only one of many well-studied *summability* techniques that weaken the conventional definition of convergence for infinite sequences.[3] These are techniques which try to assign "limits" to sequences that do not converge in the conventional sense. There is a general framework for summability techniques, which we now explain. (See, for example, [PS72] for further details.)

Let $A = (a_{ij})$ be an infinite square matrix; that is, a_{ij} is a (possibly complex) number for each pair of natural numbers i, j. Let $(s_i) = s_1, s_2, s_3, \ldots$ be an infinite sequence. Suppose that, for all i, the series $\sum_{j=1}^{\infty} a_{ij}s_j$ converges, say to sum S_i. Then the new sequence (S_i) is called the *A-transform* of (s_i). The idea is that (S_i) may converge to a limit, even if (s_i) does not. The standard notion of limit can be obtained by taking $a_{ii} = 1$ and $a_{ij} = 0$ if $i \neq j$. The Cesàro limit can be obtained by taking $a_{ij} = 1/i$ if $j \leq i$, and $a_{ij} = 0$ otherwise.[4]

Not every transform makes intuitive sense as a weakened notion of convergence. It would seem reasonable to require, at the very least, the following conditions of a matrix transform A.

- *Computability.* There should be a recursive function f such that $f(i, j)$ is the entry a_{ij} of the matrix A. It is difficult to see how we could actually use a transform whose elements could not be effectively computed.

[3] Summability theory is so named because one application is to find a way of assigning a "sum" to series that are divergent according to the conventional notion of limit. However, the theory addresses the problem of convergence for any sequence, whether or not it arises naturally as a sequence of partial sums.

[4] One subtle problem concerning our application of summability transforms is that some terms in the sequence $\Pr_N(\varphi|KB)$ may not exist. Throughout the following, we adopt perhaps the simplest solution to this difficulty, which is to apply the transform to the subsequence generated by just those domain sizes for which the probability exists (i.e., for which KB is satisfiable).

- *Regularity.* If a sequence converges (in the conventional sense), say to limit ℓ, then the A-transform should exist and converge to ℓ. This ensures that we really do obtain a more general notion of convergence.

The regularity condition has been well studied. The following three conditions are known to be necessary and sufficient for A to be regular. (This result is known as the Silverman-Toeplitz theorem; see [PS72].)

R1. $\lim_{i\to\infty} a_{ij} = 0$, for all j,

R2. $\lim_{i\to\infty} \sum_{j=1}^{\infty} a_{ij} = 1$, and

R3. there exists M such that $\sum_{j=1}^{\infty} |a_{ij}| < M$, for all i.

In our setting—where the motivation is assigning degrees of belief—we can give an fairly intuitive interpretation to many regular summability methods. Fix a value for i and suppose that (1) for all j, a_{ij} is real and nonnegative, and (2) $\sum_{j=1}^{\infty} a_{ij} = 1$. Then, for any i, the sequence $a_{i1}, a_{i2}, \ldots$ can also be viewed as a probability distribution over possible domain sizes. Given that one accepts the basic random-worlds framework for assigning degrees of belief relative to a particular domain size, it seems plausible that $\sum_{N=1}^{\infty} a_{iN}\Pr_N(\varphi|KB)$ should be one's degree of belief in φ given KB, if the uncertainty about the correct domain size is captured by $a_{i1}, a_{i2}, \ldots$ (and $\Pr_N(\varphi|KB)$ defined for all finite N). For example, row i of the Cesàro matrix would be appropriate for someone who knows for certain that there are i or less individuals, but subject to this assigns equal degree of belief to each of the i possibilities. However, no single distribution over the natural numbers seems to accurately model the situation where *all* we know is that "the domain size is large." For one thing, any distribution gives nonzero probability to particular domain sizes, which seems to involve some commitment to scale. Instead, we can consider a sequence of distributions, such that the degree of belief in any particular domain size tends to zero. Constructions such as this always satisfy conditions R1–R3, and thus fall into the framework of regular transforms. In fact, almost all summability transforms considered in the literature are regular transforms. The main result of this section is that no summability technique covered by this framework can guarantee convergence for asymptotic conditional probabilities. This is so even if the vocabulary consists of a single binary predicate symbol.

Theorem 5.3.2: *Let A be any computable regular matrix transform, and let Φ be a vocabulary containing at least one non-unary predicate symbol. There exist $\varphi, KB \in \mathcal{L}(\Phi)$ such that the A-transform of the sequence $\Pr_N(\varphi|KB)$ exists, but does not converge.*

The proof of this theorem is based on the fact that even a single binary predicate is extremely expressive. In fact, it is well-known that, using a binary predicate symbol, we can use first-order sentences, interpreted over finite domains, to encode (arbitrarily long) prefixes of the computation of a deterministic Turing machine (see [Tra50]). That is, given a Turing machine $\mathbf{M}$, we can define a sentence $KB_{\mathbf{M}}$ such that any finite model satisfying $KB_{\mathbf{M}}$ encodes a finite prefix of the computation of $\mathbf{M}$ on empty input. The exact construction is fairly standard, but requires many details; we present an outline in Appendix B.1. This construction forms the basis for the proof of this theorem, which can be found in Appendix B.2.

This result is a very powerful one, that covers all notions of limit of which we are aware. This is the case in spite of the fact that there are a few well-known notions of limit which are not, strictly speaking, matrix transforms. Nevertheless, our theorem is applicable to these cases as well. The best example of this is *Abel convergence.* A sequence (s_j) is said to be Abel convergent if $\lim_{x \to 1^-} (1 \Leftrightarrow x) \sum_{j=1}^{\infty} s_j \, x^{(j-1)}$ exists. This is not a matrix transform, because we must consider all sequences of x that tend to 1. However, consider any particular sequence of rationals that converges to 1, say

$$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, \frac{i}{i+1}, \ldots$$

We can use these to define a matrix variant of the Abel method, by setting

$$a_{ij} = \left(1 \Leftrightarrow \frac{i}{i+1}\right) \left(\frac{i}{i+1}\right)^{j-1} .$$

This is regular and computable, and is strictly weaker than the standard Abel method. More precisely, if the Abel limit converges, then so does this matrix transform. Since our theorem shows that this new summability method does not ensure convergence for conditional probabilities, this is automatically also the case for the Abel limit.

5.4 Undecidability results

We have seen that asymptotic conditional probabilities do not always exist. We might hope that at least we can easily decide when they do exist, so that we would know when the random-worlds method is applicable. As we show in this section, this hope is not realized. In this section, we show the undecidability of several important problems associated with asymptotic conditional probabilities: deciding whether the limit is well-defined, deciding whether the limit exists, and giving some nontrivial approximation to its value (deciding whether it lies in some non-trivial interval). Liogon'kiĭ [Lio69] showed that the problem of computing the asymptotic conditional probability for the random-worlds method is undecidable. He did not consider other problems, nor did he characterize the degree of undecidability of the problem. Our undecidability results all rely on the Turing machine construction in Appendix B.1, and use a fixed finite vocabulary, consisting of equality and a single binary predicate. The proofs of the results can be found in Appendix B.3. Most of them can, in fact, be translated to a language without equality, at the cost of adding two more binary predicates (see Section 5.5).

We analyze the complexity of these problems in terms of the *arithmetic hierarchy.* This is a hierarchy that extends the notions of r.e. (recursively enumerable) and co-r.e. sets. We briefly review the relevant definitions here, referring the reader to [Rog67] for further details. Consider a formula ξ in the language of arithmetic (i.e., using $0, 1, +, \times$) having j free variables. The formula ξ, interpreted over the natural numbers, is said to define a *recursive set* if the set of j-tuples satisfying the formula is a recursive set. We can define more complex sets using quantification. We define a Σ_k^0 *prefix* as a block of quantifiers of the form $\exists x_1 \ldots x_h \forall y_1 \ldots y_m \ldots$, where there are k alternations of quantifiers (but there is no restriction on the number of quantifiers of the same type that appear consecutively). A Π_k^0 prefix is defined similarly, except that the quantifier block starts with a universal quantifier. A set A of natural numbers is in Σ_k^0 if there is a first-order formula $\xi(x) = Q\xi'$ in the language of arithmetic with one free variable x, where Q is a Σ_k^0 quantifier block and ξ' defines a recursive set, such that $n \in A$ iff $\xi(n)$ is true. We can similarly define what it means for a set to be in Π_k^0. The notion of *completeness* for these classes is defined in a standard fashion, using recursive reductions between problems. A set is in Σ_1^0 iff it is r.e., and it is in Π_1^0 iff it is co-r.e. The hierarchy is known to be strict; higher levels of the hierarchy correspond problems which are strictly harder ("more undecidable").

We start with the problem of deciding if the asymptotic probability is well-defined; this is certainly a prerequisite for deciding whether the limit exists. Of course, this depends in part on which definition of well-definedness we use.

Theorem 5.4.1: *Let Φ be a vocabulary containing at least one non-unary predicate symbol.*

(a) The problem of deciding whether a sentence in $\mathcal{L}(\Phi)$ is satisfiable for infinitely many domain sizes is Π_2^0-complete.

(b) The problem of deciding whether a sentence in $\mathcal{L}(\Phi)$ is satisfiable for all but finitely many domain sizes is Σ_2^0-complete.

Corollary 5.4.2: *Let Φ be a vocabulary containing at least one non-unary predicate symbol. For $KB \in \mathcal{L}(\Phi)$, the problem of deciding whether $\Box\Diamond\Pr_\infty(*|KB)$ is well-defined is Π_2^0-complete, and the problem of deciding whether $\Diamond\Box\Pr_\infty(*|KB)$ is well-defined is Σ_2^0-complete.*

If deciding well-definedness were the only difficulty in computing, then there might still be hope. In many cases, it might be obvious that the sentence we are conditioning on is satisfiable in all (or, at least, in infinitely many) domain sizes. As we are about to show, the situation is actually much worse. Deciding if the limit exists is even more difficult than deciding well-definedness; in fact, it is Π_3^0-complete.

Theorem 5.4.3: *Let Φ be a vocabulary containing at least one non-unary predicate symbol. For sentences $\varphi, KB \in \mathcal{L}(\Phi)$, the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) exists is Π_3^0-complete. The lower bound holds even if we have an oracle that tells us whether the limit is well-defined and its value if it exists.*

Even if we have an oracle that will tell us whether the conditional probability is well-defined and whether it exists, it is difficult to compute the asymptotic probability. Indeed, given any nontrivial interval (one which is not the interval $[0, 1]$), it is even difficult to tell whether the asymptotic probability is in the interval.

Theorem 5.4.4: *Let Φ be a vocabulary containing at least one non-unary predicate symbol, and let $r, r_1, r_2 \in [0,1]$ be rational numbers such that $r_1 \leq r_2$. For sentences $\varphi, KB \in \mathcal{L}(\Phi)$, given an oracle for deciding whether $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB)$ (resp., $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$) exists,*

(a) the problem of deciding whether $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB) = r$ (resp., $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB) = r$) is Π^0_2-complete,

(b) if $[r_1, r_2] \neq [0,1]$, then the problem of deciding whether $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB) \in [r_1, r_2]$ (resp., $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB) \in [r_1, r_2]$) is Π^0_2-complete,

(c) if $r_1 \neq r_2$, then the problem of deciding if $\Diamond\Box\mathrm{Pr}_\infty(\varphi|KB) \in (r_1, r_2)$ (resp., $\Box\Diamond\mathrm{Pr}_\infty(\varphi|KB)$ $\in (r_1, r_2)$) is Σ^0_2-complete.

5.5 Eliminating Equality

The proofs of all the negative results above depend on the Turing machine construction presented in Appendix B.1. It seems that this construction makes heavy use of equality, so that one might suspect that the problems disappear if we disallow equality. This is not the case. As we now show, we can eliminate the use of equality from most of these results, at the price of adding two more binary predicate symbols to the vocabulary. Intuitively, we add one predicate E to replace the equality predicate $=$, and one predicate G that is used to force E to behave like equality.

Theorem 5.5.1: *Suppose G and E are binary predicate symbols not appearing in Φ, and $\varphi, KB \in \mathcal{L}(\Phi)$ are such that $\#worlds^\Phi_N(KB)$ is a non-decreasing function of N. Then we can find sentences $\varphi', KB' \in \mathcal{L}^-(\Phi \cup \{G, E\})$ such that*

$$\lim_{N\to\infty} (\mathrm{Pr}_N(\varphi|KB) \Leftrightarrow \mathrm{Pr}_N(\varphi'|KB')) = 0 \ .$$

Using Theorem 5.5.1, we can show analogues to most of our results for the language with equality. First, we can immediately deduce the following corollary to Theorem 5.3.2.

Corollary 5.5.2: *Let A be any computable regular matrix transform, and let Φ be a vocabulary containing at least three non-unary predicate symbols. There exist $\varphi, KB \in \mathcal{L}(\Phi)$ such that the A-transform of the sequence $\mathrm{Pr}_N(\varphi|KB)$ exists, but does not converge.*

It is easy to show that similar analogues to most of the complexity results of this chapter also hold. The exceptions are Theorem 5.4.1 and Corollary 5.4.2.

For KB that does not use equality, $\Box\Diamond\mathrm{Pr}_\infty(*|KB)$ is well-defined iff $\Diamond\Box\mathrm{Pr}_\infty(*|KB)$ is well-defined iff KB is satisfiable for some model. This is true because if KB is satisfied in some model of size N, then it is also satisfied in some model of size N' for every $N' > N$. As a consequence, we can show:

Theorem 5.5.3: *Let Φ be a vocabulary containing at least two non-unary predicate symbols. For $KB \in \mathcal{L}^-(\Phi)$, the problem of deciding if $\Box\Diamond\Pr_\infty(*|KB)$ (resp., $\Diamond\Box\Pr_\infty(*|KB)$) is well-defined is r.e.-complete.*

In Appendix B.4 we formally state and prove the theorems asserting that the remaining complexity results do carry over.

5.6 Is there any hope?

These results show that most interesting problems regarding asymptotic probabilities are badly undecidable in general. Are there restricted sublanguages for which these questions become tractable, or at least decidable?

All of our negative results so far depend on having at least one non-unary predicate symbol in the vocabulary. In fact, it clearly suffices to have the non-unary predicate symbols appear only in KB. However, as we indicated in the introduction, this additional expressive power of KB is essential. If we restrict KB to refer only to unary predicates and constants, many of the problems we encounter in the general case disappear. This holds even if φ can refer to arbitrary predicates. In the next chapter, we focus on this important special case. Here, we consider one other case.

A close look at our proofs in the previous sections shows that we typically started by constructing sentences of low quantification depth, that use (among other things) an unbounded number of unary predicates. For example, the original construction of the sentences encoding computations of Turing machines used a unary predicate for every state of the machine. We then explained how to encode everything using only one binary predicate. In the process of doing this encoding, we had to introduce additional quantifiers (for example, an existential quantifier for every unary predicate eliminated). Thus, our undecidability results seem to require one of two things: an unbounded vocabulary (in terms of either the number of predicates or of their arity), or unbounded quantification depth. Do we really need both? It is actually easy to show that the answer is yes.

Definition 5.6.1: Define $d(\xi)$ to be the *depth of quantifier nesting* in the formula ξ:

- $d(\xi) = 0$ for any atomic formula ξ,
- $d(\neg\xi) = d(\xi)$,
- $d(\xi_1 \wedge \xi_2) = \max(d(\xi_1), d(\xi_2))$,
- $d(\forall y\, \xi) = d(\xi) + 1$. ▮

Let $\mathcal{L}_d(\Phi)$ consist of all sentences $\varphi \in \mathcal{L}(\Phi)$ such that φ has quantification depth at most d.

Theorem 5.6.2: *For every finite vocabulary Φ and every d, there exists a Turing machine $\mathbf{M}_d^\Phi$ such that for all $\varphi, KB \in \mathcal{L}_d(\Phi)$, $\mathbf{M}_d^\Phi$ decides in time linear in the length of φ and KB whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) is well-defined, if so whether it exists, and if it exists computes an arbitrarily good rational approximation to its value.*

The proof of this theorem, which can be found in Appendix B.5, is based on the fact that there are only finitely many non-equivalent formulas of depth $\leq d$ over a fixed vocabulary. That is, there exists a finite set Σ_d of formulas such that every formula in $\mathcal{L}_i^d(\Phi)$ is equivalent to some member of Σ_d. There exists a lookup table containing the answer for any pair of formulas φ', KB' in Σ_d. For any pair of formulas φ, KB, we can find their equivalents φ', KB' in Σ_d in linear time, and then simply use the lookup table. Since the size of the table is taken to be a constant, this last step can also be done in linear time (in the sizes of φ and KB).

This proof asserts that, for each d, there exist lookup tables that effectively determine the behavior of the asymptotic probability for sentences in $\mathcal{L}_d(\Phi)$. Moreover, it shows that we can effectively construct a finite set of lookup tables, one of which is bound to be the right one. Unfortunately, we cannot effectively determine which one is the right one, for if we could, we could effectively construct $\mathbf{M}_d^\Phi$ given Φ and d, and this would contradict our earlier undecidability results. Thus, even for this extremely restrictive sublanguage we cannot effectively construct algorithms for computing asymptotic conditional probabilities.

5.7 Discussion

The results in this chapter show that the applicability of the random-worlds method is not as wide as we might have hoped. There are problems for which random worlds will not be able to assign a degree of belief (because of nonexistence of the limit). Furthermore, deciding whether this is the case, and computing the degree of belief if it does exist, are highly undecidable.

However, as we mentioned, there is one interesting special case where these results do not hold. Liogon'kiĭ [Lio69] has shown that if the (first-order) knowledge base contains only unary predicate symbols, the asymptotic conditional probability does exist and can be effectively computed. In the next chapter, we extend these results significantly (although still within the first-order framework), and prove a number of related complexity results for the problem. In Chapter 7, we extend these results to deal with knowledge bases containing statistical statements. As we explain in these two chapters, the unary case is an important special case, covering many practical problems, especially those involving statistical information.

It is interesting to note that in [Car52], where Carnap considers a continuum of methods for inductive reasoning (which includes the random-worlds method), he considers only the unary case for all of them, without any comment or justification. He does provide some justification in [Car50], as well as expressing concern that the case of non-unary predicates may cause difficulties (although he presents no technical justification for this claim):

> ... the bulk of our inductive logic will deal only with properties of individuals [i.e., unary predicates], not with relations between individuals, except for those relations

> which are defined on the basis of properties. At the present time, this restriction seems natural and well justified, in view of the fact that deductive logic took more than two thousand years from its start with Aristotle to the first logic of relations (De Morgan, 1860). Inductive logic ... is only a few hundred years old. Therefore, it is not surprising to see that so far nobody has made an attempt to apply it to relations. ... The inclusion of relations in deductive logic causes obviously a certain increase in complexity. The corresponding increase in complexity for inductive logic is very much greater.

Carnap's allusion to the difficulty of adding relations to deductive logic is perhaps the observation — known at the time — that while first-order logic over a vocabulary with only unary predicate symbols is decidable, it becomes undecidable when we add non-unary predicates [DG79, Lew79]. The fact that there is an increase in complexity in inductive logic when we add non-unary predicates is not substantiated by Carnap, other than by the observation that very difficult combinatorial questions arise. As our results show, Carnap's concern about the difficulty of doing inductive reasoning with relations (non-unary predicates) is well founded.

Chapter 6

Unary First-Order Knowledge Bases

6.1 Introduction

In the previous chapter, we investigated the problem of computing asymptotic conditional probabilities for the case where φ and KB are arbitrary first-order formulas. The negative results we presented all depend on the fact that the vocabulary contains at least one non-unary predicate symbol. It is this non-unary predicate that gives the knowledge base the expressive power that causes non-existence and undecidability. In this chapter, we show that unary predicates do not have this expressive power. For a knowledge base containing only unary predicate symbols, constant symbols, and equality, the problems described in the previous chapter disappear. This is the case even if φ contains predicate symbols of arbitrary arity. In this chapter, we continue to concentrate on the case of first-order formulas. The issues concerning statistical statements are quite different, and are investigated in the next chapter. However, it turns out that the techniques we describe here will also help us in the case of the full language (with statistical statements).

As we remarked in the previous chapter, the issue of asymptotic conditional probabilities for first-order formulas has already been investigated by Liogon'kiĭ. His results for the unary case involve conditioning on formulas involving unary predicates only (but no constants or equality). For this case, he proves that the asymptotic conditional probability does exist and can be effectively computed, even if the left side of the conditional has predicates of arbitrary arity and equality. This gap between unary predicates and binary predicates is somewhat reminiscent of the fact that first-order logic over a vocabulary with only unary predicates (and constant symbols) is decidable, while if we allow even a single binary predicate symbol, then it becomes undecidable [DG79, Lew79]. This similarity is not coincidental; some of the techniques used to show that first-order logic over a vocabulary with unary predicate symbols is decidable are used by us to show that asymptotic probabilities exist.

In this chapter, we extend the results of Liogon'kiĭ [Lio69] for the unary case. We first show

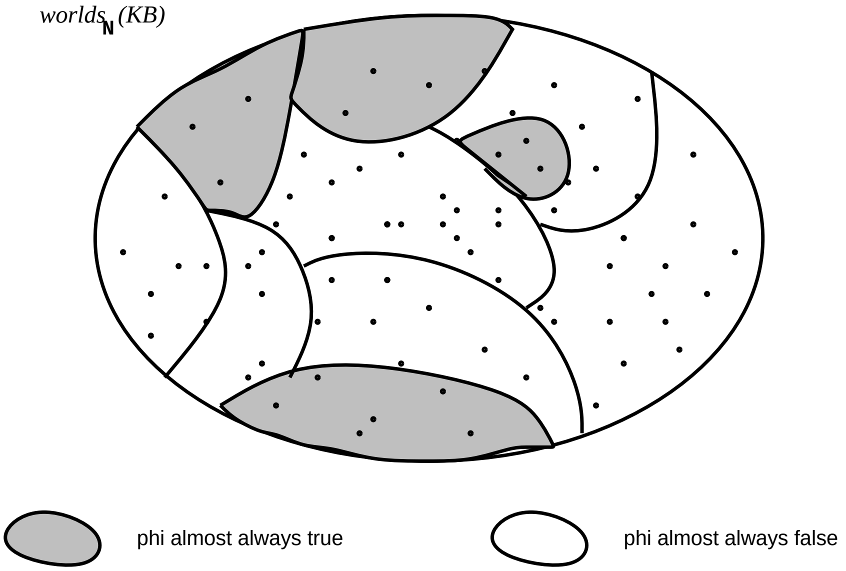

Figure 6.1: Division of the worlds into uniform classes

(in Section 6.3) that, if we condition on a formula involving only unary predicates, constants, and equality that is satisfiable in arbitrarily large worlds, the asymptotic conditional probability exists. We also present an algorithm for computing this limit. The main idea we use is the following: To compute $\Pr_\infty(\varphi|KB)$, we examine the behavior of φ in finite worlds of KB. It turns out that we can partition the worlds of KB into a finite collection of classes, such that φ behaves *uniformly* in any individual class. By this we mean that almost all worlds in the class satisfy φ or almost none do; i.e., there is a 0-1 law for the asymptotic probability of φ when we restrict attention to worlds in a single class (see Figure 6.1). Computing $\Pr_\infty(\varphi|KB)$ reduces to first identifying the classes, computing the relative weight of each class (which is required because the classes are not necessarily of equal relative size), and then deciding, for each class, whether the asymptotic probability of φ is zero or one.

In Section 6.2 we show how the lack of expressivity of a unary vocabulary allows us to define an appropriate finite collection of classes. In Section 6.3 we prove the existence of a 0-1 law within each class, and compute the relative weight of the classes. This allows us to give a formula for the asymptotic conditional probability of φ given KB, based on the 0-1 probabilities for the individual classes.

In Section 6.4 we turn our attention to the complexity of computing the asymptotic probability in this case. Our results depend on several factors: whether the vocabulary is finite or infinite, whether there is a bound on the depth of quantifier nesting, whether equality is used in KB, whether non-unary predicates are used, and whether there is a bound on predicate arities. For a fixed and finite vocabulary, there are just two cases: if there is no bound on the depth of quantifier nesting then computing asymptotic conditional probabilities is PSPACE-complete, otherwise the computation can be done in linear time. The case in which the vocabulary is not

	depth ≤ 1	**restricted**	**general case**
existence	NP-complete	NEXPTIME-complete	NEXPTIME-complete
compute	#P/PSPACE	#EXP-complete	#TA(EXP,LIN)-complete
approximate	(co-)NP-hard	(co-)NEXPTIME-hard	TA(EXP,LIN)-hard

Table 6.1: Complexity of asymptotic conditional probabilities

fixed (which is the case more typically considered in complexity theory) is more complex. The results for this case are summarized in Table 6.1. Perhaps the most interesting aspect of this table is the factors that cause the difference in complexity between #EXP and #TA(EXP,LIN). Here, #TA(EXP,LIN) is the counting class corresponding to alternating Turing machines that take exponential time and make only a linear number of alternations; a formal definition is provided in Section 6.4.6. If we allow the use of equality in *KB*, then we need to restrict both φ and *KB* to using only unary predicates to get the #EXP upper bound. On the other hand, if *KB* does not mention equality, then the #EXP upper bound is attained as long as there is some fixed bound on the arity of the predicates appearing in φ. If we have no bound on the arity of the predicates that appear in φ, or if we allow equality in *KB* and predicates of arity 2 in φ, then the #EXP upper bound no longer holds, and we move to #TA(EXP,LIN).

Our results showing that computing the asymptotic probability is hard can be extended to show that finding a nontrivial estimate of the probability (i.e., deciding if it lies in a nontrivial interval) is almost as difficult. The lower bounds for arity-bounded case and the general case require formulas of quantification depth 2 or more. For unquantified sentences or depth one quantification, things seem to become an exponential factor easier. We do not have tight bounds for the complexity of computing the degree of belief in this case; we have a #P lower bound and a PSPACE upper bound.

We observe that apart from our precise classification of the complexity of these problems, our analysis provides an effective algorithm for computing the asymptotic conditional probability. The complexity of this algorithm is, in general, double-exponential in the number of unary predicates used and in the maximum arity of any predicate symbol used; it is exponential in the overall size of the vocabulary and in the lengths of φ and *KB*.

6.2 Unary expressivity

The success of the approach outlined above depends on the lack of expressivity of unary languages. For a vocabulary Φ, we take $\mathcal{P}$ to be the set of all unary predicates in Φ, $\mathcal{C}$ to be the set of all constant symbols in Φ, and define $\Psi = \mathcal{P} \cup \mathcal{C}$ to be the unary fragment of Φ. Finally, if φ is a formula, we use Φ_φ to denote those symbols in Φ that appear in φ; we can similarly define $\mathcal{C}_\varphi$, $\mathcal{P}_\varphi$, and Ψ_φ.

In this section we show that sentences in $\mathcal{L}(\Psi)$ can only assert a fairly limited class of constraints. For instance, one corollary of our general result will be the well-known theorem [DG79] that, if $KB \in \mathcal{L}(\Psi)$ is satisfiable at all, it is satisfiable in a "small" model (one of size at most exponential in the size of the KB). Furthermore, if it is satisfiable in a "large" model, then it is satisfiable in every large model. This last fact allows us to considerably simplify the definition of well-definedness used in the previous chapter. There, we differentiated between the case where $\Pr_N(\varphi|KB)$ is well-defined for all but finitely many N's, and the case where it is well-defined for infinitely many N's. As we have just claimed (and will prove later in this chapter), this distinction need not be made when KB is a unary formula. Thus, for the purposes of this chapter, we use the following definition of well-definedness, which is simpler than that of the previous chapter.

Definition 6.2.1: The asymptotic conditional probability according to the random worlds method, denoted $\Pr_\infty(\varphi|KB)$, is *well-defined* if $\#worlds_N^\Phi(KB) \neq 0$ for all but finitely many N. ▮

6.2.1 Atomic descriptions

In order to analyze the expressivity of a unary formula, a number of definitions are necessary.

Definition 6.2.2: Given a vocabulary Φ and a finite set of variables $\mathcal{X}$, a *complete description* D over Φ and $\mathcal{X}$ is an unquantified conjunction of formulas such that:

- For every predicate $R \in \Phi \cup \{=\}$ of arity m, and for every $z_1, \ldots, z_m \in \mathcal{C} \cup \mathcal{X}$, D contains exactly one of $R(z_1, \ldots, z_m)$ or $\neg R(z_1, \ldots, z_m)$ as a conjunct.
- D is consistent.[1] ▮

We can think of a complete description as being a formula that describes as fully as possible the behavior of the predicate symbols in Φ over the constant symbols in Φ and the variables in $\mathcal{X}$.

We can also consider complete descriptions over subsets of Φ. The case when we look just at the unary predicates and a single variable x will be extremely important:

Definition 6.2.3: Let $\mathcal{P}$ be $\{P_1, \ldots, P_k\}$. An *atom* over $\mathcal{P}$ is a complete description over $\mathcal{P}$ and some variable $\{x\}$. More precisely, it is a conjunction of the form $P_1'(x) \wedge \ldots \wedge P_k'(x)$, where each P_i' is either P_i or $\neg P_i$. Since the variable x is irrelevant to our concerns, we typically suppress it and describe an atom as a conjunction of the form $P_1' \wedge \ldots \wedge P_k'$. ▮

Note that there are $2^k = 2^{|\mathcal{P}|}$ atoms over $\mathcal{P}$, and that they are mutually exclusive and exhaustive. We use $A_1, \ldots, A_{2^{|\mathcal{P}|}}$ to denote the atoms over $\mathcal{P}$, listed in some fixed order. For example, there are four atoms over $\mathcal{P} = \{P_1, P_2\}$: $A_1 = P_1 \wedge P_2$, $A_2 = P_1 \wedge \neg P_2$, $A_3 = \neg P_1 \wedge P_2$, $A_4 = \neg P_1 \wedge \neg P_2$.

We now want to define the notion of *atomic description* which is, roughly speaking, a maximally expressive formula in the unary vocabulary Ψ. Fix a natural number M. An atomic

[1] Inconsistency is possible because of the use of equality. For example, if D includes $z_1 = z_2$ as well as both $R(z_1, z_3)$ and $\neg R(z_2, z_3)$, it is inconsistent.

description of size M consists of two parts. The first part, the *size description with bound M*, specifies exactly how many elements in the domain should satisfy each atom A_i, except that if there are M or more elements satisfying the atom it only expresses that fact, rather than giving the exact count. More formally, given a formula $\xi(x)$ with a free variable x, we take $\exists^m x\, \xi(x)$ to be the sentence that says there are precisely m domain elements satisfying ξ:

$$\exists^m x\, \xi(x) =_{\text{def}} \exists x_1 \ldots x_m \left(\bigwedge_i (\xi(x_i) \wedge \bigwedge_{j \neq i} (x_j \neq x_i)) \wedge \forall y(\xi(y) \Rightarrow \vee_i (y = x_i)) \right).$$

Similarly, we define $\exists^{\geq m} x\, \xi(x)$ to be the formula that says that there are at least m domain elements satisfying ξ:

$$\exists^{\geq m} x\, \xi(x) =_{\text{def}} \exists x_1 \ldots x_m \left(\bigwedge_i (\xi(x_i) \wedge \bigwedge_{j \neq i} (x_j \neq x_i)) \right).$$

Definition 6.2.4: A *size description with bound M (over $\mathcal{P}$)* is a conjunction of $2^{|\mathcal{P}|}$ formulas: for each atom A_i over $\mathcal{P}$, it includes either $\exists^{\geq M} x\, A_i(x)$ or a formula of the form $\exists^m x\, A_i(x)$ for some $m < M$. ∎

The second part of an atomic description is a complete description that specifies the properties of constants and free variables.

Definition 6.2.5: A *size M atomic description* (over Φ and $\mathcal{X}$) is a conjunction of:

- a size description with bound M over $\mathcal{P}$, and
- a complete description over Ψ and $\mathcal{X}$. ∎

Note that an atomic description is a finite formula, and there are only finitely many size M atomic descriptions over Ψ and $\mathcal{X}$ (for fixed M). For the purposes of counting atomic descriptions (as we do in Section 6.3.2), we assume some arbitrary but fixed ordering of the conjuncts in an atomic description. Under this assumption, we cannot have two distinct atomic descriptions that differ only in the ordering of conjuncts. Given this, it is easy to see that atomic descriptions are mutually exclusive. Moreover, atomic descriptions are exhaustive—the disjunction of all consistent atomic descriptions of size M is valid.

Example 6.2.6: Consider the following size description σ with bound 4 over $\mathcal{P} = \{P_1, P_2\}$:

$$\exists^1 x\, A_1(x) \wedge \exists^3 x\, A_2(x) \wedge \exists^{\geq 4} x\, A_3(x) \wedge \exists^{\geq 4} x\, A_4(x).$$

Let $\Psi = \{P_1, P_2, c_1, c_2, c_3\}$. It is possible to augment σ into an atomic description in many ways. For example, one consistent atomic description ψ_* of size 4 over Ψ and $\emptyset$ (no free variables) is:[2]

$$\sigma \wedge A_2(c_1) \wedge A_3(c_2) \wedge A_3(c_3) \wedge c_1 \neq c_2 \wedge c_1 \neq c_3 \wedge c_2 = c_3.$$

[2] In our examples, we use the commutativity of equality in order to avoid writing down certain superfluous disjuncts. In this example, for instance, we do not write down both $c_1 \neq c_2$ and $c_2 \neq c_1$.

On the other hand, the atomic description

$$\sigma \wedge A_1(c_1) \wedge A_1(c_2) \wedge A_3(c_3) \wedge c_1 \neq c_2 \wedge c_1 \neq c_3 \wedge c_2 \neq c_3$$

is an inconsistent atomic description, since σ dictates that there is precisely one element in the atom A_1 whereas the second part of the atomic description implies that there are two distinct domain elements in that atom. ▮

As we explained, an atomic description is, intuitively, a maximally descriptive sentence over a unary vocabulary. The following theorem formalizes this idea by showing that each unary formula is equivalent to a disjunction of atomic descriptions. For a given M and set $\mathcal{X}$ of variables, let $\mathcal{A}^{\Psi}_{M,\mathcal{X}}$ be the set of consistent atomic descriptions of size M over Ψ and $\mathcal{X}$.

Theorem 6.2.7: *If ξ is a formula in $\mathcal{L}(\Psi)$ whose free variables are contained in $\mathcal{X}$, and $M \geq d(\xi) + |\mathcal{C}| + |\mathcal{X}|$,[3] then there exists a set of atomic descriptions $\mathcal{A}^{\Psi}_{\xi} \subseteq \mathcal{A}^{\Psi}_{M,\mathcal{X}}$ such that ξ is equivalent to $\bigvee_{\psi \in \mathcal{A}^{\Psi}_{\xi}} \psi$.*

The proof of this theorem can be found in Appendix C.1.

For the remainder of this chapter we will be interested in sentences. Thus, we restrict attention to atomic descriptions over Ψ and the empty set of variables. Moreover, we assume that all formulas mentioned are in fact sentences, and have no free variables.

Definition 6.2.8: For $\Psi = \mathcal{P} \cup \mathcal{C}$, and a sentence $\xi \in \mathcal{L}(\Psi)$, we define $\mathcal{A}^{\Psi}_{\xi}$ to be the set of consistent atomic descriptions of size $d(\xi) + |\mathcal{C}|$ over Ψ such that ξ is equivalent to the disjunction of the atomic descriptions in $\mathcal{A}^{\Psi}_{\xi}$. ▮

It will be useful for our later results to prove a simpler analogue of Theorem 6.2.7 for the case where the sentence ξ does not use equality or constant symbols. A *simplified atomic description over $\mathcal{P}$* is simply a size description with bound 1. Thus, it consists of a conjunction of $2^{|\mathcal{P}|}$ formulas of the form $\exists^{\geq 1} x\, A_i(x)$ or $\exists^{0} x\, A_i(x)$, one for each atom over $\mathcal{P}$. Using the same techniques as those of Theorem 6.2.7, we can prove:

Theorem 6.2.9: *If $\xi \in \mathcal{L}^{-}(\mathcal{P})$, then ξ is equivalent to a disjunction of simplified atomic descriptions over $\mathcal{P}$.*

6.2.2 Named elements and model descriptions

Recall that we are attempting to divide the worlds satisfying KB into classes such that:

- φ is uniform in each class, and

[3] Recall that $d(\xi)$ denotes the depth of quantifier nesting of ξ. See Definition 5.6.1.

- the relative weight of the classes is easily computed.

In the previous section, we defined the concept of atomic description, and showed that a sentence $KB \in \mathcal{L}(\Psi)$ is equivalent to some disjunction of atomic descriptions. This suggests that atomic descriptions might be used to classify models of KB. Liogon'kiĭ [Lio69] has shown that this is indeed a successful approach, as long as we consider languages without constants and condition only on sentences that do not use equality. In Theorem 6.2.9 we showed that, for such languages, each sentence is equivalent to the disjunction of simplified atomic descriptions. The following theorem, due to Liogon'kiĭ, says that classifying models according to which simplified atomic description they satisfy leads to the desired uniformity property. This result will be a corollary of a more general theorem that we prove later.

Proposition 6.2.10: [Lio69] *Suppose that $\mathcal{C} = \emptyset$. If $\varphi \in \mathcal{L}(\Phi)$ and ψ is a consistent simplified atomic description over $\mathcal{P}$, then $\Pr_\infty(\varphi|\psi)$ is either 0 or 1.*

If $\mathcal{C} \neq \emptyset$, then we do not get an analogue to Proposition 6.2.10 if we simply partition the worlds according to the atomic description they satisfy. For example, consider the atomic description ψ_* from Example 6.2.6, and the sentence $\varphi = R(c_1, c_1)$ for some binary predicate R. Clearly, by symmetry, $\Pr_\infty(\varphi|\psi_*) = 1/2$, and therefore φ is not uniform over the worlds satisfying ψ_*. We do not even need to use constant symbols, such as c_1, to construct such counterexamples. Recall that the size description in ψ_* included the conjunct $\exists^1 x\, A_1(x)$. So if $\varphi' = \exists x\, (A_1(x) \wedge R(x,x))$ then we also get $\Pr_\infty(\varphi'|\psi_*) = 1/2$.

The general problem is that, given ψ_*, φ can refer "by name" to certain domain elements and thus its truth can depend on their properties. In particular, φ can refer to domain elements that are denotations of constants in $\mathcal{C}$ as well as to domain elements that are the denotations of the "fixed-size" atoms—those atoms whose size is fixed by the atomic description. In the example above, we can view "the x such that $A_1(x)$" as a name for the unique domain element satisfying atom A_1. In any model of ψ_*, we call the denotations of the constants and elements of the fixed-size atoms the *named elements* of that model. The discussion above indicates that there is no uniformity theorem if we condition only on atomic descriptions, because an atomic expression does not fix the denotations of the non-unary predicates with respect to the named elements. This analysis suggests that we should augment an atomic description with complete information about the named elements. This leads to a finer classification of models which does have the uniformity property. To define this classification formally, we need the following definitions.

Definition 6.2.11: The *characteristic* of an atomic description ψ of size M is a tuple C_ψ of the form $\langle (f_1, g_1), \ldots, (f_{2^{|\mathcal{P}|}}, g_{2^{|\mathcal{P}|}}) \rangle$, where

- $f_i = m$ if exactly $m < M$ domain elements satisfy A_i according to ψ,
- $f_i = *$ if at least M domain elements satisfy A_i according to ψ,

- g_i is the number of distinct domain elements which are interpretations of elements in $\mathcal{C}$ that satisfy A_i according to ψ. ▮

Note that we can compute the characteristic of ψ immediately from the syntactic form of ψ.

Definition 6.2.12: Suppose $C_\psi = \langle (f_1, g_1), ..., (f_{2^{|\mathcal{P}|}}, g_{2^{|\mathcal{P}|}}) \rangle$ is the characteristic of ψ. We say that an atom A_i is *active* in ψ if if $f_i = *$; otherwise A_i is *passive*. Let $\mathbf{A}(\psi)$ be the set $\{i \ : \ A_i \text{ is active in } \psi\}$. ▮

We can now define named elements:

Definition 6.2.13: Given an atomic description ψ and a model W of ψ, the *named elements* in W are the elements satisfying the passive atoms and the elements that are denotations of constants.

The number of named elements in any model of ψ is

$$\nu(\psi) = \sum_{i \in \mathbf{A}(\psi)} g_i + \sum_{i \notin \mathbf{A}(\psi)} f_i,$$

where $C_\psi = \langle (f_1, g_1), ..., (f_{2^{|\mathcal{P}|}}, g_{2^{|\mathcal{P}|}}) \rangle$, as before. ▮

As we have discussed, we wish to augment ψ with information about the named elements. We accomplish this using the following notion of *model fragment* which is, roughly speaking, the projection of a model onto the named elements.

Definition 6.2.14: Let $\psi = \sigma \wedge D$ for a size description σ and a complete description D over Ψ. A *model fragment* $\mathcal{V}$ for ψ is a model over the vocabulary Φ with domain $\{1, \ldots, \nu(\psi)\}$ such that:

- $\mathcal{V}$ satisfies D, and

- $\mathcal{V}$ satisfies the conjuncts in σ defining the sizes of the passive atoms. ▮

We can now define what it means for a model W to satisfy a model fragment $\mathcal{V}$.

Definition 6.2.15: Let W be a model of ψ, and let $i_1, \ldots, i_{\nu(\psi)} \in \{1, \ldots, N\}$ be the named elements in W, where $i_1 < i_2 < \ldots < i_{\nu(\psi)}$. The model W is said to *satisfy* the model fragment $\mathcal{V}$ if the function $F(j) = i_j$ from the domain of $\mathcal{V}$ to the domain of W is an isomorphism between $\mathcal{V}$ and the submodel of W formed by restricting to the named elements. ▮

Example 6.2.16: Consider the atomic description ψ_* from Example 6.2.6. Its characteristic C_{ψ_*} is $\langle (1,0), (3,1), (*,1), (*,0) \rangle$. The active atoms are thus A_3 and A_4. Note that $g_3 = 1$ because c_2 and c_3 are constrained to denote the same element. Thus, the number of named elements $\nu(\psi_*)$ in a model of ψ_* is $1+3+1 = 5$. Therefore each model fragment for ψ_* will have domain $\{1,2,3,4,5\}$. The elements in the domain will be the named elements; these correspond to the single element in A_1, the three elements in A_2, and the unique element denoting both c_2 and c_3 in A_3.

Let Φ be $\{P_1, P_2, c_1, c_2, c_3, R\}$ where R is a binary predicate symbol. One possible model fragment $\mathcal{V}_*$ for ψ_* over Φ gives the symbols in Φ the following interpretation:

$$\begin{array}{lll} c_1^{\mathcal{V}_*} = 4 & c_2^{\mathcal{V}_*} = 3 & c_3^{\mathcal{V}_*} = 3 \\ P_1^{\mathcal{V}_*} = \{1,2,4,5\} & P_2^{\mathcal{V}_*} = \{1,3\} & R^{\mathcal{V}_*} = \{(1,3),(3,4)\}. \end{array}$$

It is easy to verify that $\mathcal{V}_*$ satisfies the properties of the constants as prescribed by the description D in ψ_* as well as the two conjuncts $\exists^1 x\, A_1(x)$ and $\exists^3 x\, A_2(x)$ in the size description in ψ_*.

Now, let W be a world satisfying ψ_*, and assume that the named elements in W are $3,8,9,14,17$. Then W satisfies $\mathcal{V}_*$ if this 5-tuple of elements has precisely the same properties in W as the 5-tuple $1,2,3,4,5$ does in $\mathcal{V}_*$. ∎

Although a model fragment is a semantic structure, the definition of satisfaction just given also allows us to regard it as a logical assertion that is true or false in any model over Φ whose domain is a subset of the natural numbers. In the following, we use this view of a model description as an assertion frequently. In particular, we freely use assertions which are the conjunction of an ordinary first-order ψ and a model fragment $\mathcal{V}$, even though the result is not a first-order formula. Under this viewpoint it makes perfect sense to use an expression such as $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$.

Definition 6.2.17: A *model description* augmenting ψ over the vocabulary Φ is a conjunction of ψ and a model fragment $\mathcal{V}$ for ψ over Φ. Let $\mathcal{M}^\Phi(\psi)$ be the set of model descriptions augmenting ψ. (If Φ is clear from context, we omit the subscript and write $\mathcal{M}(\psi)$ rather than $\mathcal{M}^\Phi(\psi)$.) ∎

It should be clear that model descriptions are mutually exclusive and exhaustive. Moreover, as for atomic descriptions, each unary sentence KB is equivalent to some disjunction of model descriptions. From this, and elementary probability theory, we conclude the following fact, which forms the basis of our techniques for computing asymptotic conditional probabilities.

Proposition 6.2.18: *For any $\varphi \in \mathcal{L}(\Phi)$ and $KB \in \mathcal{L}(\Psi)$*

$$\Pr_\infty(\varphi|KB) = \sum_{\psi \in \mathcal{A}^\Psi_{KB}} \sum_{(\psi\wedge\mathcal{V}) \in \mathcal{M}(\psi)} \Pr_\infty(\varphi|\psi \wedge \mathcal{V}) \cdot \Pr_\infty(\psi \wedge \mathcal{V}|KB),$$

if all limits exist.

As we show in the next section, model descriptions have the uniformity property so the first term in the product will always be either 0 or 1.

It might seem that the use of model fragments is a needless complication and that any model fragment, in its role as a logical assertion, will be equivalent to some first-order sentence. Consider the following definition:

Definition 6.2.19: Let $n = \nu(\psi)$. The *complete description capturing* $\mathcal{V}$, denoted $D_{\mathcal{V}}$, is a formula that satisfies the following:[4]

- $D_{\mathcal{V}}$ is a complete description over Φ and the variables $\{x_1, \ldots, x_n\}$ (see Definition 6.2.2),
- for each $i \neq j$, $D_{\mathcal{V}}$ contains a conjunct $x_i \neq x_j$, and
- $\mathcal{V}$ satisfies $D_{\mathcal{V}}$ when i is assigned to x_i for each $i = 1, \ldots, n$. ▮

Example 6.2.20: The complete description $D_{\mathcal{V}_*}$ capturing the model fragment $\mathcal{V}_*$ from the previous example has conjuncts such as $P_1(x_1)$, $\neg P_1(x_3)$, $R(x_1, x_3)$, $\neg R(x_1, x_2)$, and $x_4 = c_1$. ▮

The distinction between a model fragment and the complete description capturing it is subtle. Clearly if a model satisfies $\mathcal{V}$, then it also satisfies $\exists x_1, \ldots, x_n\, D_{\mathcal{V}}$. The converse is not necessarily true. A model fragment places additional constraints on which domain elements are denotations of the constants and passive atoms. For example, a model fragment might entail that, in any model over the domain $\{1, \ldots, N\}$, the denotation of constant c_1 is less than that of c_2. Clearly, no first-order sentence can assert this. The main implication of this difference is combinatorial; it turns out that counting model fragments (rather than the complete descriptions that capture them) simplifies many computations considerably. Although we typically use model fragments, there are occasions where it is important to remain within first-order logic and use the corresponding complete descriptions instead. For instance, this is the case in the next subsection. Whenever we do this we will appeal to the following result, which is easy to prove:

Proposition 6.2.21: *For any $\varphi \in \mathcal{L}(\Phi)$ and model description $\psi \wedge \mathcal{V}$ over Φ, we have*

$$\Pr_\infty(\varphi | \psi \wedge \mathcal{V}) = \Pr_\infty(\varphi | \psi \wedge \exists x_1, \ldots, x_{\nu(\psi)}\, D_{\mathcal{V}}).$$

6.3 Asymptotic conditional probabilities

6.3.1 A conditional 0-1 law

In the previous section, we showed how to partition KB into model descriptions. We now show that φ is uniform over each model description. That is, for any sentence $\varphi \in \mathcal{L}(\Phi)$ and any model description $\psi \wedge \mathcal{V}$, the probability $\Pr_\infty(\varphi | \psi \wedge \mathcal{V})$ is either 0 or 1. The technique we use to prove this is a generalization of Fagin's proof of the 0-1 law for first-order logic without constant or function symbols [Fag76]. This result (independently proved by Glebskiĭ et al. [GKLT69]) states that if φ is a first-order sentence in a vocabulary without constant or function symbols, then $\Pr_\infty(\varphi)$ is either 0 or 1. It is well known that we can get asymptotic probabilities that are neither 0 nor 1 if we use constant symbols, or if we look at general conditional probabilities.

[4] Note that there will, in general, be more than one complete description capturing $\mathcal{V}$. We choose one of them arbitrarily for $D_{\mathcal{V}}$.

However, in the special case where we condition on a model descriptions there is still a 0-1 law. Throughout this section let $\psi \wedge \mathcal{V}$ be a fixed model description with at least one active atom, and let $n = \nu(\psi)$ be the number of named elements according to ψ.

As we said earlier, the proof of our 0-1 law is based on Fagin's proof. Like Fagin, our strategy involves constructing a theory T which, roughly speaking, states that any finite fragment of a model can be extended to a larger fragment in all possible ways. We then prove two propositions:

1. T is complete; that is, for each $\varphi \in \mathcal{L}(\Phi)$, either $T \models \varphi$ or $T \models \neg\varphi$. This result, in the case of the original 0-1 law, is due to Gaifman [Gai64].

2. For any $\varphi \in \mathcal{L}(\Phi)$, if $T \models \varphi$ then $\Pr_\infty(\varphi|\psi \wedge \mathcal{V}) = 1$.

Using the first proposition, for any sentence φ, either $T \models \varphi$ or $T \models \neg\varphi$. Therefore, using the second proposition, either $\Pr_\infty(\varphi|\psi \wedge \mathcal{V}) = 1$ or $\Pr_\infty(\neg\varphi|\psi \wedge \mathcal{V}) = 1$. The latter case immediately implies that $\Pr_\infty(\varphi|\psi \wedge \mathcal{V}) = 0$. Thus, these two propositions suffice to prove the conditional 0-1 law.

We begin by defining several concepts which will be useful in defining the theory T.

Definition 6.3.1: Let $\mathcal{X}' \supseteq \mathcal{X}$, let D be a complete description over Φ and $\mathcal{X}$, and let D' be a complete description over Φ and $\mathcal{X}'$. We say that D' *extends* D if every conjunct of D is a conjunct of D'. ▮

The core of the definition of T is the concept of an *extension axiom*, which asserts that any finite substructure can be extended to a larger structure containing one more element.

Definition 6.3.2: Let $\mathcal{X} = \{x_1, \ldots, x_j\}$ for some k, let D be a complete description over Φ and $\mathcal{X}$, and let D' be any complete description over Φ and $\mathcal{X} \cup \{x_{j+1}\}$ that extends D. The sentence:

$$\forall x_1, x_2, \ldots, x_j\ (D \Rightarrow \exists x_{j+1} D')$$

is an *extension axiom.* ▮

In the original 0-1 law, Fagin considered the theory consisting of all the extension axioms. In our case, we must consider only those extension axioms whose components are consistent with ψ, and which extend $D_{\mathcal{V}}$.

Definition 6.3.3: Given $\psi \wedge \mathcal{V}$, we define T to consist of $\psi \wedge \exists x_1, \ldots, x_n D_{\mathcal{V}}$ together with all extension axioms

$$\forall x_1, x_2, \ldots, x_j\ (D \Rightarrow \exists x_{j+1} D')$$

in which D (and hence D') extends $D_{\mathcal{V}}$ and in which D' (and hence D) is consistent with ψ. ▮

We have used $D_{\mathcal{V}}$ rather than $\mathcal{V}$ in this definition so that T is a first-order theory. Note that the consistency condition above is not redundant, even given that the components of an extension axiom extend $D_{\mathcal{V}}$. However, inconsistency can arise only if D' asserts the existence of a new element in some passive atom (because this would contradict the size description in ψ).

The statements and proofs of the two propositions that imply the 0-1 law can be found in Appendix C.2. As outlined above, these allow us to prove the main theorem of this section:

Theorem 6.3.4: *For any sentence $\varphi \in \mathcal{L}(\Phi)$ and model description $\psi \wedge \mathcal{V}$, $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ is either 0 or 1.*

Note that if ψ is a simplified atomic description, then there are no named elements in any model of ψ. Therefore, the only model description augmenting ψ is simply ψ itself. Thus Proposition 6.2.10, which is Liogon'kiĭ's result, is a corollary of the above theorem.

6.3.2 Computing the relative weights of model descriptions

We now want to compute the relative weights of model descriptions. It will turn out that certain model descriptions are dominated by others, so that their relative weight is 0, while all the dominating model descriptions have equal weight. Thus, the problem of computing the relative weights of model descriptions reduces to identifying the dominating model descriptions. There are two factors that determine which model descriptions dominate. The first, and more significant, is the number of active atoms; the second is the number of named elements. Let $\alpha(\psi)$ denote the number of active atoms according to ψ.

To compute these relative weights of the model descriptions, we must evaluate $\#worlds_N^\Phi(\psi \wedge \mathcal{V})$. The following lemma (whose proof is in Appendix C.3) gives a precise expression for the asymptotic behavior of this function as N grows large.

Lemma 6.3.5: *Let ψ be a consistent atomic description of size $M \geq |\mathcal{C}|$ over Ψ, and let $(\psi \wedge \mathcal{V}) \in \mathcal{M}^\Phi(\psi)$.*

(a) If $\alpha(\psi) = 0$ and $N > \nu(\psi)$, then $\#worlds_N^\Psi(\psi) = 0$. In particular, this holds for all $N > 2^{|\mathcal{P}|}M$.

(b) If $\alpha(\psi) > 0$, then

$$\#worlds_N^\Phi(\psi \wedge \mathcal{V}) \sim \binom{N}{n} a^{N-n} 2^{\sum_{i \geq 2} b_i (N^i - n^i)},$$

where $a = \alpha(\psi)$, $n = \nu(\psi)$, and b_i is the number of predicates of arity i in Φ.

The asymptotic behavior described in this lemma motivates the following definition:

Definition 6.3.6: Given an atomic description ψ over Ψ, let the *degree* of ψ, written $\Delta(\psi)$, be the pair $(\alpha(\psi), \nu(\psi))$, and let degrees be ordered lexicographically. We extend this definition to sentences as follows. For $KB \in \mathcal{L}(\Psi)$, we define the *degree of KB over* Ψ, written $\Delta^{\Psi}(KB)$, to be $\max_{\psi \in \mathcal{A}^{\Psi}_{KB}} \Delta(\psi)$, and the *activity count* of KB, to be $\alpha^{\Psi}(KB)$ (i.e., the first component of $\Delta^{\Psi}(KB)$). ▮

One important conclusion of this lemma justifies our treatment of well-definedness (Definition 6.2.1) when conditioning on unary formulas. It shows that if KB is satisfied in some "sufficiently large" model, then it is satisfiable over all "sufficiently large" domains. It thus allows us to avoid dealing with persistent vs. intermittent limits when conditioning on monadic formulas.

Lemma 6.3.7: *Suppose that* $KB \in \mathcal{L}(\Psi)$, *and* $M = d(KB) + |\mathcal{C}_{KB}|$. *Then the following conditions are equivalent:*

(a) KB is satisfied in some model of cardinality greater than $2^{|\mathcal{P}|}M$,

(b) $\alpha^{\Psi}(KB) > 0$,

(c) for all $N > 2^{|\mathcal{P}|}M$, *KB is satisfiable in some model of cardinality* N,

(d) $\Pr_{\infty}(*|KB)$ *is well-defined.*

For the case of sentences in the languages without equality or constants, the condition for well-definedness simplifies considerably.

Corollary 6.3.8: *If* $KB \in \mathcal{L}^{-}(\mathcal{P})$, *then* $\Pr_{\infty}(*|KB)$ *is well-defined iff KB is satisfiable.*

Proof: The only if direction is obvious. For the other, if KB is consistent, then it is equivalent to a non-empty disjunction of consistent simplified atomic descriptions. Any consistent simplified atomic description has arbitrarily large models. ▮

We remark that we can extend our proof techniques to show that Corollary 6.3.8 holds even if $\mathcal{C} \neq \emptyset$, although we must still require that KB does not mention equality. We omit details here.

For the remainder of this chapter, we will consider only sentences KB such that $\alpha^{\Psi}(KB) > 0$. For the case of unary first-order sentences, this assumption is equivalent to our assumption of Chapter 3 that the knowledge base is always eventually consistent.

Lemma 6.3.5 shows that, asymptotically, the number of worlds satisfying $\psi \wedge \mathcal{V}$ is completely determined by the degree of ψ. Model descriptions of higher degree have many more worlds, and therefore dominate. On the other hand, model descriptions with the same degree have the same number of worlds at the limit, and are therefore equally likely. This observation allows us to compute the relative weights of different model descriptions.

Definition 6.3.9: For any degree $\delta = (a, n)$, let $\mathcal{A}_{KB}^{\Psi,\delta}$ be the set of atomic descriptions $\psi \in \mathcal{A}_{KB}^{\Psi}$ such that $\Delta(\psi) = \delta$. For any set of atomic descriptions $\mathcal{A}'$, we use $\mathcal{M}(\mathcal{A}')$ to denote $\cup_{\psi \in \mathcal{A}'} \mathcal{M}(\psi)$. ▮

Theorem 6.3.10: *Let $KB \in \mathcal{L}(\Psi)$ and $\Delta^{\Psi}(KB) = \delta$. Let ψ be an atomic description in $\mathcal{A}_{KB}^{\Psi}$, and let $\psi \wedge \mathcal{V} \in \mathcal{M}^{\Phi}(\psi)$.*

(a) If $\Delta(\psi) < \delta$ then $\Pr_\infty(\psi \wedge \mathcal{V} | KB) = 0$.

(b) If $\Delta(\psi) = \delta$ then $\Pr_\infty(\psi \wedge \mathcal{V} | KB) = 1/|\mathcal{M}^{\Phi}(\mathcal{A}_{KB}^{\Psi,\delta})|$.

Combining this result with Proposition 6.2.18, we deduce

Theorem 6.3.11: *For any $\varphi \in \mathcal{L}(\Phi)$ and $KB \in \mathcal{L}(\Psi)$,*

$$\Pr_\infty(\varphi | KB) = \sum_{(\psi \wedge \mathcal{V}) \in \mathcal{M}(\mathcal{A}_{KB}^{\Psi,\delta})} \Pr_\infty(\varphi | \psi \wedge \mathcal{V}) / |\mathcal{M}(\mathcal{A}_{KB}^{\Psi,\delta})|.$$

This result, together with the techniques of the next section, will allow us compute asymptotic conditional probabilities.

The results of Liogon'kiĭ are a simple corollary of the above theorem. For an activity count a, let $\mathcal{A}_{KB}^{\Psi,a}$ denote the set of atomic descriptions $\psi \in \mathcal{A}_{KB}^{\Psi}$ such that $\alpha(\psi) = a$.

Theorem 6.3.12: [Lio69] *Assume that $\mathcal{C} = \emptyset$, $\varphi \in \mathcal{L}(\Phi)$, $KB \in \mathcal{L}^{-}(\mathcal{P})$, and $\alpha^{\mathcal{P}}(KB) = a$. Then* $\Pr_\infty(\varphi | KB) = \sum_{\psi \in \mathcal{A}_{KB}^{\mathcal{P},a}} \Pr_\infty(\varphi | \psi) / |\mathcal{A}_{KB}^{\mathcal{P},a}|$.

Proof: By Lemma 6.2.9, a sentence $KB \in \mathcal{L}^{-}(\mathcal{P})$ is the disjunction of the simplified atomic descriptions in $\mathcal{A}_{KB}^{\mathcal{P}}$. A simplified atomic description ψ has no named elements, and therefore $\Delta(\psi) = (\alpha(\psi), 0)$. Moreover, $\mathcal{M}(\psi) = \{\psi\}$ for any $\psi \in \mathcal{A}_{KB}^{\mathcal{P}}$. The result now follows trivially from the previous theorem. ▮

This calculation simplifies somewhat if φ and KB are both monadic. In this case, we assume without loss of generality that $d(\varphi) = d(KB)$. (If not, we can replace φ with $\varphi \wedge KB$ and KB with $KB \wedge (\varphi \vee \neg\varphi)$.) This allows us to assume that $\mathcal{A}_{\varphi \wedge KB}^{\Psi} \subseteq \mathcal{A}_{KB}^{\Psi}$, thus simplifying the presentation.

Corollary 6.3.13: *Assume that $\varphi, KB \in \mathcal{L}^{-}(\mathcal{P})$, and $\alpha^{\mathcal{P}}(KB) = a$. Then*

$$\Pr_\infty(\varphi | KB) = \frac{|\mathcal{A}_{\varphi \wedge KB}^{\mathcal{P},a}|}{|\mathcal{A}_{KB}^{\mathcal{P},a}|}.$$

Proof: Since φ is monadic, $\varphi \wedge KB$ is equivalent to a disjunction of the atomic descriptions $\mathcal{A}_{\varphi \wedge KB}^{\mathcal{P}} \subseteq \mathcal{A}_{KB}^{\mathcal{P}}$. Atomic descriptions are mutually exclusive; thus, for $\psi \in \mathcal{A}_{KB}^{\Psi}$, $\Pr_\infty(\varphi | \psi) = 1$ if $\psi \in \mathcal{A}_{\varphi \wedge KB}^{\Psi}$ and $\Pr_\infty(\varphi | \psi) = 0$ otherwise. The result then follows immediately from Theorem 6.3.12. ▮

6.4 Complexity analysis

In this section we investigate the computational complexity of problems associated with asymptotic conditional probabilities. In fact, we consider three problems: deciding whether the asymptotic probability is well-defined, computing it, and approximating it.

Our computational approach is based on Theorem 6.3.11, which tells us that

$$\Pr_\infty(\varphi|KB) = \frac{1}{|\mathcal{M}(\mathcal{A}^{\Psi,\delta}_{KB})|} \cdot \sum_{(\psi\wedge\mathcal{V})\in\mathcal{M}(\mathcal{A}^{\Psi,\delta}_{KB})} \Pr_\infty(\varphi|\psi\wedge\mathcal{V}).$$

The basic structure of the algorithms we give for computing $\Pr_\infty(\varphi|KB)$ is simply to enumerate model descriptions $\psi \wedge \mathcal{V}$ and, for those of the maximum degree, compute the conditional probability $\Pr_\infty(\varphi|\psi\wedge\mathcal{V})$. In Section 6.4.1 we show how to compute this latter probability.

6.4.1 Computing the 0-1 probabilities

The method we give for computing $\Pr_\infty(\varphi|\psi\wedge\mathcal{V})$ is an extension of Grandjean's algorithm [Gra83] for computing asymptotic probabilities in the unconditional case. For the purposes of this section, fix a model description $\psi\wedge\mathcal{V}$ over Φ. In our proof of the conditional 0-1 law (Section 6.3.1), we defined a theory T corresponding to $\psi\wedge\mathcal{V}$. We showed that T is a complete and consistent theory, and that $\varphi \in \mathcal{L}(\Phi)$ has asymptotic probability 1 iff $T \models \varphi$. We therefore need an algorithm that decides whether $T \models \varphi$.

Grandjean's original algorithm decides whether $\Pr_\infty(\varphi)$ is 0 or 1 for a sentence φ with no constant symbols. For this case, the theory T consists of all possible extension axioms, rather than just the ones involving model descriptions extending $D_\mathcal{V}$ and consistent with ψ (see Definition 6.3.3). The algorithm has a recursive structure, which at each stage attempts to decide something more general than whether $T \models \varphi$. It decides whether $T \models D \Rightarrow \xi$, where

- D is a complete description over Φ and the set $\mathcal{X}_j = \{x_1,\ldots,x_j\}$ of variables,
- $\xi \in \mathcal{L}(\Phi)$ is any formula whose only free variables (if any) are in $\mathcal{X}_j$.

The algorithm begins with $j = 0$. In this case, D is a complete description over $\mathcal{X}_0$ and Φ. Since Φ contains no constants and $\mathcal{X}_0$ is the empty set, D must in fact be the empty conjunction, which is equivalent to the formula *true*. Thus, for $j = 0$, $T \models D \Rightarrow \varphi$ iff $T \models \varphi$. While $j = 0$ is the case of real interest, the recursive construction Grandjean uses forces us to deal with the case $j > 0$ as well. In this case, the formula $D \Rightarrow \varphi$ contains free variables; these variables are treated as being universally quantified for purposes of determining if $T \models D \Rightarrow \varphi$.

Our algorithm is the natural extension to Grandjean's algorithm for the case of conditional probabilities and for a language with constants. The chief difference is that we begin by considering $T \models D_\mathcal{V} \Rightarrow \varphi$ (where $\mathcal{V}$ is the model fragment on which we are conditioning). Suppose $D_\mathcal{V}$ uses the variables $x_1,\ldots,x_n$, where $n = \nu(\psi)$. We have said that $T \models D_\mathcal{V} \Rightarrow \varphi$ is interpreted as $T \models \forall x_1,\ldots,x_n\ (D_\mathcal{V} \Rightarrow \varphi)$, and this is equivalent to $T \models (\exists x_1,\ldots,x_n\ D_\mathcal{V}) \Rightarrow \varphi$ because

Procedure `Compute01`($D \Rightarrow \xi$)

1. If ξ is of the form ξ' or $\neg\xi'$ for an atomic formula ξ' then:
    - Return(*true*) if ξ is a conjunct of D,
    - Return(*false*) otherwise.
2. If ξ is of the form $\xi_1 \wedge \xi_2$ then:
    - Return(*true*) if `Compute01`($D \Rightarrow \xi_1$) and `Compute01`($D \Rightarrow \xi_2$),
    - Return(*false*) otherwise.
3. If ξ is of the form $\xi_1 \vee \xi_2$ then:
    - Return(*true*) if `Compute01`($D \Rightarrow \xi_1$) or `Compute01`($D \Rightarrow \xi_2$),
    - Return(*false*) otherwise.
4. If ξ is of the form $\exists y\, \xi'$ and D is a complete description over Φ and $\{x_1, \ldots, x_j\}$ then:
    - Return(*true*) if `Compute01`($D' \Rightarrow \xi'[y/x_{j+1}]$) for some complete description D' over Φ and $\{x_1, \ldots, x_{j+1}\}$ that extends D and is consistent with ψ.
    - Return(*false*) otherwise.
5. If ξ is of the form $\forall y\, \xi'$ and D is a complete description over Φ and $\{x_1, \ldots, x_j\}$ then:
    - Return(*true*) if `Compute01`($D' \Rightarrow \xi'[y/x_{j+1}]$) for all complete descriptions D' over Φ and $\{x_1, \ldots, x_{j+1}\}$ that extend D and are consistent with ψ.
    - Return(*false*) otherwise.

Figure 6.2: `Compute01` for computing 0-1 probabilities

φ is closed. Because $\exists x_1, \ldots, x_n\ D_{\mathcal{V}}$ is in T by definition, this latter assertion is equivalent to $T \models \varphi$, which is what we are really interested in.

Starting from the initial step just outlined, the algorithm then recursively examines smaller and smaller subformulas of φ, while maintaining a description D which keeps track of any new free variables that appear in the current subformula. Of course, D will also extend $D_{\mathcal{V}}$ and will be consistent with ψ.

The recursive procedure `Compute01` in Figure 6.2 implements this idea. For a complete description D over Φ and $\mathcal{X}_j$ (where D extends $D_{\mathcal{V}}$ and is consistent with ψ), and a formula $\xi \in \mathcal{L}(\Phi)$ whose free variables (if any) are in $\mathcal{X}_j$, it decides whether $T \models D \Rightarrow \xi$. The algorithm proceeds by induction on the structure of the formula, until the base case—an atomic formula or its negation—is reached. `Compute01` is called initially with the arguments $D_{\mathcal{V}}$ and φ. Without loss of generality, we assume that all negations in φ are pushed in as far as possible, so that only atomic formulas are negated. We also assume that φ does not use the variables $x_1, x_2, x_3, \ldots$.

The proof that `Compute01` is correct can be found in Appendix C.4. The appendix also describes how the algorithm can be implemented on an *alternating Turing machine* (ATM) [CKS81]. In an ATM, the nonterminal states are classified into two kinds: universal and existential. Just as with a nondeterministic TM, a nonterminal state may have one or more successors. The terminal states are classified into two kinds: accepting and rejecting. The computation of an ATM forms a tree, where the nodes are instantaneous descriptions (ID's) of the machine's state at various points in the computation, and the children of a node are the possible successor ID's. We recursively define what it means for a node in a computation tree to be an *accepting* node. Leaves are terminal states, and a leaf is accepting just if the machine is in an accepting state in the corresponding ID. A node whose ID is in an existential state is accepting iff at least one of its children is accepting. A node whose ID is in a universal state is accepting iff all of its children are accepting. The entire computation is accepting if the root is an accepting node.

We use several different measures for the complexity of an ATM computation. The time of the computation is the number of steps taken by its longest computation branch. The number of *alternations* of a computation of an ATM is the maximum number of times, over all branches, that the type of state switched (from universal to existential or vice versa). The number of *branches* is simply the number of distinct computation paths. The number of branches is always bounded by an exponential in the computation time, but sometimes we can find tighter bounds.

`Compute01` is easily implemented on an ATM (as is Grandjean's original algorithm). The complexity analysis of the resulting algorithm is summarized in the following theorem, which forms the basis for almost all of our upper bounds in this section.

Theorem 6.4.1: *There exists an alternating Turing machine that takes as input a finite vocabulary Φ, a model description $\psi \wedge \mathcal{V}$ over Φ, and a formula $\varphi \in \mathcal{L}(\Phi)$, and decides whether $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ is 0 or 1. The machine uses time $O(|\Phi|2^{|\mathcal{P}|}(\nu(\psi) + |\varphi|)^\rho)$ and $O(|\varphi|)$ alternations, where ρ is the maximum arity of predicates in Φ. If $\rho > 1$, the number of branches is $2^{O(|\Phi|(\nu(\psi)+|\varphi|)^\rho)}$. If $\rho = 1$, the number of branches is $O((2^{|\Phi|} + \nu(\psi))^{|\varphi|})$.*

An alternating Turing machine can also be simulated by a deterministic Turing machine. This allows us to prove the following important corollary.

Corollary 6.4.2: *There exists a deterministic Turing machine that takes as input a finite vocabulary Φ, a model description $\psi \wedge \mathcal{V}$ over Φ, and a formula $\varphi \in \mathcal{L}(\Phi)$, and decides whether $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ is 0 or 1. If $\rho > 1$ the machine uses time $2^{O(|\Phi|(\nu(\psi)+|\varphi|)^\rho)}$ and space $O(|\Phi|(\nu(\psi) + |\varphi|)^\rho)$. If $\rho = 1$ the machine uses time $2^{O(|\varphi||\Phi|\log(\nu(\psi)+1))}$ and space $O(|\varphi||\Phi|\log(\nu(\psi) + 1))$.*

6.4.2 Computing asymptotic conditional probabilities

Our overall goal is to compute $\Pr_\infty(\varphi|KB)$ for some $\varphi \in \mathcal{L}(\Phi)$ and $KB \in \mathcal{L}(\Psi)$. To do this, we enumerate model descriptions over Φ of size $d(KB) + |\mathcal{C}|$, and check which are consistent with

```
Procedure $\texttt{Compute-Pr}_\infty(\varphi|KB)$

  $\delta \leftarrow (0,0)$
  For each model description $\psi \wedge \mathcal{V}$ do:
    Compute $\Pr_\infty(KB|\psi \wedge \mathcal{V})$ using $\texttt{Compute01}(D_{\mathcal{V}} \Rightarrow KB)$
    If $\Delta(\psi) = \delta$ and $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$ then
      $\mathit{count}(KB) \leftarrow \mathit{count}(KB) + 1$
      Compute $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ using $\texttt{Compute01}(D_{\mathcal{V}} \Rightarrow \varphi)$
      $\mathit{count}(\varphi) \leftarrow \mathit{count}(\varphi) + \Pr_\infty(\varphi|\psi \wedge \mathcal{V})$
    If $\Delta(\psi) > \delta$ and $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$ then
      $\delta \leftarrow \Delta(\psi)$
      $\mathit{count}(KB) \leftarrow 1$
      Compute $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ using $\texttt{Compute01}(D_{\mathcal{V}} \Rightarrow \varphi)$
      $\mathit{count}(\varphi) \leftarrow \Pr_\infty(\varphi|\psi \wedge \mathcal{V})$
  If $\delta = (0,0)$ then output "$\Pr_\infty(\varphi|KB)$ not well-defined"
    otherwise output "$\Pr_\infty(\varphi|KB) = \mathit{count}(\varphi)/\mathit{count}(KB)$".
```

Figure 6.3: `Compute-Pr`$_\infty$ for computing asymptotic conditional probabilities.

KB. Among those model descriptions that are of maximal degree, we compute the fraction of model descriptions for which $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ is 1.

More precisely, let $\delta_{KB} = \Delta^{\Psi}(KB)$. Theorem 6.3.11 tells us that

$$\Pr_\infty(\varphi|KB) = \frac{1}{|\mathcal{M}(\mathcal{A}_{KB}^{\Psi,\delta_{KB}})|} \sum_{(\psi \wedge \mathcal{V}) \in \mathcal{M}(\mathcal{A}_{KB}^{\Psi,\delta_{KB}})} \Pr_\infty(\varphi|\psi \wedge \mathcal{V}).$$

The procedure `Compute-Pr`$_\infty$, described in Figure 6.3, generates one by one all model descriptions $\psi \wedge \mathcal{V}$ of size $d(KB) + |\mathcal{C}|$ over Φ. The algorithm keeps track of three things, among the model descriptions considered thus far: (1) the highest degree δ of a model description consistent with KB, (2) the number $\mathit{count}(KB)$ of model descriptions of degree δ consistent with KB, and (3) among the model descriptions of degree δ consistent with KB, the number $\mathit{count}(\varphi)$ of descriptions such that $\Pr_\infty(\varphi|\psi \wedge \mathcal{V}) = 1$. Thus, for each model description $\psi \wedge \mathcal{V}$ generated, the algorithm computes $\Delta(\psi)$. If $\Delta(\psi) < \delta$ or $\Pr_\infty(KB|\psi \wedge \mathcal{V})$ is 0, then the model description is ignored. Otherwise, if $\Delta(\psi) > \delta$, then the count for lower degrees is irrelevant. In this case, the algorithm erases the previous counts by setting $\delta \leftarrow \Delta(\psi)$, $\mathit{count}(KB) \leftarrow 1$, and $\mathit{count}(\varphi) \leftarrow \Pr_\infty(\varphi|\psi \wedge \mathcal{V})$. If $\Delta(\psi) = \delta$, then the algorithm updates $\mathit{count}(KB)$ and $\mathit{count}(\varphi)$ appropriately.

Different variants of this algorithm are the basis for most of the upper bounds in the remainder of this chapter.

6.4.3 Finite vocabulary

We now consider the complexity of various problems related to $\Pr_\infty(\varphi|KB)$ for a fixed finite vocabulary Φ. The input for such problems is φ and KB, and so the input length is the sum of the lengths of φ and KB. Since, for the purposes of this section, we view the vocabulary Φ as fixed (independent of the input), its size and maximum arity can be treated as constants.

We first consider the issue of well-definedness.

Theorem 6.4.3: *Fix a finite vocabulary Φ with at least one unary predicate symbol. For $KB \in \mathcal{L}(\Psi)$, the problem of deciding whether $\Pr_\infty(*|KB)$ is well-defined is PSPACE-complete. The lower bound holds even if $KB \in \mathcal{L}^-(\{P\})$.*

In order to compute asymptotic conditional probabilities in this case, we simply use the function $\texttt{Compute-Pr}_\infty$. In fact, since $\texttt{Compute-Pr}_\infty$ can also be used to determine well-definedness, we could also have used it to prove the previous theorem.

Theorem 6.4.4: *Fix a finite vocabulary Φ. For $\varphi \in \mathcal{L}(\Phi)$ and $KB \in \mathcal{L}(\Psi)$, the problem of computing $\Pr_\infty(\varphi|KB)$ is PSPACE-complete. Indeed, deciding if $\Pr_\infty(\varphi|true) = 1$ is PSPACE-hard even if $\varphi \in \mathcal{L}^-(\{P\})$ for some unary predicate symbol P.*

Since for $\varphi \in \mathcal{L}^-(\{P\})$, the probability $\Pr_\infty(\varphi|true)$ is either 0 or 1, it follows immediately from Theorem 6.4.4 that we cannot approximate the limit. Indeed, if we fix ϵ with $0 < \epsilon < 1$, the problem of deciding whether $\Pr_\infty(\varphi|KB) \in [0, 1 \Leftrightarrow \epsilon]$ is PSPACE-hard even for $\varphi, KB \in \mathcal{L}^-(\{P\})$. We might hope to prove that for any nontrivial interval $[r_1, r_2]$, it is PSPACE-hard to decide if $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$. This stronger lower bound does not hold for the language $\mathcal{L}^-(\{P\})$. Indeed, it follows from Theorem 6.3.12 that if Φ is any fixed vocabulary then, for $\varphi \in \mathcal{L}(\Phi)$ and $KB \in \mathcal{L}^-(\Psi)$, $\Pr_\infty(\varphi|KB)$ must take one of a finite number of values (the possible values being determined entirely by Φ). So the approximation problem is frequently trivial; in particular, this is the case for any $[r_1, r_2]$ that does not contain one of the possible values. To see that there are only a finite number of values, first note that there is a fixed collection of atoms over Φ. If KB does not use equality, an atomic description can only say, for each atom A over Φ, whether $\exists x\, A(x)$ or $\neg\exists x\, A(x)$ holds. There is also a fixed set of constant symbols to describe. Therefore, there is a fixed set of possible atomic descriptions. Finally, note that the only named elements are the constants, and so there is also a fixed (and finite) set of model fragments. This shows that the set of model descriptions is finite, from which it follows that $\Pr_\infty(\varphi|KB)$ takes one of finitely many values fixed by Φ. Thus, in order to have have $\Pr_\infty(\varphi|KB)$ assume infinitely many values, we must allow equality in the language. Moreover, even with equality in the language, one unary predicate does not suffice. Using Theorem 6.3.11, it can be shown that two unary predicates are necessary to allow the asymptotic conditional probability to assume infinitely many possible values. As the following result shows, this condition also suffices.

Theorem 6.4.5: *Fix a finite vocabulary Φ that contains at least two unary predicates and rational numbers $0 \le r_1 \le r_2 \le 1$ such that $[r_1, r_2] \neq [0, 1]$. For $\varphi, KB \in \mathcal{L}(\mathcal{P})$, the problem* *of deciding whether* $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ *is PSPACE-hard, even given an oracle that tells us whether the limit is well-defined.*

These results show that simply assuming that the vocabulary is fixed and finite is not by itself enough to lead to computationally easy problems. Nevertheless, there is some good news. We observed in the previous chapter that if Φ is fixed and finite, and we bound the depth of quantifier nesting, then there exists a linear time algorithm for computing asymptotic probabilities. In general, as we observed, we cannot effectively construct this algorithm, although we know that it exists. As we now show, for the case of conditioning on a unary formula, we can effectively construct this algorithm.

Theorem 6.4.6: *Fix* $d \geq 0$. *For* $\varphi \in \mathcal{L}(\Phi)$, $KB \in \mathcal{L}(\Psi)$ *such that* $d(\varphi), d(KB) \leq d$, *we can effectively construct a linear time algorithm that decides if* $\Pr_\infty(\varphi|KB)$ *is well-defined and computes it if it is.*

6.4.4 Infinite vocabulary—restricted cases

Up to now we have assumed that the vocabulary Φ is finite. In many standard complexity arguments it is important that the vocabulary be infinite. For example, satisfiability for propositional logic formulas is decidable in linear time if we consider a fixed finite vocabulary; we need to consider the class of formulas definable over some infinite vocabulary of propositional symbols to get NP-completeness. In the next three sections we consider formulas over an infinite vocabulary Ω. As we observed in Section 3.2, the probability $\Pr_N(\varphi|KB)$ is independent of our choice of vocabulary. Therefore, when φ and KB are drawn from an infinite vocabulary, it is simplest to define $\Pr_N(\varphi|KB)$ relative to the set of vocabulary symbols actually appearing in φ and in KB. Thus, the assumption of an infinite vocabulary only affects the complexity analysis.

As before, we are interested in computing the complexity of the same three problems: deciding whether the asymptotic probability is well-defined, computing it, and approximating it. As we mentioned earlier, the complexity is quite sensitive to a number of factors. One factor, already observed in the unconditional case [BGK85, Gra83], is whether there is a bound on the arity of the predicates in Ω. Without such a bound, the problem is complete for the class #TA(EXP,LIN). Unlike the unconditional case, however, simply putting a bound on the arity of the predicates in Ω is not enough to improve the complexity (unless the bound is 1); we also need to restrict the use of equality, so that it cannot appear in the right-hand side of the conditional. Roughly speaking, with equality, we can use the named elements to play the same role as the predicates of unbounded arity. In this section, we consider what happens if we in fact restrict the language so that either (1) Ω has no predicate of arity ≥ 2, or (2) there is a bound (which may be greater than 1) on the arity of the predicates in Ω, but we never condition on formulas that use equality. As we now show, these two cases turn out to be quite similar. In particular, the same complexity results hold.

Throughout this section, we take Ω to be a fixed infinite vocabulary such that all predicate symbols in Ω have arity less than some fixed bound ρ. Let $\mathcal{Q}$ be the set of all unary predicate symbols in Ω, let $\mathcal{D}$ be the set of all constant symbols in Ω, and let $\Upsilon = \mathcal{Q} \cup \mathcal{D}$.

We start with the problem of deciding whether the asymptotic probability is well defined. Since well-definedness depends only on the right-hand side of the conditional, which we already assume is restricted to mentioning only unary predicates, its complexity is independent of the bound ρ. Thus, the well-definedness problem is NEXPTIME-complete even if we do not use the assumptions that we are making throughout the rest of this section.

Theorem 6.4.7: *For $KB \in \mathcal{L}(\Upsilon)$, the problem of deciding if $\Pr_\infty(*|KB)$ is well-defined is NEXPTIME-complete. The NEXPTIME lower bound holds even for $KB \in \mathcal{L}^-(\mathcal{Q})$ where $d(KB) \leq 2$.*

We next consider the problem of computing the asymptotic probability $\Pr_\infty(\varphi|KB)$, given that it is well-defined. We show that this problem is #EXP-complete. Recall that #P (see [Val79a]) is the class of integer functions computable as the number of accepting computations of a nondeterministic polynomial-time Turing machine. More precisely, a function $f : \{0,1\}^* \to I\!N$ is said to be in #P if there is a nondeterministic Turing machine $\mathbf{M}$ such that for any w, the number of accepting paths of $\mathbf{M}$ on input w is $f(w)$. The class #EXP is the exponential time analogue.

The function we are interested in is $\Pr_\infty(\varphi|KB)$, which is not integer valued. Nevertheless, we want to show that it is in #EXP. In the spirit of similar definitions for #P (see, for example, [Val79b, PB83]) and NP (e.g., [GJ79]) we extend the definition of #EXP to apply also to non-integer-valued functions.

Definition 6.4.8: An arbitrary function f is said to be *#EXP-easy* if there exists an integer-valued function g in #EXP and a polynomial-time-computable function h such that for all x, $f(x) = h(g(x))$. (In particular, we allow h to involve divisions, so that $f(x)$ may be a rational function.) A function f is *#EXP-hard* if, for every #EXP-easy function g, there exist polynomial-time functions h_1 and h_2 such that, for all x, $g(x) = h_2(f(h_1(x)))$.[5] A function f is *#EXP-complete* if it is #EXP-easy and #EXP-hard. ∎

We can similarly define analogues of these definitions for the class #P.

We now show that for an infinite arity-bounded vocabulary in which equality is not used, or for any unary vocabulary, the problem of computing the asymptotic conditional probability is #EXP-complete. We start with the upper bound.

Theorem 6.4.9: *If either (a) $\varphi, KB \in \mathcal{L}(\Upsilon)$ or (b) $\varphi \in \mathcal{L}(\Omega)$ and $KB \in \mathcal{L}^-(\Upsilon)$, then computing $\Pr_\infty(\varphi|KB)$ is #EXP-easy.*

[5] Notice that we need the function h_2 as well as h_1. For example, if g is an integer-valued function and f always returns a rational value between 0 and 1, as is the case for us, then there is no function h_1 such that $g(x) = f(h_1(x))$.

We now want to prove a matching lower bound. Just as for Theorem 6.4.7, we show that the lower bound actually holds for $\varphi, KB \in \mathcal{L}^-(\mathcal{Q})$ of quantifier depth 2.

Theorem 6.4.10: *Given $\varphi, KB \in \mathcal{L}^-(\mathcal{Q})$ of depth at least 2, the problem of computing $\Pr_\infty(\varphi|KB)$ is #EXP-hard, even given an oracle for deciding whether the limit exists.*

As in Theorem 6.4.5, we can also show that any nontrivial approximation of the asymptotic probability is hard, even if we restrict to sentences of depth 2.

Theorem 6.4.11: *Fix rational numbers $0 \leq r_1 \leq r_2 \leq 1$ such that $[r_1, r_2] \neq [0,1]$. For $\varphi, KB \in \mathcal{L}^-(\mathcal{Q})$ of depth at least 2, the problem of deciding whether $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ is both NEXPTIME-hard and co-NEXPTIME-hard, even given an oracle for deciding whether the limit exists.*

6.4.5 Sentences of depth 1

The lower bounds of the previous section all hold provided we consider formulas whose quantification depth is at least 2. Can we do better if we restrict to formulas of quantification depth at most 1? As we show in this section, we can. The complexities typically drop by an exponential factor. For example, checking well-definedness is now NP-complete rather than NEXPTIME-complete. We can also prove #P-hardness for the problem of computing probabilities for depth 1 sentences, and can give a matching upper bound for a subclass of such sentences. For the full generality of depth 1 sentences, we have not proved a #P upper bound for computing the asymptotic probability. The best algorithm we have found for general depth one sentences is in PSPACE. We observe that the depth 1 case is strongly related to the case of computing probabilities over a propositional language. In fact, our lower bounds are proved for an essentially propositional language. A related paper by Roth [Rot93] extends our hardness results to restricted propositional languages.

We begin with the lower bounds. In fact, all of our lower bounds rely only on quantifier-free sentences, over a vocabulary consisting of unary predicates and a single constant c.

Theorem 6.4.12: *For a quantifier-free sentence $KB \in \mathcal{L}^-(\mathcal{Q} \cup \{c\})$, the problem of deciding whether $\Pr_\infty(*|KB)$ is well-defined is NP-hard.*

Theorem 6.4.13: *For quantifier-free sentences $\varphi, KB \in \mathcal{L}^-(\mathcal{Q} \cup \{c\})$, the problem of computing $\Pr_\infty(\varphi|KB)$ is #P-hard.*

The next result shows that it is difficult even to approximate conditional probabilities in $\mathcal{L}^-(\mathcal{Q} \cup \{c\})$.

Theorem 6.4.14: *Fix rational numbers $0 \leq r_1 \leq r_2 \leq 1$ such that $[r_1, r_2] \neq [0,1]$. For quantifier-free sentences $\varphi, KB \in \mathcal{L}^-(\mathcal{Q} \cup \{c\})$, deciding whether $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ is both NP-hard and co-NP-hard.*

It follows from the proof of this theorem that, for any $\epsilon > 0$, it is both NP-hard and co-NP-hard to find a value v that approximates the asymptotic probability to within less than $1/2 \Leftrightarrow \epsilon$. It is straightforward to show that this subsumes a similar result of Paris and Vencovska [PV89], where it is proven that approximating asymptotic probabilities for a richer language (which includes statistical information) is NP-hard.

We now state the upper bound corresponding to Theorem 6.4.12.

Theorem 6.4.15: *For $KB \in \mathcal{L}(\Upsilon)$ of quantifier depth 1, the problem of deciding whether $\Pr_\infty(*|KB)$ is well-defined is in NP.*

We have not been able to prove a matching upper bound for Theorem 6.4.13; all we can prove is a PSPACE upper bound. We can, however, prove a #P upper bound under certain restrictions (see Theorem C.5.9 in Appendix C.5.3). To prove these results, we would like to use the same techniques used in Theorem 6.4.4. That is, we would like to generate model descriptions and for each of these compute the probability of KB and $\varphi \wedge KB$ given the model description. However, we cannot accomplish this in polynomial space, since model descriptions can have exponential size. This is not due to the number of named elements because, as we show later, the only named elements in an atomic description of maximal degree (that is consistent with a depth 1 formula) are the constants. However, an atomic description must still list the (potentially exponentially many) finite atoms, and a model fragment must list the properties of the constants which can also require exponential space. For the latter, observe that if Ω_φ contains a predicate R of arity r, describing the denotation of R over the constants could take as much as $|\mathcal{C}|^r$ space, where $\mathcal{C} = \mathcal{D}_{\varphi \wedge KB}$. Since r can be as large as $O(|\varphi|)$, this is exponential in the size of the input. It follows that we need some shorter alternative to the use of complete model descriptions. Fortunately it turns out that, in the case of formulas of depth 1, we can find a polynomial-length substitute. The appropriate definitions, and the proof of the following theorem, can be found in Appendix C.5.3.

Theorem 6.4.16: *For sentences $KB \in \mathcal{L}(\Upsilon)$ and $\varphi \in \mathcal{L}(\Omega)$ of quantifier depth 1, the problem of computing $\Pr_\infty(\varphi|KB)$ is in PSPACE.*

6.4.6 Infinite vocabulary—the general case

In Section 6.4.4 we investigated the complexity of asymptotic conditional probabilities when the (infinite) vocabulary satisfies certain restrictions. As we now show, the results there were tight in the sense that the restrictions cannot be weakened. We examine the complexity of the general case, in which the vocabulary is infinite with no bound on predicates' arities and/or in which equality can be used.

The problem of checking if $\Pr_\infty(\varphi|KB)$ is well defined is still NEXPTIME-complete. Theorem 6.4.7 (which had no restrictions) still applies. However, the complexity of the other problems we consider does increase. It can be best described in terms of the complexity class TA(EXP,LIN)—the class of problems that can be solved by an exponential time ATM with a linear number of alternations. The class TA(EXP,LIN) also arises in the study of unconditional probabilities where there is no bound on the arity of the predicates. Grandjean [Gra83] proved a TA(EXP,LIN) upper bound for computing whether the unconditional probability is 0 or 1 in this case, and Immerman [BGK85] proved a matching lower bound. Of course, Grandjean's result can be viewed as a corollary of Theorem 6.4.1. Immerman's result, which has not, to the best of our knowledge, appeared in print, is a corollary of Theorem C.5.10 which we state and prove in Appendix C.5.4.

To capture the complexity of computing the asymptotic probability in the general case, we use a counting class #TA(EXP,LIN) that corresponds to TA(EXP,LIN). To define this class, we restrict attention to the class of ATM's whose initial states are existential. Given such an ATM **M**, we define an *initial existential path* in the computation tree of **M** on input w to be a path in this tree, starting at the initial state, such that every node on the path corresponds to an existential state except for the last node, which corresponds to a universal or an accepting state. That is, an initial existential path is a maximal path that starts at the root of the tree and contains only existential nodes except for the last node in the path. We say that an integer-valued function $f : \{0,1\}^* \to I\!N$ is in #TA(EXP,LIN) if there is a machine **M** in the class TA(EXP,LIN) such that, for all w, $f(w)$ is the number of existential paths in the computation tree of **M** on input w whose last node is accepting (recall that we defined a notion of "accepting" for any node in the tree in Section 6.4.1). We extend the definition of #TA(EXP,LIN) to apply to non-integer valued functions and define *#TA(EXP,LIN)-easy* just as we did before with #P and #EXP in Section 6.4.4.

We start with the upper bound.

Theorem 6.4.17 : *For $\varphi \in \mathcal{L}(\Omega)$ and $KB \in \mathcal{L}(\Upsilon)$, the function* $\Pr_\infty(\varphi|KB)$ *is in #TA(EXP,LIN).*

We now want to state the matching lower bound. Moreover, we would like to show that the restrictions from Section 6.4.4 cannot be weakened. Recall from Theorem 6.4.9 that the #EXP upper bound held under one of two conditions: either (a) φ and KB are both unary, or (b) the vocabulary is arity-bounded and KB does not use equality. To show that (a) is tight, we show that the #TA(EXP,LIN) lower bound holds even if we allow φ and KB to use only binary predicates and equality. (The use of equality is necessary, since without it we know from (b) that the problem is #EXP-easy.) To show that (b) is tight, we show that the lower bound holds for a non-arity-bounded vocabulary, but without allowing equality in KB. Neither lower bound requires the use of constants.

Theorem 6.4.18: *For $\varphi \in \mathcal{L}(\Omega)$ and $KB \in \mathcal{L}(\Upsilon)$, computing* $\Pr_\infty(\varphi|KB)$ *is #TA(EXP,LIN)-hard. The lower bound holds even if φ, KB do not mention constant symbols and either (a) φ uses no predicate of arity > 2, or (b) KB uses no equality.*

6.5 Discussion

In this chapter, we have focused on the case of a first-order language, where the formula we are conditioning on is unary. We have presented an algorithm for computing asymptotic conditional probabilities in this case, and investigated the complexity of the problem under a number of assumptions.

Clearly, we are ultimately interested in the full language, allowing the use of statistical statements. In this context, the first-order language is a very narrow special case, so that one might ask why these results are of interest. One answer is obvious. The different lower bounds that hold for a first-order language, clearly carry over to the general case. However, the results of this chapter turn out to be useful in the process of computing degrees of belief even for the full language (under the assumption, of course, that the knowledge base is unary). As we show in the next chapter, some of our algorithms can be combined with a maximum entropy computation in order to compute degrees of belief in the general case.

Chapter 7

The Maximum Entropy Connection

In the two previous chapters, we studied the problem of computing asymptotic conditional probabilities in the first-order case. In this chapter, we focus on the much more useful case where the knowledge base has statistical as well as first-order information. In light of the results of Chapters 5 and 6, for most of the chapter we restrict attention to the case when the knowledge base is expressed in a unary language. Our major result involves showing that asymptotic conditional probabilities can often by computed using *the principle of maximum entropy* [Jay78].

The idea of maximizing *entropy* has played an important role in many fields, including the study of probabilistic models for inferring degrees of belief [Jay57]. In the simplest setting, we can view entropy as a real-valued function on finite probability spaces. If Ω is a finite set and μ is a probability measure on Ω, the entropy $H(\mu)$ is defined to be $\Leftrightarrow\sum_{\omega\in\Omega}\mu(\omega)\ \ln\ \mu(\omega)$ (we take $0\ln\ 0 = 0$).

One standard application of entropy is the following. Suppose we know the space Ω, but have only partial information about μ, expressed in the form of constraints. For example, we might have a constraint such as $\mu(\omega_1) + \mu(\omega_2) \geq 1/3$. Although there may be many measures μ that are consistent with what we know, the *principle of maximum entropy* suggests that we adopt that μ^* which has the largest entropy among all the possibilities. Using the appropriate definitions, it can be shown that there is a sense in which this μ^* incorporates the "least" additional information [Jay57]. For example, if we have no constraints on μ, then μ^* will be the measure that assigns equal probability to all elements of Ω. Roughly speaking, μ^* assigns probabilities as equally as possible given the constraints.

Like maximum entropy, the random-worlds method also is used to determine degrees of belief (i.e., probabilities) relative to a knowledge base. Aside from this, there seems to be no obvious connection between the two approaches. Even the form of the "knowledge base" differs: the principle of maximum entropy applies to algebraic constraints on a probability distribution, whereas random-worlds uses assertions in a formal logic. Indeed, as long as the knowledge base makes use of a binary predicate symbol (or unary function symbol), we suspect that there is no useful connection between the two at all; see Section 7.3 for some discussion.

To understand the use of maximum entropy, suppose the vocabulary consists of the unary predicate symbols $P_1, \ldots, P_k$. We can consider the 2^k *atoms* that can be formed from these predicate symbols, namely, the formulas of the form $Q_1 \wedge \ldots \wedge Q_k$, where each Q_i is either P_i or $\neg P_i$. We can view the knowledge base as placing constraints on the proportion of domain elements satisfying each atom. For example, the constraint $\|P_1(x)|P_2(x)\|_x = 1/2$ says that the proportion of the domain satisfying some atom that contains P_2 as a conjunct is twice the proportion satisfying atoms that contain both P_1 and P_2 as conjuncts. Given a model of KB, we can define the entropy of this model as the entropy of the vector denoting the proportions of the different atoms. We show that, as N grows large, there are many more models with high entropy than with lower entropy. Therefore, models with high entropy dominate. We use this *concentration phenomenon* to show that our degree of belief in φ given KB according to the random-worlds method is closely related to the assignment of proportions to atoms that has maximum entropy among all assignments consistent with the constraints imposed by KB.

The concentration phenomenon relating entropy to the random-worlds method is well-known [Jay82]. In physics, the "worlds" are the possible configurations of a system typically consisting of many particles or molecules, and the mutually exclusive properties (our atoms) can be, for example, quantum states. The corresponding entropy measure is at the heart of statistical mechanics and thermodynamics. There are subtle but important differences between our viewpoint and that of the physicists. The main one lies in our choice of language. We want to express some intelligent agent's knowledge (which is why we take first-order logic as our starting point). The most specific difference concerns constant symbols. We need these because the most interesting questions for us arise when we have some knowledge about — and wish to assign degrees of belief to statements concerning — a particular individual. The parallel in physics would address properties of a single particle, which is generally considered to be well outside the scope of statistical mechanics.

Another work that examines the connection between random worlds and entropy from our point of view — computing degrees of belief for formulas in a particular logic — is that of Paris and Vencovska [PV89]. They restrict the knowledge base to consist of a conjunction of constraints that (in our notation) have the form $\|\alpha(x)|\beta(x)\|_x \approx r$ and $\|\alpha(x)\|_x \approx r$, where β and α are quantifier-free formulas involving unary predicates only, with no constant symbols. Not only is most of the expressive power of first-order logic not available in their approach, but the statistical information that can be expressed is quite limited. For example, it is not possible to make general assertions about statistical independence. Paris and Vencovska show that, for this language, the degree of belief can be computed using maximum entropy. As we have already suggested, we believe that a much richer language than this is called for. Our language allows arbitrary first-order assertions, full Boolean logic, arbitrary polynomial combinations of statistical expressions, and more.

In Section 7.1, we show that the connection between maximum entropy and random worlds can still be made in this much richer setting, although the results are much more difficult to prove. We then show, in Section 7.2, how maximum entropy can be used to compute degrees of belief in a large number of interesting cases. Using the techniques of Chapter 6, we show how maximum entropy can even be used to assign probabilities to non-unary formulas, so long as the knowledge base is unary and satisfies certain assumptions. We cannot make the connection for the full language, though. For one thing, as we hinted earlier, there are problems if we try to condition on a knowledge base that includes non-unary predicates. In addition, there are subtleties that arise involving the interaction between statistical information and first-order quantification. We feel that an important contribution of this chapter lies in pointing out the limitations of maximum entropy methods, particularly in the presence of non-unary predicates. Although the random-worlds method makes sense regardless of the vocabulary, it seems that once we allow non-unary predicate symbols in the language, we completely lose all connection between the random-worlds method and maximum entropy. We return to this last point in Section 7.3.

7.1 Degrees of belief and entropy

Let $\mathcal{L}_1^{\approx}$ be the sublanguage of $\mathcal{L}^{\approx}$ where only unary predicate symbols and constant symbols appear in formulas; in particular, we assume that equality ($=$) does not occur in formulas in $\mathcal{L}_1^{\approx}$.[1] Let $\mathcal{L}_1^{=}$ be the corresponding sublanguage of $\mathcal{L}^{=}$. In this section, we show that the expressive power of a knowledge base *KB* in the language $\mathcal{L}_1^{\approx}$ is quite limited. In fact, such a *KB* can essentially only place constraints on the proportions of the atoms. If we then think of these as constraints on the "probabilities of the atoms" we have the necessary ingredients to apply maximum entropy. We then show that there is a strong connection between the maximum entropy distribution found this way and the degree of belief generated by random-worlds method.

7.1.1 Unary Expressivity

To see what constraints a formula places on the probabilities of atoms, it is useful to convert the formula to a certain canonical form. As a first step to doing this, we recall the formal definition of atom from Definition 6.2.3. Let $\mathcal{P} = \{P_1, \ldots, P_k\}$ consist of the unary predicate symbols in the vocabulary Φ.

Definition 7.1.1: An *atom (over $\mathcal{P}$)* is conjunction of the form $P_1'(x) \wedge \ldots \wedge P_k'(x)$, where each P_i' is either P_i or $\neg P_i$. Since the variable x is irrelevant to our concerns, we suppress it and describe an atom as a conjunction of the form $P_1' \wedge \ldots \wedge P_k'$. ▮

Note that there are $2^{|\mathcal{P}|} = 2^k$ atoms over $\mathcal{P}$, and that they are mutually exclusive and exhaustive. Throughout this chapter, we use K to denote 2^k. We use $A_1, \ldots, A_K$ to denote the atoms over $\mathcal{P}$, listed in some fixed order.

Example 7.1.2: There are $K = 4$ atoms over $\mathcal{P} = \{P_1, P_2\}$: $A_1 = P_1 \wedge P_2$, $A_2 = P_1 \wedge \neg P_2$, $A_3 = \neg P_1 \wedge P_2$, $A_4 = \neg P_1 \wedge \neg P_2$. ▮

[1] We remark that many of our results can be extended to the case where the *KB* mentions equality, but the extra complexity added obscures many of the essential ideas.

The *atomic proportion terms* $||A_1(x)||_x, \ldots, ||A_K(x)||_x$ will play a significant role in our technical development. It turns out that $\mathcal{L}_1^{\approx}$ is a rather weak language; a formula $KB \in \mathcal{L}_1^{\approx}$ does little more than constrain the proportion of the atoms. In other words, for any such KB we can find an equivalent formula in which the only proportion expressions are these unconditional proportions of atoms. In particular, all of the more complex syntactic machinery in $\mathcal{L}_1^{\approx}$ — proportions over tuples, first-order quantification, nested proportions, and conditional proportions — does not add expressive power. (It does add convenience, however; knowledge can often be expressed far more succinctly if the full power of the language is used.)

Given any KB, the first step towards applying maximum entropy is to use $\mathcal{L}_1^{\approx}$'s lack of expressivity and replace all proportions by atomic proportions. It is also useful to make various other simplifications to KB that will help us in Section 7.2. We combine these steps and require that KB be transformed into a special *canonical form* which we now describe.

Definition 7.1.3: An *atomic term t over $\mathcal{P}$* is a polynomial over terms of the form $||A(x)||_x$, where A is an atom over $\mathcal{P}$. Such an atomic term t is *positive* if every coefficient of the polynomial t is positive. ▮

Definition 7.1.4: A (closed) sentence $\chi \in \mathcal{L}_1^{=}$ is in *canonical form* if it is a disjunction of conjunctions, where each conjunct is one of the following:

- $t' = 0$, $t' > 0 \wedge t \leq t'\varepsilon_i$, or $t' > 0 \wedge \neg(t \leq t'\varepsilon_i)$, where t and t' are atomic terms and t' is positive,
- $\exists x\, A_i(x)$ or $\neg\exists x\, A_i(x)$ some some atom A_i, or
- $A_i(c)$ for some atom A_i and some constant c.

Furthermore, a disjunct cannot contain both $A_i(c)$ and $A_j(c)$ for $i \neq j$ as conjuncts, nor can it contain both $A_i(c)$ and $\neg\exists x\, A_i(x)$. (This last condition is a minimal consistency requirement.) ▮

We now prove the following theorem, that extends Theorem 6.2.7 from Chapter 6 to the language $\mathcal{L}_1^{=}$.

Theorem 7.1.5: *Every formula in $\mathcal{L}_1^{=}$ is equivalent to a formula in canonical form. Moreover, there is an effective procedure that, given a formula $\xi \in \mathcal{L}_1^{=}$ constructs an equivalent formula $\widehat{\xi}$ in canonical form.*

We remark that the length of the formula $\widehat{\xi}$ is typically exponential in the length of ξ. Such a blowup seems inherent in any scheme defined in terms of atoms.

Theorem 7.1.5 is a generalization of Claim 5.7.1 in [Hal90]. It, in turn, is a generalization of a well-known result which says that any first-order formula with only unary predicates is equivalent to one with only depth-one quantifier nesting. Roughly speaking, this is because for a a quantified formula such as $\exists x\, \xi'$, subformulas talking about a variable y other than x can be moved outside the scope of the quantifier. This is possible because no literal subformula can talk about x and y together. Our proof uses the same idea and extends it to proportion statements. In particular, it shows that for any $\xi \in \mathcal{L}_1^{\approx}$ there is an equivalent $\hat{\xi}$ which has no nested quantifiers or nested proportions.

Notice, however, that such a result does not hold once we allow even a single binary predicate in the language. For example, the formula $\forall y\, \exists x\, R(x,y)$ clearly needs nested quantification because $R(x,y)$ talks about both x and y and so must remain within the scope of both quantifiers. With binary predicates, each additional depth of nesting really does add expressive power. This shows that there can be no "canonical form" theorem quite like Theorem 7.1.5 for richer languages. This issue is one of the main reasons why we restrict the *KB* to a unary language in this chapter. (See Section 7.3 for further discussion.)

Given any formula in in canonical form we can immediately derive from it, in a syntactic manner, a set of constraints on the possible proportions of atoms.

Definition 7.1.6: Let *KB* be in canonical form. We construct a formula , (KB) in the language of real closed fields (i.e., over the vocabulary $\{0, 1, +, \times\}$) as follows, where $u_1, \ldots, u_K$ are variables (distinct from the tolerance variables ε_j):

- we replace each occurrence of the formula $A_i(c)$ by $u_i > 0$,
- we replace each occurrence of $\exists x\, A_i(x)$ by $u_i > 0$ and replace each occurrence of $\neg\exists x\, A_i(x)$ by $u_i = 0$.
- we replace each occurrence of $||A_i(x)||_x$ by u_i. ∎

Notice that , (KB) has two types of variables: the new variables that we introduced, and the tolerance variables ε_i. In order to eliminate the dependence on the latter, we often consider the formula $KB[\vec{\tau}]$ for some tolerance vector $\vec{\tau}$.

Definition 7.1.7: Given a formula γ over the variables $u_1, \ldots, u_K$, let $Sol[\gamma]$ be the set of vectors in $\Delta^K = \{\vec{u} \in [0,1]^K : \sum_i^K u_i = 1\}$ satisfying γ. Formally, if $(a_1, \ldots, a_K) \in \Delta^K$, then $(a_1, \ldots, a_K) \in Sol[\gamma]$ iff $(I\!R, V) \models \gamma$, where V is a valuation such that $V(u_i) = a_i$. ∎

Definition 7.1.8: The *solution space of KB given* $\vec{\tau}$, denoted $S^{\vec{\tau}}[KB]$, is defined to be the closure of $Sol[, (KB[\vec{\tau}])]$. (We typically use $\overline{A}$ to denote the closure of a set in $I\!R^K$.) ∎

If *KB* is not in canonical form, we define , (KB) and $S^{\vec{\tau}}[KB]$ to be , $(\widehat{KB})$ and $S^{\vec{\tau}}[\widehat{KB}]$, respectively, where $\widehat{KB}$ is the formula in canonical form equivalent to *KB* obtained by the procedure in Theorem 7.1.5.

Example 7.1.9: Let $\mathcal{P}$ be $\{P_1, P_2\}$, with the atoms ordered as in Example 7.1.2. Consider

$$KB = \forall x\, P_1(x) \wedge 3||P_1(x) \wedge P_2(x)||_x \preceq_i 1.$$

The canonical formula $\widehat{KB}$ equivalent to KB is:[2]

$$\neg\exists x\, A_3(x) \wedge \neg\exists x\, A_4(x) \wedge 3||A_1(x)||_x \Leftrightarrow 1 \leq \varepsilon_i.$$

As expected, $\widehat{KB}$ constrains both $||A_3(x)||_x$ and $||A_4(x)||_x$ (i.e., u_3 and u_4) to be 0. We also see that $||A_1(x)||_x$ (i.e., u_1) is (approximately) at most 1/3. Therefore:

$$S^{\vec{\tau}}[KB] = \left\{(u_1, \ldots, u_4) \in \Delta^4 : u_1 \leq 1/3 + \tau_i/3, u_3 = u_4 = 0\right\}. \quad \blacksquare$$

7.1.2 The concentration phenomenon

With every world $W \in \mathcal{W}^*$, we can associate a particular tuple $(u_1, \ldots, u_K)$ where u_i is the fraction of the domain satisfying atom A_i in W.

Definition 7.1.10: Given a world $W \in \mathcal{W}^*$, we define $\pi(W) \in \Delta^K$ to be

$$(||A_1(x)||_x, ||A_2(x)||_x, \ldots, ||A_K(x)||_x)$$

where the values of the proportions are interpreted over W. The vector $\pi(W)$ is also defined to be the *point* associated with W. ∎

We define the *entropy* of any model W to be the entropy of $\pi(W)$; that is, if $\pi(W) = (u_1, \ldots, u_K)$, then the entropy of W is $H(u_1, \ldots, u_K)$. As we are about to show, the entropy of $\vec{u}$ turns out to be a very good (asymptotic) indicator of how many worlds W there are such that $\pi(W) = \vec{u}$. In fact, there are so many more worlds near points of high entropy that we can ignore all the other points when computing degrees of belief. This *concentration* phenomenon, as Jaynes [Jay82] has called it, which is essentially the content of the next lemma, justifies our interest in the the maximum entropy point(s) of $S^{\vec{\tau}}[KB]$.

For any $\mathcal{S} \subseteq \Delta^K$ let $\#worlds^{\vec{\tau}}_N[\mathcal{S}](KB)$ denote the number of worlds W of size N such that $(W, \vec{\tau}) \models KB$ and such that $\pi(W) \in \mathcal{S}$; for any $\vec{u}$, let $\#worlds^{\vec{\tau}}_N[\vec{u}](KB)$ abbreviate $\#worlds^{\vec{\tau}}_N[\{\vec{u}\}](KB)$. Of course $\#worlds^{\vec{\tau}}_N[\vec{u}](KB)$ is zero unless all components of $\vec{u}$ are multiples of $1/N$. However, if there are any models associated with $\vec{u}$ at all, we can estimate their number quite accurately using the entropy function:

Lemma 7.1.11: *There exist some function $h : I\!N \rightarrow I\!N$ and two strictly positive polynomial functions $f, g : I\!N \rightarrow I\!R$ such that, for $KB \in \mathcal{L}_1^{\approx}$ and $\vec{u} \in \Delta^K$, if $\#worlds^{\vec{\tau}}_N[\vec{u}](KB) \neq 0$, then in fact*

$$(h(N)/f(N))e^{NH(\vec{u})} \leq \#worlds^{\vec{\tau}}_N[\vec{u}](KB) \leq h(N)g(N)e^{NH(\vec{u})}.$$

Of course, it follows from the lemma that tuples whose entropy is near maximum have overwhelmingly more worlds associated with them than tuples whose entropy is further from maximum. This is essentially the concentration phenomenon.

Lemma 7.1.11 is actually fairly straightforward to prove. The following simple example illustrates the basic idea.

[2] Note that here we are viewing KB as a formula in $\mathcal{L}^{=}$, under the tranlsation defined earlier; we do this throughout the chapter without further comment.

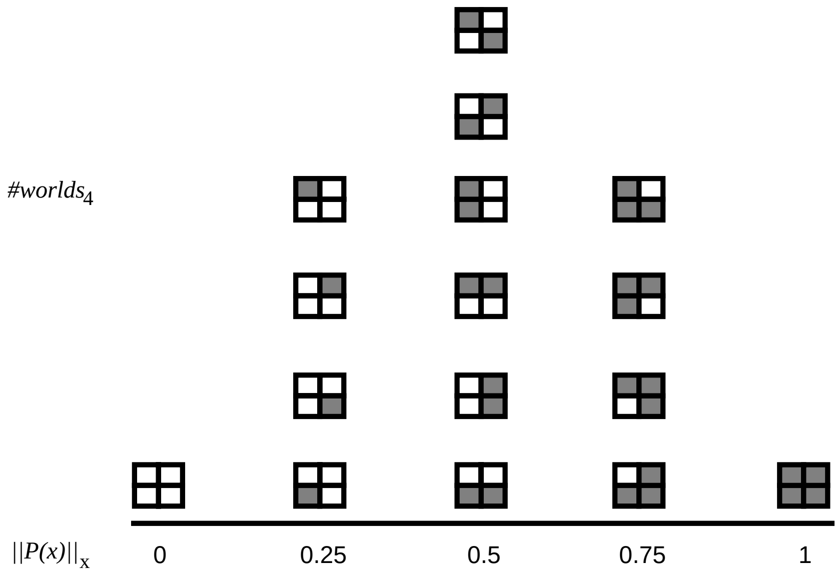

Figure 7.1: Partition of $\mathcal{W}_4$ according to $\pi(W)$.

Example 7.1.12: Suppose $\Phi = \{P\}$ and $KB = \mathit{true}$. We have

$$\Delta^K = \Delta^2 = \{(u_1, 1 \Leftrightarrow u_1) : \quad 0 \leq u_1 \leq 1\},$$

where the atoms are $A_1 = P, A_2 = \neg P$. For any N, partition the worlds in $\mathcal{W}_N$ according to the point to which they correspond. For example, the graph in Figure 7.1 shows us the partition of $\mathcal{W}_4$. In general, consider some point $\vec{u} = (r/N, (N \Leftrightarrow r)/N)$. The number of worlds corresponding to $\vec{u}$ is simply the number of ways of choosing the denotation of P. That is, we need to choose which r elements satisfy P; hence, the number of such worlds is $\binom{N}{r} = \frac{N!}{r!(N-r)!}$. Figure 7.2 shows the qualitative behavior of this function for large values of N.

We can estimate the factorials appearing in this expression using Stirling's approximation, which asserts that the factorial $m!$ is approximately $m^m = e^{m \ln m}$. So, after substituting for the three factorials, we can estimate $\binom{N}{r}$ as $e^{N \log N - (r \log r + (N-r)\log(N-r))}$, which reduces to $e^{NH(\vec{u})}$. The entropy term in the general case arises from the use of Stirling's approximation in an analogous way. (A more careful estimate is done in the proof of Lemma 7.1.11 in the appendix.) ▮

Because of the exponential dependence on N times the entropy, the number of worlds associated with points of high entropy swamp all other worlds as N grows large. This concentration phenomenon, well-known in the field of statistical physics, forms the basis for our main result in this section. It asserts that it is possible to compute degrees of belief according to random worlds while ignoring all but those worlds whose entropy is near maximum. The next theorem is a formal statement of precisely this phenomenon.

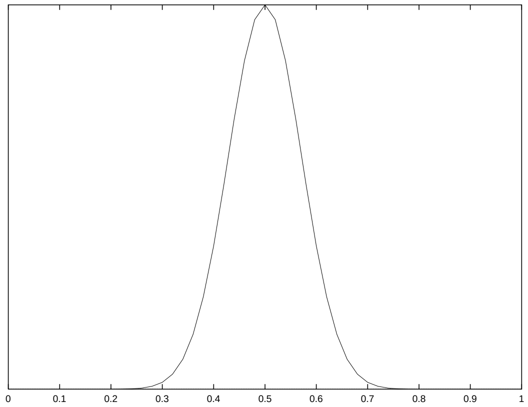

Figure 7.2: Concentration phenomenon for worlds of size N.

Theorem 7.1.13: *For all sufficiently small $\vec{\tau}$, the following is true. Let $\mathcal{Q}$ be the points with greatest entropy in $S^{\vec{\tau}}[KB]$ and let $\mathcal{O} \subseteq I\!\!R^K$ be any open set containing $\mathcal{Q}$. Then for all $\theta \in \mathcal{L}^{\approx}$ and for* $\lim^* \in \{\limsup, \liminf\}$*:*

$$\lim_{N\to\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\theta|KB) = \lim_{N\to\infty}{}^{*} \frac{\#\mathit{worlds}_N^{\vec{\tau}}[\mathcal{O}](\theta \wedge KB)}{\#\mathit{worlds}_N^{\vec{\tau}}[\mathcal{O}](KB)}.$$

In general Theorem 7.1.13 may seem to be of limited usefulness: knowing that we only have to look at worlds near the maximum entropy point does not substantially reduce the number of worlds we need to consider. (Indeed, the whole point of the concentration phenomenon is that almost all worlds have high entropy.) Nevertheless, as the rest of this chapter shows, this result can be very useful when combined with the following two results. The first of these says that if all the worlds near the maximum entropy points have a certain property, then we should have degree of belief 1 that this property is true.

Corollary 7.1.14: *For all sufficiently small $\vec{\tau}$, the following is true. Let $\mathcal{Q}$ be the points with greatest entropy in $S^{\vec{\tau}}[KB]$, let $\mathcal{O} \subseteq I\!\!R^K$ be an open set containing $\mathcal{Q}$, and let $\theta[\mathcal{O}] \in \mathcal{L}^{=}$ be an assertion that holds for any world W such that $\pi(W) \in \mathcal{O}$. Then*

$$\Pr{}_\infty^{\vec{\tau}}(\theta[\mathcal{O}]|KB) = 1.$$

Example 7.1.15: For the knowledge base *true* in Example 7.1.12, it is easy to see that the maximum entropy point is $(0.5, 0.5)$. Fix some arbitrary $\epsilon > 0$. Clearly, there is some open set

$\mathcal{O}$ around this point such that the assertion $\theta = ||P(x)||_x \in [0.5 \Leftrightarrow \epsilon, 0.5 + \epsilon]$ holds for any world in $\mathcal{O}$. Therefore, we can conclude that

$$\mathrm{Pr}^{\vec{\tau}}_{\infty}(||P(x)||_x \in [0.5 \Leftrightarrow \epsilon, 0.5 + \epsilon] \, |\mathit{true}) = 1. \ \blacksquare$$

This corollary turns out to be surprisingly useful, for the following reason. As we showed in Section 4.1, formulas θ with degree of belief 1 can essentially be treated just like other knowledge in KB. That is, for any other formula φ, the degrees of belief relative to KB and $KB \wedge \theta$ will be identical (even if KB and $KB \wedge \theta$ are not logically equivalent). However, as we show in the next section, it may be possible to apply certain techniques to $KB \wedge \theta$ which cannot be used for KB. For convenience, we restate and prove a slightly more general version of this fact:

Theorem 7.1.16: *If* $\mathrm{Pr}^{\vec{\tau}}_{\infty}(\theta|KB) = 1$ *and* $\lim^* \in \{\limsup, \liminf\}$*, then:*

$$\lim_{N \to \infty}{}^{*} \mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB) = \lim_{N \to \infty}{}^{*} \mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB \wedge \theta).$$

Proof: Basic probabilistic reasoning shows that, for any N and $\vec{\tau}$:

$$\mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB) = \mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB \wedge \theta)\, \mathrm{Pr}^{\vec{\tau}}_{N}(\theta|KB) \ + \mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB \wedge \neg\theta)\, \mathrm{Pr}^{\vec{\tau}}_{N}(\neg\theta|KB).$$

By assumption, $\mathrm{Pr}^{\vec{\tau}}_{N}(\theta|KB)$ tends to 1 when we take limits, so the first term tends to $\mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB \wedge \theta)$. On the other hand, $\mathrm{Pr}^{\vec{\tau}}_{N}(\neg\theta|KB)$ has limit 0. Because $\mathrm{Pr}^{\vec{\tau}}_{N}(\varphi|KB \wedge \neg\theta)$ is bounded, we conclude that the second product also tends to 0. The result follows. $\blacksquare$

7.2 Computing degrees of belief

Although the concentration phenomenon is interesting, its application to actually computing degrees of belief may not be obvious. Since we know that almost all worlds will have high entropy, a direct application of Theorem 7.1.13 does not substantially reduce the number of worlds we must consider. Yet, as we show in this section, the concentration theorem can form the basis of a practical technique for computing degrees of belief in many cases. We begin in Section 7.2.1 by presenting the intuitions underlying this technique. In Section 7.2.2 we build on these intuitions by presenting results for a restricted class of formulas: those queries which are quantifier-free formulas over a unary language with a single constant symbol. In spite of this restriction, many of the issues arising in the general case can be seen here. Moreover, as we show in Section 7.2.3, this restricted sublanguage is rich enough to allow us to embed two well-known propositional approaches that make use of maximum entropy: *probabilistic logic*, due to Nilsson [Nil86]; and the maximum-entropy extension of ϵ-semantics [GP90] due to Goldszmidt, Morris, Pearl [GMP90]. In Section 7.2.4, we consider to what extend the results for the restricted language can be extended. We show that they can, but numerous subtleties arise.

7.2.1 The general strategy

Although the random-worlds method is defined by counting worlds, we can sometimes find more direct ways to calculate the degrees of belief it yields. In Chapter 4 we present a number of such techniques, most of which only apply in very special cases. One of the simplest and most intuitive is the basic *direct inference* theorem, which we restate for convenience.

Theorem 4.2.1: *Let KB be a knowledge base of the form $\psi(\vec{c}) \wedge KB'$, and assume that for all sufficiently small tolerance vectors $\vec{\tau}$,*

$$KB[\vec{\tau}] \models \|\varphi(\vec{x})|\psi(\vec{x})\|_{\vec{x}} \in [\alpha, \beta].$$

If no constant in $\vec{c}$ appears in KB', in $\varphi(\vec{x})$, or in $\psi(\vec{x})$, then $\Pr_\infty(\varphi(\vec{c})|KB) \in [\alpha, \beta]$.

This result, in combination with the results of the previous section, provides us with a very powerful tool. Roughly speaking, we propose to use the following strategy: The basic concentration phenomenon says that most worlds are very similar in a certain sense. As shown in Corollary 7.1.14, we can use this to find some assertions that are "almost certainly" true (i.e., with degree of belief 1) even if they are not logically implied by KB. Theorem 7.1.16 then tells us that we can treat these new assertions as if they are in fact known with certainly. When these new assertions state statistical "knowledge", they can vastly increase our opportunities to apply direct inference. The following example illustrates this idea.

Example 7.2.1: Consider a very simple knowledge base over a vocabulary containing the single unary predicate $\{P\}$:

$$KB = (||P(x)||_x \preceq_1 0.3).$$

There are two atoms A_1 and A_2 over $\mathcal{P}$, with $A_1 = P$ and $A_2 = \neg P$. The solution space of this KB given $\vec{\tau}$ is clearly

$$S^{\vec{\tau}}[KB] = \{(u_1, u_2) \in \Delta^2 \ : \ u_1 \leq 0.3 + \tau_1\}.$$

A straightforward computation shows that, for $\tau_1 < 0.2$, this has a unique maximum entropy point $\vec{v} = (0.3 + \tau_1, 0.7 \Leftrightarrow \tau_1)$.

Now, consider the query $P(c)$. For any $\epsilon > 0$, let $\theta[\epsilon]$ be the formula $||P(x)||_x \in [(0.3+\tau_1) \Leftrightarrow \epsilon, (0.3+\tau_1)+\epsilon]$. This satisfies the condition of Corollary 7.1.14, so it follows that $\Pr^{\vec{\tau}}_\infty(\theta[\epsilon]|KB) = 1$. Using Theorem 7.1.16, we know that for $\lim^* \in \{\liminf, \limsup\}$,

$$\lim_{N\to\infty}{}^* \Pr^{\vec{\tau}}_N(P(c)|KB) = \lim_{N\to\infty}{}^* \Pr^{\vec{\tau}}_N(P(c)|KB \wedge \theta[\epsilon]).$$

But now we can use direct inference. (Note that here, our "knowledge" about c is vacuous, i.e., "$true(c)$".) We conclude that, if there is any limit at all, then necessarily

$$\Pr^{\vec{\tau}}_\infty(P(c)|KB \wedge \theta[\epsilon]) \in [(0.3 + \tau_1) \Leftrightarrow \epsilon, (0.3 + \tau_1) + \epsilon].$$

So, for any $\epsilon > 0$,

$$\Pr^{\vec{\tau}}_\infty(P(c)|KB) \in [(0.3 + \tau_1) \Leftrightarrow \epsilon, (0.3 + \tau_1) + \epsilon].$$

Since this is true for all ϵ, the only possible value for $\Pr_\infty^{\vec{\tau}}(P(c)|KB)$ is $0.3 + \tau_1$, which is the value of u_1 (which represents $||P(x)||_x$) *at* the maximum entropy point. Note that it is also clear what happens as $\vec{\tau}$ tends to $\vec{0}$. Thus, $\Pr_\infty(P(c)|KB)$ is 0.3. ∎

This example demonstrates the main steps of one strategy for computing degrees of belief. First the maximum entropy points of the space $S^{\vec{\tau}}[KB]$ are computed as a function of $\vec{\tau}$. Then, these are used to compute $\Pr_\infty^{\vec{\tau}}(\varphi|KB)$, assuming the limit exists (if not, the lim sup and lim inf of $\Pr_N(\varphi|KB)$ are computed instead). Finally, we compute the limit of this probability as $\vec{\tau}$ goes to zero.

This strategy has a serious potential problem: we clearly cannot compute $\Pr_\infty^{\vec{\tau}}(\varphi|KB)$ separately for each of the infinitely many tolerance vectors $\vec{\tau}$ and then take the limit as $\vec{\tau}$ goes to 0. We might hope to *parametrically* compute this probability as a function of $\vec{\tau}$, and then compute the limit. But there is no reason to believe that $\Pr_\infty^{\vec{\tau}}(\varphi|KB)$ is, in general, an easily characterizable function of $\vec{\tau}$, which will make computing the limit as $\vec{\tau}$ goes to 0 difficult. We can, however, often avoid this limiting process, if the maximum entropy points of $S^{\vec{\tau}}[KB]$ converge to the maximum entropy points of $S^{\vec{0}}[KB]$. (For future reference, notice that $S^{\vec{0}}[KB]$ is the space defined by the closure of the constraints obtained from KB by replacing all occurrences of $\approx_i$ by $=$ and all occurrences of $\preceq_i$ by $\leq$.) In many such cases, we can compute $\Pr_\infty(\varphi|KB)$ directly in terms of the maximum entropy points of $S^{\vec{0}}[KB]$, without taking limits at all.

As the following example shows, this type of continuity does not hold in general: the maximum entropy points of $S^{\vec{\tau}}[KB]$ do not necessarily converge to those of $S^{\vec{0}}[KB]$.

Example 7.2.2: Consider the knowledge base

$$KB = (||P(x)||_x \approx_1 0.3 \vee ||P(x)||_x \approx_2 0.4) \wedge ||P(x)||_x \not\approx_3 0.4 \ .$$

It is easy to see that $S^{\vec{0}}[KB]$ is just $\{(0.3.0.7)\}$: The point $(0.4, 0.6)$ is disallowed by the second conjunct. Now, consider $S^{\vec{\tau}}[KB]$ for $\vec{\tau} > \vec{0}$. If $\tau_2 \leq \tau_3$, then $S^{\vec{\tau}}[KB]$ indeed does not contain points where u_1 is near 0.4; the maximum entropy point of this space is easily seen to be $0.3+\tau_1$. However, if $\tau_2 > \tau_3$ then there will be points in $S^{\vec{\tau}}[KB]$ where u_1 is around 0.4: those where $0.4+\tau_3 < u_1 \leq 0.4+\tau_2$. Since these points have a higher entropy than the points in the vicinity of 0.3, the former will dominate. Thus, the set of maximum entropy points of $S^{\vec{\tau}}[KB]$ does not converge to a single well-defined set. What it converges to (if anything) depends on how $\vec{\tau}$ goes to $\vec{0}$. This nonconvergence has consequences as far as degrees of belief go. It is not hard to show $\Pr_\infty^{\vec{\tau}}(P(c)|KB)$ can be $0.3 + \tau_1$, $0.4 + \tau_2$, or 0.5, depending on the precise relationship between τ_1, τ_2, and τ_3. It follows that $\Pr_\infty(P(c)|KB)$ does not exist. ∎

We say that a degree of belief $\Pr_\infty(\varphi|KB)$ is not *robust* if the behavior of $\Pr_\infty^{\vec{\tau}}(\varphi|KB)$ (or of $\liminf \Pr_N^{\vec{\tau}}(\varphi|KB)$ and $\limsup \Pr_N^{\vec{\tau}}(\varphi|KB)$) as $\vec{\tau}$ goes to $\vec{0}$ depends on *how* $\vec{\tau}$ goes to $\vec{0}$. In other worlds, nonrobustness describes situations when $\Pr_\infty(\varphi|KB)$ does not exist because of sensitivity to the exact choice of tolerances. We shall see a number of other examples of nonrobustness in later sections.

It might seem that the notion of robustness is an artifact of our approach. In particular, it seems to depend on the fact that our language has the expressive power to say that the two tolerances represent a different degree of approximation, simply by using different subscripts ($\approx_2$ vs. $\approx_3$ in the example). In an approach to representing approximate equality that does not make these distinctions (as, for example, the approach taken in [PV89]), we are bound to get the answer 0.3 in the example above, since then $||P(x)||_x \not\approx_3 0.4$ really would be the negation of $||P(x)||_x \approx_2 0.4$. We would argue that the answer 0.3 is not as reasonable as it might at first seem. Suppose one of the two different instances of 0.4 in the previous example had been slightly different; for example, suppose we had used 0.399 rather than 0.4 in the first of them. In this case, the second conjunct is essentially vacuous, and can be ignored. The maximum entropy point in $S^{\vec{0}}[KB]$ is now 0.399, and we will, indeed, derive a degree of belief of 0.399 in $P(c)$. Thus, arbitrarily small changes to the numbers in the original knowledge base can cause large changes in our degrees of belief. But these numbers are almost always the result of approximate observations; this is reflected by our decision to use approximate equality rather then equality when referring to them. It therefore does not seem reasonable to base actions on a degree of belief that can change so drastically in the face of small changes in the measurement of data. Note that, if we know that the two instances of 0.4 do, in fact, denote exactly the same number, we can represent this by using the same approximate equality connective in both disjuncts. In this case, it is easy to see that we do get the answer 0.3.

A close look at the example shows that the nonrobustness arises because of the negated proportion expression $||P(x)||_x \not\approx_3 0.4$. Indeed, we can show that if we start with a KB in canonical form that does not contain negated proportion expressions, then in a precise sense the set of maximum entropy points of $S^{\vec{\tau}}[KB]$ does converge to the set of maximum entropy points of $S^{\vec{0}}[KB]$. An argument can be made that we should eliminate negated proportion expressions from the language altogether. It is one thing to argue that sometimes we have statistical values whose accuracy we are unsure about, so that we want to make logical assertions less stringent than exact numerical equality. It is harder to think of cases in which the opposite is true, and *all* we know is that some statistic is "not even approximately" equal to some value. We do not eliminate negated proportion expressions from the language, however, since without them we would not be able to prove an analogue to Theorem 7.1.5. (They arise when we try to flatten nested proportion expressions, for example.) Instead, we have identified a weaker condition that is sufficient to prevent problems such as that seen in Example 7.2.2. *Essential positivity* simply tests that negations are not interacting with the maximum entropy computation in a harmful way.

Definition 7.2.3: Let $,^{\le}(KB[\vec{0}])$ be the result of replacing each strict inequality in $,(KB[\vec{0}])$ with its weakened version. More formally, we replace each subformula of the form $t < 0$ with $t \le 0$, and each subformula of the form $t > 0$ with $t \ge 0$. (Recall that these are the only constraints possible in $,(KB[\vec{0}])$, since all tolerance variables ε_i are assigned 0.) Let $S^{\le\vec{0}}[KB]$ be $Sol[,^{\le}(KB[\vec{0}])]$. We say that KB is *essentially positive* if the sets $S^{\le\vec{0}}[KB]$ and $S^{\vec{0}}[KB]$ have the same maximum entropy points. ▮

Example 7.2.4: Consider again the knowledge base KB from Example 7.2.2. The constraint formula , $(KB[\vec{0}])$ is (after simplification):

$$(u_1 = 0.3 \vee u_1 = 0.4) \wedge (u_1 < 0.4 \vee u_1 > 0.4).$$

Its "positive" version is , $^{\leq}(KB[\vec{0}])$:

$$(u_1 = 0.3 \vee u_1 = 0.4) \wedge (u_1 \leq 0.4 \vee u_1 \geq 0.4),$$

which is clearly equivalent to $u_1 = 0.3 \vee u_1 = 0.4$. Thus, $S^{\vec{0}}[KB] = \{(u_1, u_2) \in \Delta^2 \ : \ u_1 \leq 0.3\}$ whereas $S^{\leq\vec{0}}[KB] = S^{\vec{0}}[KB] \cup \{(0.4, 0.6)\}$. Since the two spaces have different maximum entropy points, the knowledge base KB is not essentially positive. ▮

As the following result shows, essential positivity suffices to guarantee that the maximum entropy points of $S^{\vec{\tau}}[KB]$ converge to those of $S^{\vec{0}}[KB]$.

Proposition 7.2.5: *Assume that KB is essentially positive and let Q be the set of maximum entropy points of $S^{\vec{0}}[KB]$ (and thus also of $S^{\leq\vec{0}}[KB]$). Then for all $\epsilon > 0$ and all sufficiently small tolerance vectors $\vec{\tau}$ (where "sufficiently small" may depend on ϵ), every maximum entropy point of $S^{\vec{\tau}}[KB]$ is within ϵ of some maximum entropy point in Q.*

7.2.2 Queries for a single individual

We now show how to compute $\Pr_\infty(\varphi|KB)$ for a certain restricted class of first-order formulas φ and knowledge bases KB. Most significantly, the query φ is a quantifier-free (first-order) sentence over the vocabulary $\mathcal{P} \cup \{c\}$; thus, it is a query about a single individual — c. While this class is rather restrictive, it suffices to express a large body of real-life examples. Moreover, it is significantly richer than the language considered by Paris and Vencovska [PV89].

The following definition helps define the class of interest.

Definition 7.2.6: A formula is *essentially propositional* if it is a quantifier-free and proportion-free formula in the language $\mathcal{L}^{\approx}(\{P_1, \ldots, P_k\})$ (so that it has no constant symbols) and has only one free variable, namely x. ▮

In this section, we focus on computing the degree of belief $\Pr_\infty(\varphi(c)|KB)$ for formulas We say that $\varphi(c)$ is a *simple query for KB* if:

- $\varphi(x)$ is essentially propositional,
- KB is of the form $\psi(c) \wedge KB'$, where $\psi(x)$ is essentially propositional and KB' does not mention c.

Thus, just as in Theorem 4.2.1, we assume that $\psi(c)$ summarizes all this is known about c. In this section, we focus on computing the degree of belief $\Pr_\infty(\varphi(c)|KB)$ for a formula $\varphi(c)$ which is a simple query for KB.

Note that any essentially propositional formula $\xi(x)$ is equivalent to a disjunction of atoms. For example, over the vocabulary $\{P_1, P_2\}$, the formula $P_1(x) \vee P_2(x)$ is equivalent to $A_1(x) \vee A_2(x) \vee A_3(x)$ (where the atoms are ordered as in Example 7.1.2). For any essentially propositional formula ξ, we take $\mathcal{A}(\xi)$ be the (unique) set of atoms such that ξ is equivalent to $\bigvee_{A_j \in \mathcal{A}(\xi)} A_j(x)$.

If we view a tuple $\vec{u} \in \Delta^K$ as a probability assignment to the atoms, we can extend $\vec{u}$ to a probability assignment on all essentially propositional formulas using this identification of an essentially propositional formula with a set of atoms:

Definition 7.2.7: Let ξ be an essentially propositional formula. We define a function $F_{[\xi]} : \Delta^K \to I\!R$ as follows:

$$F_{[\xi]}(\vec{u}) = \sum_{A_j \in \mathcal{A}(\xi)} u_j.$$

For essentially propositional formulas $\varphi(x)$ and $\psi(x)$ we define the (partial) function $F_{[\varphi|\psi]} : \Delta^K \to I\!R$ to be:

$$F_{[\varphi|\psi]}(\vec{u}) = \frac{F_{[\varphi \wedge \psi]}(\vec{u})}{F_{[\psi]}(\vec{u})}.$$

Note that this function is undefined when $F_{[\psi]}(\vec{u}) = 0$. ▮

As the following result shows, if φ is a simple query for *KB*, then all that matters in computing $\Pr_\infty(\varphi|KB)$ is $F_{[\varphi|\psi]}(\vec{u})$ for tuples $\vec{u}$ of maximum entropy. Thus, in a sense, we are only using KB' to determine the space over which we maximize entropy. Having defined this space, we can focus on ψ and φ in determining the degree of belief.

Theorem 7.2.8: *Suppose $\varphi(c)$ is a simple query for KB. For all $\vec{\tau}$ sufficiently small, if $\mathcal{Q}$ is the set of maximum entropy points in $S^{\vec{\tau}}[KB]$ and $F_{[\psi]}(\vec{v}) > 0$ for all $\vec{v} \in \mathcal{Q}$, then for* $\lim^* \in \{\limsup, \liminf\}$

$$\lim_{N\to\infty}{}^{*} \Pr{}^{\vec{\tau}}_N(\varphi(c)|KB) \in \left[\inf_{\vec{v}\in\mathcal{Q}} F_{[\varphi|\psi]}(\vec{v}), \sup_{\vec{v}\in\mathcal{Q}} F_{[\varphi|\psi]}(\vec{v})\right].$$

The following is an immediate but important corollary of this theorem. It asserts that, if the space $S^{\vec{\tau}}[KB]$ has a unique maximum entropy point, then its value uniquely determines the probability $\Pr^{\vec{\tau}}_\infty(\varphi(c)|KB)$.

Corollary 7.2.9: *Suppose $\varphi(c)$ is a simple query for KB. For all $\vec{\tau}$ sufficiently small, if $\vec{v}$ is the unique maximum entropy point in $S^{\vec{\tau}}[KB]$ and $F_{[\psi]}(\vec{v}) > 0$, then*

$$\Pr{}^{\vec{\tau}}_\infty(\varphi(c)|KB) = F_{[\varphi|\psi]}(\vec{v}).$$

We are interested in $\Pr_\infty(\varphi(c)|KB)$, which means that we are interested in the limit of $\Pr^{\vec{\tau}}_\infty(\varphi(c)|KB)$ as $\vec{\tau} \to \vec{0}$. As we observed in the previous section, if *KB* is essentially positive, then by continuity we can compute this by looking directly at the maximum entropy points of $S^{\vec{0}}[KB]$. By combining Theorem 7.2.8 with Proposition 7.2.5, we can show:

Theorem 7.2.10: *Suppose $\varphi(c)$ is a simple query for KB. If the space $S^{\vec{0}}[KB]$ has a unique maximum entropy point $\vec{v}$, KB is essentially positive, and $F_{[\psi]}(\vec{v}) > 0$, then*

$$\Pr_\infty(\varphi(c)|KB) = F_{[\varphi|\psi]}(\vec{v}).$$

How applicable is this theorem? As our examples and the discussion in the next section shows, we often do get simple queries and knowledge bases that are essentially positive. What about the assumption of a unique maximum entropy point? Since the entropy function is convex, this assumption is automatically satisfied if $S^{\vec{0}}[KB]$ is a *convex* space. Recall that a space S is convex if for all $\vec{u}, \vec{u}' \in S$, and all $\alpha \in [0,1]$, it is also the case that $\alpha\vec{u} + (1 \Leftrightarrow \alpha)\vec{u}' \in S$. Furthermore, the space $S^{\vec{0}}[KB]$ is clearly convex if it is defined using a conjunction of linear constraints. While it is clearly possible to create knowledge bases where $S^{\vec{0}}[KB]$ has multiple maximum entropy points (for example, using disjunctions), we expect that such knowledge bases arise rarely in practical applications. Perhaps the most restrictive assumption made by this theorem is the seemingly innocuous requirement that $F_{[\psi]}(\vec{v}) > 0$. This assumption is obviously necessary to the theorem as stated: without it, the function $F_{[\varphi|\psi]}$ is simply not defined. Unfortunately, we show in Section 7.2.4 that this requirement is, in fact, a severe one; in particular, it prevents the theorem from being applied to most examples derived from default reasoning, using our statistical interpretation of defaults (described in Section 3.3).

We close this subsection with an example of the theorem in action.

Example 7.2.11: Let the language consist of $\mathcal{P} = \{\mathit{Hepatitis}, \mathit{Jaundice}, \mathit{BlueEyed}\}$, and the constant *Eric*. There are eight atoms in this language. We use $A_{Q_1Q_2Q_3}$ to denote the atom $Q_1(x) \wedge Q_2(x) \wedge Q_3(x)$, where Q_1 is either H (denoting *Hepatitis*) or $\overline{H}$ (denoting $\neg\mathit{Hepatitis}$), Q_2 is J or $\overline{J}$ (for *Jaundice* and $\neg\mathit{Jaundice}$, respectively), and Q_3 is B or $\overline{B}$ (for *BlueEyed* and $\neg\mathit{BlueEyed}$, respectively).

Consider the knowledge base KB_{hep}:

$$\begin{array}{l}
\forall x\,(\mathit{Hepatitis}(x) \Rightarrow \mathit{Jaundice}(x)) \wedge \\
||\mathit{Hepatitis}(x)|\mathit{Jaundice}(x)||_x \approx_1 0.8 \wedge \\
||\mathit{BlueEyed}(x)||_x \approx_2 0.25 \wedge \\
\mathit{Jaundice}(\mathit{Eric}).
\end{array}$$

If we list the atoms in the order

$$A_{HJB}, A_{HJ\overline{B}}, A_{H\overline{J}B}, A_{H\overline{J}\overline{B}}, A_{\overline{H}JB}, A_{\overline{H}J\overline{B}}, A_{\overline{H}\overline{J}B}, A_{\overline{H}\overline{J}\overline{B}},$$

then it is not hard to show that , (KB_{hep}) is:

$$\begin{array}{lcll}
u_3 & = & 0 & \wedge \\
u_4 & = & 0 & \wedge \\
(u_1 + u_2) & \leq & (0.8 + \varepsilon_1)(u_1 + u_2 + u_5 + u_6) & \wedge \\
(u_1 + u_2) & \geq & (0.8 \Leftrightarrow \varepsilon_1)(u_1 + u_2 + u_5 + u_6) & \wedge \\
(u_1 + u_3 + u_5 + u_7) & \leq & (0.25 + \varepsilon_2) & \wedge \\
(u_1 + u_3 + u_5 + u_7) & \geq & (0.25 \Leftrightarrow \varepsilon_2) & \wedge \\
(u_1 + u_2 + u_5 + u_6) & > & 0. &
\end{array}$$

To find the space $S^{\vec{0}}[KB_{hep}]$ we simply set $\varepsilon_1 = \varepsilon_2 = 0$. Then it is quite straightforward to find the maximum entropy point in this space, which, taking $\gamma = 2^{1.6}$, is:

$$(v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8) = \left(\frac{1}{5+\gamma}, \frac{3}{5+\gamma}, 0, 0, \frac{1}{4(5+\gamma)}, \frac{3}{4(5+\gamma)}, \frac{\gamma}{4(5+\gamma)}, \frac{3\gamma}{4(5+\gamma)} \right).$$

Using $\vec{v}$, we can compute various asymptotic probabilities very easily. For example,

$$\begin{aligned} \Pr_\infty(\mathit{Hepatitis}(\mathit{Eric}) | KB_{hep}) &= F_{[\mathit{Hepatitis}|\mathit{Jaundice}]}(\vec{v}) \\ &= \frac{v_1 + v_2}{v_1 + v_2 + v_5 + v_6} \\ &= \frac{\frac{1}{5+\gamma} + \frac{3}{5+\gamma}}{\frac{1}{5+\gamma} + \frac{3}{5+\gamma} + \frac{1}{4(5+\gamma)}, \frac{3}{4(5+\gamma)}} = 0.8, \end{aligned}$$

as expected. Similarly,

$$\Pr_\infty(\mathit{BlueEyed}(\mathit{Eric}) | KB_{hep}) = 0.25$$

and

$$\Pr_\infty(\mathit{BlueEyed}(\mathit{Eric}) \wedge \mathit{Hepatitis}(\mathit{Eric}) | KB_{hep} = 0.2\ .$$

Note that the first two answers also follow from the direct inference principle (Theorem 4.2.1), which happens to be applicable in this case. ∎

7.2.3 Probabilistic propositional logic

In this section we consider two well-known approaches to probabilistic propositional logic. As the following discussion shows, probabilistic propositional logic in general, and these two approaches in particular, can easily be embedded in our framework as simple queries. Let $p_1, \ldots, p_k$ be a set of primitive propositions. Nilsson [Nil86] considered the problem of reasoning about the probabilities of certain formulas over these propositions, given constraints over the probabilities of other formulas. For example, we might know that $\Pr(\mathit{fly}|\mathit{bird}) \geq 0.7$, that $\Pr(\mathit{yellow}) \leq 0.2$, and be interested in the probability $\Pr(\mathit{fly}|\mathit{bird} \wedge \mathit{yellow})$. Although it is not always easy to see what $\Pr(\mathit{fly}|\mathit{bird})$ means in the "real world", at least the formal semantics for such statements is straightforward. Consider the set Ω of $K = 2^k$ truth assignments for the propositions $p_1, \ldots, p_k$. We give semantics to probabilistic statements over this language in terms of a probability distribution μ over the set Ω (see [FHM90] for details). Since each truth assignment $\omega \in \Omega$ determines the truth value of any propositional formula β, we can determine the probability of any such formula:

$$\Pr_\mu(\beta) = \sum_{\omega \models \beta} \mu(\omega).$$

Clearly, we can determine if a probability distribution μ satisfies a set Λ of probabilistic constraints. The standard notion of probabilistic inference would say that $\Lambda \models \Pr(\beta) \in [\lambda_1, \lambda_2]$ if $\Pr_\mu(\beta)$ is within the range $[\lambda_1, \lambda_2]$ for any distribution μ that satisfies the constraints in Λ.

Unfortunately, while this is a very natural definition, the constraints that one can derive from it are typically quite weak. For that reason, Nilsson suggested strengthening this notion of inference by applying the *principle of maximum entropy* [Jay78]: Rather than considering all distributions μ satisfying Λ, we consider only the distribution(s) μ^* that have the greatest entropy among those satisfying the constraints. As we now show, one implication of our results is that the random worlds method provides a principled motivation for this introduction of maximum entropy to probabilistic propositional reasoning. In fact, the connection between probabilistic propositional reasoning and random worlds should be now fairly clear:

- The primitive propositions $p_1, \ldots, p_k$ correspond to unary predicates $P_1, \ldots, P_k$.

- A propositional formula β over $p_1, \ldots, p_k$ corresponds uniquely to an essentially propositional formula ξ_β as follows: we replace each occurrence of the propositional symbol p_i with $P_i(x)$.

- The set Λ of probabilistic constraints corresponds to a knowledge base KB'_Λ — a constant-free knowledge base containing only proportion expressions. The correspondence is as follows:

  - A probability expression of the form $\Pr(\beta)$ appearing in Λ is replaced by the proportion expression $||\xi_\beta(x)||_x$. Similarly, a conditional probability expression $\Pr(\beta|\beta')$ is replaced by $\|\xi_\beta(x)|\xi_{\beta'}(x)\|_x$.
  - Each comparison connective $=$ is replaced by $\approx_i$ for some i, and each $\leq$ with $\preceq_i$. (The particular choices for the approximate equality connectives do not matter in this context.)

  The other elements that can appear in a proportion formula (such as rational numbers and arithmetical connectives) remain unchanged.

- There is a one-to-one correspondence between truth assignments and atoms: the truth assignment ω corresponds to the atom $A = P'_1 \wedge \ldots P'_k$ where P'_i is P_i if $\omega(p_i) = \mathit{true}$ and $\neg P_i$ otherwise. Let $\omega_1, \ldots, \omega_K$ be the truth assignments corresponding to the atoms $A_1, \ldots, A_K$, respectively.

- There is a one-to-one correspondence between probability distributions over the set Ω of truth assignments and points in Δ^K. For each point $\vec{u} \in \Delta^K$, let $\mu_{\vec{u}}$ denote the corresponding probability distribution over Ω, where $\mu_{\vec{u}}(\omega_i) = u_i$.

Remark 7.2.12: Clearly, $\omega_j \models \beta$ iff $A_j \in \mathcal{A}(\xi_\beta)$. Therefore, for any $\vec{u}$, we have

$$F_{[\xi_\beta]}(\vec{u}) = \Pr_{\mu_{\vec{u}}}(\beta). \quad \blacksquare$$

The following result demonstrates the tight connection between probabilistic propositional reasoning using maximum entropy and random worlds.

Theorem 7.2.13: *Let Λ be a conjunction of constraints of the form $\Pr(\beta|\beta') = \lambda$ or $\Pr(\beta|\beta') \in [\lambda_1, \lambda_2]$. There is a unique probability distribution μ^* of maximum entropy satisfying Λ. Moreover, for all β and β', if $\Pr_{\mu^*}(\beta') > 0$, then*

$$\Pr_\infty(\xi_\beta(c)|\xi_{\beta'}(c) \wedge KB'_\Lambda) = \Pr_{\mu^*}(\beta|\beta').$$

Theorem 7.2.13 is an easy corollary of Theorem 7.2.10. Because the constraints in Λ are linear, the space $S^{\vec{0}}[KB'_\Lambda]$ has a unique point $\vec{v}$ of maximum entropy. In fact, it is easy to show that $\mu_{\vec{v}}$ is the (unique) maximum entropy probability distribution (over Ω) satisfying the constraints Λ. And because there are no negated proportion expressions in Λ, the formula $KB = \xi_{\beta'}(c) \wedge KB'_\Lambda$ is easily seen to be essentially positive.

Most applications of probabilistic propositional reasoning consider simple constraints of the form used in the theorem, and so such applications can be viewed as very special cases of the random-words approach. In fact, this theorem is essentially a very old one. The connection between counting "worlds" and the entropy maximum in a space defined as a conjunction of linear constraints is very well-known. It has been extensively formalized in the field of thermodynamics as early as the 19th century work of Maxwell and Gibbs. Recently, this type of reasoning has been applied to problems in an AI context by Paris and Vencovska [PV89] and Shastri [Sha89]. The work of Paris and Venconvska is particularly relevant because they also realize the necessity of adopting a formal notion of "approximation", although the precise details of their approach differ from ours.

To the best of our knowledge, most of the work on probabilistic propositional reasoning and all *formal* presentations of the entropy/worlds connection (in particular, those of [PV89, Sha89]) have limited themselves to conjunctions of linear constraints. Our more general language gives us a great deal of additional expressive power. For example, it is quite reasonable to want the ability to express that properties are (approximately) statistically independent. For example, we may wish to assert that $Bird(x)$ and $Yellow(x)$ are independent properties by saying $||Bird(x) \wedge Yellow(x)||_x \approx ||Bird(x)||_x \cdot ||Yellow(x)||_x$. Clearly, such constraints are not linear. Nevertheless, our Theorem 7.2.10 covers such cases and much more.

A type of probabilistic propositional reasoning has also been applied in the context of giving probabilistic semantics to default reasoning [Pea89]. Here also, the connection to random-worlds is of interest. In particular, it follows from Corollary 7.2.9 that the recent work of Goldszmidt, Morris, and Pearl [GMP90] can be embedded in the random-worlds framework. In the rest of this subsection, we explain their approach and the embedding.

Consider a language consisting of propositional formulas over the propositional variables $p_1, \ldots, p_k$, and default rules of the form $B \rightarrow C$ (read "B's are typically C's"), where B and C are propositional formulas. A distribution μ is said to *ϵ-satisfy* a default rule $B \rightarrow C$ if $\mu(C|B) \geq 1 \Leftrightarrow \epsilon$. In addition to default rules, the framework also permits the use of material implication in a rule, as in $B \Rightarrow C$. A distribution μ is said to satisfy such a rule if $\mu(C|B) = 1$. A *parameterized probability distribution* (PPD) is a collection $\{\mu_\epsilon\}_{\epsilon>0}$ of probability distributions over Ω, parameterized by ϵ. A PPD $\{\mu_\epsilon\}_{\epsilon>0}$ ϵ-satisfies a set $\mathcal{R}$ of rules if for every ϵ, μ_ϵ ϵ-satisfies every default rule $r \in \mathcal{R}$ and satisfies every non-default rule $r \in \mathcal{R}$. A set $\mathcal{R}$ of default rules ϵ-entails $B \rightarrow C$ if for every PPD that ϵ-satisfies $\mathcal{R}$, $\lim_{\epsilon\rightarrow 0} \mu_\epsilon(C|B) = 1$.

As shown in [GP90], ϵ-entailment possesses a number of reasonable properties typically associated with default reasoning, including a preference for more specific information. However, there are a number of desirable properties that it does not have. Among other things, irrelevant information is not ignored (see Section 2.2.3 for an extensive discussion of this issue).

To obtain additional desirable properties, ϵ-semantics is extended in [GMP90] by an application of the principle of maximum entropy. Instead of considering all possible PPD's, as above, we consider only the PPD $\left\{\mu^*_{\epsilon,\mathcal{R}}\right\}_{\epsilon>0}$ such that, for each ϵ, $\mu^*_{\epsilon,\mathcal{R}}$ has the maximum entropy among distributions that ϵ-satisfy all the rules in $\mathcal{R}$ (see [GMP90] for precise definitions and technical details). Note that since the constraints used to define $\mu^*_{\epsilon,\mathcal{R}}$ are all linear, as we mentioned before, there is indeed a unique such point of maximum entropy. A rule $B \rightarrow C$ is an *ME-plausible consequence* of $\mathcal{R}$ if $\lim_{\epsilon\rightarrow 0} \mu^*_{\epsilon,\mathcal{R}}(C|B) = 1$. The notion of ME-plausible consequence is analyzed in detail in [GMP90], where it is shown to inherit all the nice properties of ϵ-entailment (such as the preference for more specific information), while successfully ignoring irrelevant information. Equally importantly, algorithms are provided for computing the ME-plausible consequences of a set of rules in certain cases.

Our maximum entropy results can be used to show that the approach of [GMP90] can be embedded in our framework in a straightforward manner. We simply translate a default rule r of the form $B \rightarrow C$ into a first-order default rule

$$\theta_r =_{\rm def} \|\xi_C(x)|\xi_B(x)\|_x \approx_1 1,$$

as in our earlier translation of Nilsson's approach. Note that the formulas that arise under this translation all use the same approximate equality connective $\approx_1$. The reason is that the approach of [GMP90] uses the same ϵ for all default rules. We can similarly translate a (non-default) rule r of the form $B \Rightarrow C$ into a first-order constraint using universal quantification:

$$\theta_r =_{\rm def} \forall x\, (\xi_B(x) \Rightarrow \xi_C(x)).$$

Under this translation, we can prove the following theorem.

Theorem 7.2.14: *Let c be a constant symbol. Using the translation described above, for any set $\mathcal{R}$ of defeasible rules, $B \rightarrow C$ is an ME-plausible consequence of $\mathcal{R}$ iff*

$$\Pr_\infty\left(\xi_C(c) \,\middle|\, \xi_B(c) \wedge \bigwedge_{r\in\mathcal{R}} \theta_r\right) = 1.$$

In particular, this theorem implies that all the computational techniques and results described in [GMP90] carry over to this special case of the random-worlds method. It also shows that random-world provides a principled justification for the approach [GMP90] present (one which is quite different from the justification given in [GMP90] itself).

7.2.4 Beyond simple queries

In Section 7.2.2 we restricted attention to simple queries. Our main result, Theorem 7.2.10 made yet more assumptions: essential positivity, the existence of a unique maximum entropy point $\vec{v}$, and the requirement that $F_{[\psi]}(\vec{v}) > 0$. We believe that this theorem is widely applicable, as demonstrated by our discussion in Section 7.2.3. However, it allows us to take advantage of only a small fragment of our rich language. In contrast, the concentration result (Theorem 7.1.13) holds with essentially no restrictions. So, it is reasonable to hope that we can extend Theorem 7.2.10 to a more general setting as well. As we show in this section, while this result can be extended substantially, there are serious limitations and subtleties. We illustrate these subtleties by means of examples, and then show how the result can indeed be extended.

We first consider the restrictions we placed on the *KB*, and show the difficulties that arise if we drop them. We start with the restriction to a single maximum entropy point. As the concentration theorem (Theorem 7.1.13) shows, the entropy of almost every world is near maximum. But it does not follow that all the maximum entropy points are surrounded by similar numbers of worlds. Thus, in the presence of more than one maximum entropy point, we face the problem of finding the relative importance, or weighting, of each maximum entropy point. As the following example illustrates, this weighting is often sensitive to the tolerance values. For this reason, non-unique entropy maxima often lead to nonrobustness.

Example 7.2.15: Suppose $\Phi = \{P, c\}$, and consider the knowledge base

$$KB = (\|P(x)\|_x \preceq_1 0.3) \vee (\|P(x)\|_x \succeq_2 0.7).$$

Assume we want to compute $\Pr_\infty(P(c)|KB)$. In this case, $S^{\vec{\tau}}[KB]$ is

$$\{(u_1, u_2) \in \Delta^2 \;:\; u_1 \leq 0.3 + \tau_1 \text{ or } u_1 \geq 0.7 \Leftrightarrow \tau_2\},$$

and $S^{\vec{0}}[KB]$ is

$$\{(u_1, u_2) \in \Delta^2 \;:\; u_1 \leq 0.3 \text{ or } u_1 \geq 0.7\}.$$

Note that $S^{\vec{0}}[KB]$ has two maximum entropy points: $(0.3, 0.7)$ and $(0.7, 0.3)$.

Now consider the maximum entropy points of $S^{\vec{\tau}}[KB]$ for $\vec{\tau} > \vec{0}$. It is not hard to show that if $\tau_1 > \tau_2$, then this space has a unique maximum entropy point, $(0.3 + \tau_1, 0.7 \Leftrightarrow \tau_1)$. In this case, $\Pr^{\vec{\tau}}_\infty(P(c)|KB) = 0.3 + \tau_1$. On the other hand, if $\tau_1 < \tau_2$, then the unique maximum entropy point of this space is $(0.7 + \tau_2, 0.3 \Leftrightarrow \tau_2)$, in which case $\Pr^{\vec{\tau}}_\infty(P(c)|KB) = 0.7 + \tau_2$. If $\tau_1 = \tau_2$, then the space $S^{\vec{\tau}}[KB]$ also has two maximum entropy points, and by symmetry we obtain that $\Pr^{\vec{\tau}}_\infty(P(c)|KB) = 0.5$. Again, by appropriately choosing a sequence of tolerance vectors converging to $\vec{0}$, we can make the asymptotic value of this fraction either 0.3, 0.5, or 0.7. So $\Pr_\infty(P(c)|KB)$ does not exist.

It is not disjunctions *per se* that cause the problem here: if we consider instead the database $KB' = (\|P(x)\|_x \preceq_1 0.3) \vee (\|P(x)\|_x \succeq_2 0.6)$, then there is no difficulty. There is a unique maximum entropy point of $S^{\vec{0}}[KB']$ — $(0.6, 0.4)$ — and the asymptotic probability $\Pr_\infty(P(c)|KB') = 0.6$, as we would want.[3] ▮

[3] We remark that it is also possible to construct examples of multiple maximum entropy points by using quadratic constraints rather than disjunction.

In light of this example (and many similar ones we can construct), we maintain our restriction below to the case where there is a single maximum entropy point. As we argued earlier, we expect this to be the typical case in practice.

We now turn our attention to the requirement that $F_{[\psi]}(\vec{v}) > 0$. As we have already observed, this seems to be an obvious restriction to make, in light of the fact that the function $F_{[\varphi|\psi]}(\vec{v})$ is not defined otherwise. However, this difficulty is actually a manifestation of a much deeper problem. As the following example shows, the entire approach of using the maximum entropy point of $S^{\vec{0}}[KB]$ to compute degrees of belief fails in those cases where $F_{[\psi]}(\vec{v}) = 0$.

Example 7.2.16: Consider the knowledge base KB described in Example 3.1.2, and consider the problem of computing $\Pr_\infty(\mathit{Fly}(\mathit{Tweety})|\mathit{Penguin}(\mathit{Tweety}) \wedge KB)$. We can easily conclude from Theorem 4.2.1 that this degree of belief is 0, as expected. But now, consider the maximum entropy point of $S^{\vec{0}}[KB \wedge \mathit{Penguin}(\mathit{Tweety})]$. The coordinates v_1, corresponding to $\mathit{Fly} \wedge \mathit{Penguin}$, and v_2, corresponding to $\neg\mathit{Fly} \wedge \mathit{Penguin}$, are both 0. Hence, $F_{[\mathit{Penguin}]}(\vec{v}) = 0$, so that Theorem 7.2.10 does not apply. But, as we said, the problem is more fundamental. The information that the proportion of flying penguins is zero is simply not present in the maximum entropy point $\vec{v}$. In particular, we would have obtained precisely the same maximum entropy point from the very different knowledge base KB' that asserts simply that $||\mathit{Penguin}(x)||_x \approx_1 0$. This new knowledge base tells us nothing whatsoever about the fraction of flying penguins. In fact, it is easy to show that $\Pr_\infty(\mathit{Fly}(\mathit{Tweety})|KB') = 0.5$. Thus, we cannot derive the degree of belief in $\mathit{Fly}(\mathit{Tweety})$ from this maximum entropy point; the relevant information is simply not present. ▮

Thus, we clearly cannot apply the philosophy of Theorem 7.2.10 to cases where $F_{[\psi]}(\vec{v}) = 0$. It is natural to ask, however, whether this requirement can be relaxed in the context of a different result. That is, is it possible to construct a technique for computing degrees of belief in those cases where $F_{[\psi]} = 0$? In the context of using maximum entropy as a computational tool, this seems difficult. In particular, it seems to require the type of parametric maximum entropy computation that we discussed in Section 7.2.1. The computational technique of [GMP90] does, in fact, use this type of parametric analysis for a restricted class of problems. Some of our theorems in Chapter 4 on the one hand, can be viewed as providing an alternative method for computing degrees of belief, which also applies to such cases.

An additional assumption made in Section 7.2.2, is that the knowledge base has a special form, namely $\psi(c) \wedge KB'$, where ψ is essentially propositional and KB' does not contain any occurrences of c. Our theorem makes use of a generalization of this restriction.

Definition 7.2.17: A knowledge base KB is said to be *separable with respect to query* φ if it has the form $\psi \wedge KB'$, where ψ contains neither quantifiers nor proportions, and KB' contains none of the constant symbols appearing in φ or in ψ.[4] ▮

It should be clear that if a query $\varphi(c)$ is simple for KB (as assumed in previous subsection), then the separability condition is satisfied.

[4] Clearly, since our approach is semantic, it also suffices if the knowledge base is equivalent to one of this form.

As the following example shows, if we do not assume separablility, we can again easily run into nonrobust behavior:

Example 7.2.18: Consider the following knowledge base KB over the vocabulary $\Phi = \{P, c\}$:

$$(||P(x)||_x \approx_1 0.3 \wedge P(c)) \vee (||P(x)||_x \approx_2 0.3 \wedge \neg P(c)).$$

KB is not separable with respect to the query $P(c)$. The space $S^{\vec{0}}[KB]$ consists of a unique point $(0.3, 0.7)$, which is also the maximum entropy point. Both disjuncts of KB are consistent with the maximum entropy point, so we might expect that the presence of the conjuncts $P(c)$ and $\neg P(c)$ in the disjuncts would not affect the degree of belief. That is, if it were possible to ignore or discount the role of the tolerances, we would expect $\Pr_\infty(P(c)|KB) = 0.3$. However, this is not the case. Consider the behavior of $\Pr_\infty^{\vec{\tau}}(P(c)|KB)$ for $\vec{\tau} > \vec{0}$. If $\tau_1 > \tau_2$, then the maximum entropy point of $S^{\vec{\tau}}[KB]$ is $(0.3+\tau_1, 0.7 \Leftrightarrow \tau_1)$. Now, consider some $\epsilon > 0$ sufficiently small so that $\tau_2+\epsilon < \tau_1$. By Corollary 7.1.14, we deduce that $\Pr_\infty^{\vec{\tau}}((||P(x)||_x > 0.3+\tau_2) \mid KB) = 1$. Therefore, by Theorem 7.1.16, $\Pr_\infty^{\vec{\tau}}(P(c)|KB) = \Pr_\infty^{\vec{\tau}}(P(c) \mid KB \wedge (||P(x)||_x > 0.3 + \tau_2))$ (assuming the limit exists). But since the newly added expression is inconsistent with the second disjunct, we obtain that $\Pr_\infty^{\vec{\tau}}(P(c)|KB) = \Pr_\infty^{\vec{\tau}}(P(c) \mid P(c) \wedge (||P(x)||_x \approx_1 0.3)) = 1$, and not 0.3. On the other hand, if $\tau_1 < \tau_2$, we get the symmetric behavior, where $\Pr_\infty^{\vec{\tau}}(P(c)|KB) = 0$. Only if $\tau_1 = \tau_2$ do we get the expected value of 0.3 for $\Pr_\infty^{\vec{\tau}}(P(c)|KB)$. Clearly, by appropriately choosing a sequence of tolerance vectors converging to $\vec{0}$, we can make the asymptotic value of this fraction any of 0, 0.3, or 1, or not exist at all. Again, $\Pr_\infty(P(c)|KB)$ is not robust. ∎

We now turn our attention to restrictions on the query. In Section 7.2.2, we restricted to queries of the form $\varphi(c)$, where $\varphi(x)$ is essentially propositional. Although we intend to ease this restriction, we do not intend to allow queries that involve statistical information. The following example illustrates the difficulties.

Example 7.2.19: Consider the knowledge base $KB = ||P(x)||_x \approx_1 0.3$ and the query $\varphi = ||P(x)||_x \approx_2 0.3$. It is easy to see that the unique maximum entropy point of $S^{\vec{\tau}}[KB]$ is $(0.3+\tau_1, 0.7 \Leftrightarrow \tau_1)$. First suppose $\tau_2 < \tau_1$. From Corollary 7.1.14, it follows that $\Pr_\infty^{\vec{\tau}}((||P(x)||_x > 0.3 + \tau_2) \mid KB) = 1$. Therefore, by Theorem 7.1.16, $\Pr_\infty^{\vec{\tau}}(\varphi|KB) = \Pr_\infty^{\vec{\tau}}(\varphi|KB \wedge (||P(x)||_x > 0.3 + \tau_2))$ (assuming the limit exists). On the other hand, the latter expression is clearly 0. If $\tau_1 < \tau_2$, then $KB[\vec{\tau}] \models \varphi[\vec{\tau}]$, so that $\Pr_\infty^{\vec{\tau}}(\varphi|KB) = 1$. Thus, the limiting behavior of $\Pr_\infty^{\vec{\tau}}(\varphi|KB)$ depends on how $\vec{\tau}$ goes to $\vec{0}$. Thus, $\Pr_\infty(\varphi|KB)$ is nonrobust. ∎

The real problem here is the semantics of proportion expressions in queries. While the utility of the $\approx$ connective in expressing statistical information in the knowledge base is surely uncontroversial, its role in *conclusions* we might draw, such as φ in Example 7.2.19, is much less clear. The formal semantics we have defined requires that we consider all possible tolerances for a proportion expression in φ, so it is not surprising that nonrobustness is the usual result. One might argue that the tolerances in queries should be allowed to depend more closely on tolerances of expressions in the knowledge base. It is possible to formalize this intuition, as is done in [KH92], to give an alternative semantics for proportion expressions (in φ) that is often more plausible. Considerations of this alternative semantics would lead us too far afield here; rather, we focus for the rest of the section on first-order queries.

In fact, our goal is to allow arbitrary first-order queries, even those that involve predicates of arbitrary arity and equality (although we still need to restrict the knowledge base to the unary language $\mathcal{L}_1^{\approx}$). However, as the following example shows, quantifiers too can cause problems.

Example 7.2.20: Let $\Phi = \{P, c\}$ and consider $KB_1 = \forall x\, \neg P(x)$, $KB_2 = ||P(x)||_x \approx_1 0$, and $\varphi = \exists x\, P(x)$. It is easy to see that $S^{\vec{0}}[KB_1] = S^{\vec{0}}[KB_2] = \{(0,1)\}$, and therefore the unique maximum entropy point in both is $\vec{v} = (0,1)$. However, $\Pr_\infty(\varphi | KB_1)$ is clearly 0, whereas $\Pr_\infty(\varphi | KB_2)$ is actually 1. To see the latter fact, observe that the vast majority of models of KB_2 around $\vec{v}$ actually satisfy $\exists x\, P(x)$. There is actually only a single world associated with $(0,1)$ at which $\exists x\, P(x)$ is false. As in Example 7.2.16, we see that the maximum entropy point of $S^{\vec{0}}[KB]$ does not always suffice to determine degrees of belief. ■

In the case of the knowledge base KB_1, the maximum entropy point $(0,1)$ is quite misleading about the nature of nearby worlds; this is the sort of "discontinuity" we must avoid when finding the degree of belief of a formula involving first-order quantifiers. The notion of stability defined below is intended to deal with this problem. To define it, we need the following simplified variant of a *size description* (defined in Chapter 6.

Definition 7.2.21: A *size description (over $\mathcal{P}$)* is a conjunction of K formulas: for each atom A_j over $\mathcal{P}$, it includes exactly one of $\exists x\, A_j(x)$ and $\neg\exists x\, A_j(x)$. For any $\vec{u} \in \Delta^K$, the *size description associated with* $\vec{u}$, written $\sigma(\vec{u})$, is that size description which includes $\neg\exists x\, A_i(x)$ if $u_i = 0$ and $\exists x\, A_i(x)$ if $u_i > 0$. ■

The problems that we want to avoid occur when there is a maximum entropy point $\vec{v}$ with size description $\sigma(\vec{v})$ such that in any neighborhood of $\vec{v}$, most of the worlds satisfying KB are associated with other size descriptions. Intuitively, the problem with this is that the coordinates of $\vec{v}$ alone give us misleading information about the nature of worlds near $\vec{v}$, and so about degrees of belief.[5] We give a sufficient condition which can be used to avoid this problem in the context of our theorems. This condition is effective and uses machinery (in particular, the ability to find solution spaces) that is needed to use the maximum entropy approach anyway.

Definition 7.2.22: Let $\vec{v}$ be a maximum entropy point of $S^{\vec{\tau}}[KB]$. We say that $\vec{v}$ is *safe* (with respect to KB and $\vec{\tau}$) if $\vec{v}$ is not contained in $S^{\vec{\tau}}[KB \wedge \neg\sigma(\vec{v})]$. We say that *$KB$ and $\vec{\tau}$ are stable for σ^** if for every maximum entropy point $\vec{v} \in S^{\vec{\tau}}[KB]$ we have that $\sigma(\vec{v}) = \sigma^*$ and that $\vec{v}$ is safe with respect to KB and $\vec{\tau}$. ■

[5] We actually conjecture that problems of this sort cannot arise in the context of a maximum entropy point of $S^{\vec{\tau}}[KB]$ for $\vec{\tau} > \vec{0}$. More precisely, for sufficiently small $\vec{\tau}$ and a maximum entropy point $\vec{v}$ of $S^{\vec{\tau}}[KB]$ with $KB \in \mathcal{L}_1^{\approx}$, we conjecture that $\Pr_\infty^{\vec{\tau}}[\mathcal{O}](\sigma(\vec{v})|KB) = 1$ where $\mathcal{O}$ is any open set that contains $\vec{v}$ but no other maximum entropy point of $S^{\vec{\tau}}[KB]$. If this is indeed the case, then the machinery of stability that we are about to introduce is unnecessary, since it holds in all cases that we need it. However, we have been unable to prove this.

The next result is the key property of stability that we need.

Theorem 7.2.23: *If KB and $\vec{\tau} > \vec{0}$ are stable for σ^* then* $\mathrm{Pr}_\infty^{\vec{\tau}}(\sigma^* | KB) = 1$.

Our theorems will use the assumption that there exists some σ^* such that, for all sufficiently small $\vec{\tau}$, KB and $\vec{\tau}$ are stable for σ^*. We note that this does not imply that σ^* is necessarily the size description associated with the maximum entropy point(s) of $S^{\vec{0}}[KB]$.

Example 7.2.24: Consider the knowledge base KB_2 in Example 7.2.20, and recall that $\vec{v} = (0, 1)$ is the maximum entropy point of $S^{\vec{0}}[KB_2]$. The size description $\sigma(\vec{v})$ is $\neg\exists x\, A_1(x) \wedge \exists x\, A_2(x)$. However the maximum entropy point of $S^{\vec{\tau}}[KB_2]$ for any $\vec{\tau} > 0$ is actually $(\tau_1, 1{\Leftrightarrow}\tau_1)$, so that the appropriate σ^* for any such $\vec{\tau}$ is $\exists x\, A_1(x) \wedge \exists x\, A_2(x)$. ▮

As we now show, the restrictions outlined above and in Section 7.2.1 suffice for our next result on computing degrees of belief. In order to state this result, we need one additional concept. Recall that in Section 7.2.2 we expressed an essentially propositional formula $\varphi(x)$ as a disjunction of atoms. Since we wish to also consider φ using more than one constant and non-unary predicates, we need a richer concept than atoms. As it turns out, the right notion is a slight generalization of the notion of *complete description* (see Definition 6.2.2).

Definition 7.2.25: Let $\mathcal{Z}$ be some set of variables and constants. A *complete description D* over Φ and $\mathcal{Z}$ is an unquantified conjunction of formulas such that:

- For every predicate $R \in \Phi \cup \{=\}$ of arity r, and for every $z_{i_1}, \ldots, z_{i_r} \in \mathcal{Z}$, D contains exactly one of $R(z_{i_1}, \ldots, z_{i_r})$ or $\neg R(z_{i_1}, \ldots, z_{i_r})$ as a conjunct.
- D is consistent.[6] ▮

In this context, complete descriptions serve to extend the role of atoms in the setting of essentially propositional formulas to the more general setting. As in the case of atoms, if we fix some arbitrary ordering of the conjuncts in a complete description, then complete descriptions are mutually exclusive and exhaustive. Clearly any formula ξ whose free variables and constants are contained in $\mathcal{Z}$, and which is is quantifier- and proportion-free, is equivalent to some disjunction of complete descriptions over $\mathcal{Z}$. For a formula ξ, let $\mathcal{A}(\xi)$ be a set of complete descriptions over $\mathcal{Z}$ such that ξ is equivalent to the disjunction $\bigvee_{D \in \mathcal{A}(\xi)} D$, where $\mathcal{Z}$ is the set of constants and free variables in ξ.

For the purposes of the remaining discussion (except within proofs), we are interested only in complete descriptions over an empty set of variables. For a set of constants $\mathcal{Z}$, we can view a description D over $\mathcal{Z}$ as describing the different properties of the constants in $\mathcal{Z}$. In our construction, we will define the set $\mathcal{Z}$ to contain precisely those constants in φ and in ψ. In particular, if KB is separable with respect to φ, then KB' will contain no constant in $\mathcal{Z}$.

[6] Inconsistency is possible because of the use of equality. For example, if D includes $z_1 = z_2$ as well as both $R(z_1, z_3)$ and $\neg R(z_2, z_3)$, it is inconsistent.

A complete description D over a set of constants $\mathcal{Z}$ can be decomposed into three parts: the *unary part* D^1 which consists of those conjuncts of D that involve unary predicates (and thus determines an atom for each of the constant symbols), the *equality part* $D^=$ which consists of those conjuncts of D involving equality (and thus determines which of the constants are equal to each other), and the *non-unary part* $D^{>1}$ which consists of those conjuncts of D involving non-unary predicates (and thus determines the non-unary properties other than equality of the constants). As we suggested, the unary part of such a complete description D extends the concept of an atom to the case of multiple constants. For this purpose, we also extend $F_{[A]}$ (for an atom A) and define $F_{[D]}$ for a description D. Intuitively, we are treating each of the individuals as independent, so that the probability that constant c_1 satisfies atom A_{j_1} and that constant c_2 satisfies A_{j_2} is just the product of the probability that c_1 satisfies A_{j_1} and the probability that c_2 satisfies A_{j_2}.

Definition 7.2.26: For a complete description D, without variables, whose unary part is equivalent to $A_{j_1}(c_1) \wedge \ldots \wedge A_{j_m}(c_m)$, and for a point $\vec{u} \in \Delta^K$, we define

$$F_{[D]}(\vec{u}) = \prod_{\ell=1}^{m} u_{j_\ell}. \quad \blacksquare$$

Note that $F_{[D]}$ is depends only on D^1, the unary part of D.

As we mentioned, we can extend our approach to deal with formulas φ that also use non-unary predicate symbols. Our computational procedure for such formulas uses the maximum-entropy approach described above combined with the techniques of Chapter 6. Recall that these latter were used in Chapter 6 to compute asymptotic conditional probabilities when conditioning on a first-order knowledge base KB_{fo}. The basic idea in that case is as follows: To compute $\Pr_\infty(\varphi|KB_{fo})$, we examine the behavior of φ in finite models of KB_{fo}. We partition the models of KB_{fo} into a finite collection of classes, such that φ behaves *uniformly* in any individual class. By this we mean that almost all worlds in the class satisfy φ or almost none do; i.e., there is a 0-1 law for the asymptotic probability of φ when we restrict attention to models in a single class. In order to compute $\Pr_\infty(\varphi|KB_{fo})$ we therefore identify the classes, compute the relative weight of each class (which is required because the classes are not necessarily of equal relative size), and then decide for each class, whether the asymptotic probability of φ is zero or one.

It turns out that much the same ideas continue to work in this framework. In this case, the classes are defined using complete descriptions and the appropriate size description σ^*. The main difference is that, rather than examining all worlds consistent with the knowledge base, we now concentrate on those worlds in the vicinity of the maximum entropy points, as outlined in the previous section. It turns out that the restriction to these worlds affects very few aspects of this computational procedure. In fact, the only difference is in computing the relative weight of the different classes. This last step can be done using maximum entropy, using the tools described in Section 7.2.2.

Theorem 7.2.27: *Let φ be a formula in $\mathcal{L}^{\approx}$, and let $KB = KB' \wedge \psi$ be an essentially positive knowledge base in $\mathcal{L}_1^{\approx}$ which is separable with respect to φ. Let $\mathcal{Z}$ be the set of constants* *appearing in φ or in ψ (so that KB' contains none of the constants in $\mathcal{Z}$), and let $\chi^{\neq}$ be the formula $\bigwedge_{c,c' \in \mathcal{Z}} c \neq c'$. Assume that there exists a size description σ^* such that, for all $\vec{\tau} > 0$, KB and $\vec{\tau}$ are stable for σ^*, and that the space $S^{\vec{0}}[KB]$ has a unique maximum entropy point $\vec{v}$, then:*

$$\Pr_\infty(\varphi|KB) = \frac{\sum_{D \in \mathcal{A}(\psi \wedge \chi^{\neq})} \Pr_\infty(\varphi|\sigma^* \wedge D) F_{[D]}(\vec{v})}{\sum_{D \in \mathcal{A}(\psi \wedge \chi^{\neq})} F_{[D]}(\vec{v})},$$

if the denominator is positive.

Both φ and $\sigma^* \wedge D$ are first-order formulas, and $\sigma^* \wedge D$ is precisely of the required form by Procedure `Compute01`of Figure 6.2. Thus, we can use this algorithm to compute this limit, in the time bounds outlined in Theorem 6.4.1 and Corollary 6.4.2.

The above theorem shows that the formula $\chi^{\neq}$ defined in its statements holds with probability 1 given any knowledge base KB of the form we are interested in. This corresponds to a default assumption of *unique names*, a property often considered to be desirable in inductive reasoning systems (see Section 4.5).

While this theorem does represent a significant generalization of Theorem 7.2.10, it still has numerous restrictions. There is no question that some of these can be loosened to some extent, although we have not been able to find a clean set of conditions significantly more general than the ones that we have stated. We leave it as an open problem whether such a set of conditions exists. Perhaps the most significant restriction we have made is that of allowing only unary predicates in the KB. This issue is the subject of the next section.

7.3 Beyond unary predicates

The random-worlds method makes complete sense for the full language $\mathcal{L}^{\approx}$ (and, indeed, for even richer languages). On the other hand, our application of maximum entropy is limited to unary knowledge bases. Is this restriction essential? While we do not have a theorem to this effect (indeed, it is not even clear what the wording of such a theorem would be), we conjecture that it is.

Certainly none of the techniques we have used in this chapter can be generalized significantly. One difficulty is that, once we have a binary or higher arity predicate, we see no analogue to the notion of atoms and no canonical form theorem. In Section 7.1.1 and in the proof of Theorem 7.1.5, we discuss why it becomes impossible to get rid of nested quantifiers and proportions when we have non-unary predicates. Even considering matters on a more intuitive level, the problems seem formidable. In a unary language, atoms are useful because they are simple descriptions that summarize everything that might be known about a domain element in a model. But consider a language with a single binary predicate $R(x, y)$. Worlds over this language include all finite graphs (where we think of $R(x, y)$ as holding if there is an edge from x to y). In this language, there are infinitely many properties that may be true or false about a domain element. For example, the assertions "the node x has m neighbors" are expressible in the language for each m. Thus, in order to partition the domain elements according to the properties they satisfy, we would need to define infinitely many partitions. Furthermore, it can be shown that "typically" (i.e., in almost all graphs of sufficiently great size) each node satisfies a different set of first-order properties. Thus, in most graphs, the nodes are all "different" from each other, so that a partition of domain elements into a finite number of "atoms" makes little sense. It is very hard to see how the basic proof strategy we have used, of summarizing a model by listing the number of elements with various properties, can possibly be useful here.

The difficulty of finding an analogue to entropy in the presence of higher-arity predicates doing is supported by our results from Chapter 5. In this chapter we have shown that maximum entropy can be a useful tool for computing degrees of belief in certain cases, if the KB involves only unary predicates. In Chapter 5 we show that there can be *no* general computational technique to compute degrees of belief once we have non-unary predicate symbols in the KB. The problem of finding degrees of belief in this case is highly undecidable. This result was proven without statistical assertions in the language, and in fact holds for quite weak sublanguages of first-order logic. (For instance, in a language without equality and with only depth-two quantifier nesting.) So even if there is some generalized version of maximum entropy, it will either be extremely restricted in application or will be useless as a computational tool.

We believe the question of how widely maximum entropy applies is quite important. Maximum entropy has been gaining prominence as a means of dealing with uncertainty both in AI and other areas. However, the difficulties of using the method once we move to non-unary predicates seem not to have been fully appreciated. In retrospect, this is not that hard to explain; in almost all applications where maximum entropy has been used (and where its application can be best justified in terms of the random-worlds method) the knowledge base is described in terms of unary predicates (or, equivalently, unary functions with a finite range). For example, in physics applications we are interested in such predicates as quantum state (see [DD85]). Similarly, AI applications and expert systems typically use only unary predicates such as symptoms and diseases [Che83]. We suspect that this is not an accident, and that deep problems will arise in more general cases. This poses a challenge to proponents of maximum entropy since, even if one accepts the maximum entropy principle, the discussion above suggests that it may simply be inapplicable in a large class of interesting examples.

Chapter 8

Conclusion

8.1 Problems: Real, Imaginary, and Complex

The principle of indifference and maximum entropy have both been subject to criticism. Any such criticism is, at least potentially, relevant to random worlds. Hence, it is important that we examine the difficulties that people have found. In this section, we consider problems relating to *causal reasoning, language dependence*, *acceptance, learning,* and *computation.*

8.1.1 Causal and temporal information

The random-worlds method can use knowledge bases which include statistical, first-order, and default information. When is this language sufficient? We suspect that it is, in fact, adequate for most traditional knowledge representation tasks. Nevertheless, the question of adequacy can be subtle. This is certainly the case for the important domain of reasoning about actions, using causal and temporal information. In principle, there would seem to be no difficulty choosing a suitable first-order vocabulary, that includes the ability to talk about time explicitly. In the semantics appropriate to many such languages, a world might model an entire temporal sequence of events. However, finding a representation with sufficient expressivity is only part of the problem: we need to know whether the degrees of belief we derive will correctly reflect our intuitions about causal reasoning. It turns out that random worlds gives unintuitive results when used with the most straightforward representations of temporal knowledge.

This observation is not really a new one. As we have observed, the random-worlds method is closely related to maximum entropy (in the context of a unary knowledge base). One of the major criticisms against maximum entropy techniques has been that they seem to have difficulty dealing with causal information [Hun89, Pea88]. Hence, it is not surprising that, if we represent causal and temporal information naively, then the random-worlds method also gives peculiar answers. On the other hand, Hunter [Hun89] has shown that maximum entropy methods can deal with causal information, provided it is represented appropriately. We have recently shown that by using an appropriate representation (related to Hunter's but quite different), the random-worlds method can also deal well with causal information [BGHK93a]. Indeed, our reprsentation allows us to (a) deal with prediction and explanation problems, (b) represent causal information of the type implicit in Bayesian causal nets [Pea88], and (c) provide a clean and concise solution to the frame problem in situation calculus [MH69]. In particular, our proposal deals well with the paradigm problems in the area, for example the Yale Shooting Problem [HM87].

The details of the proposal are beyond the scope of this thesis. However, we want to emphasize here the fact that there may be more than one reasonable way to represent our knowledge of a given domain. When one formulation does not work as we expect, we can look for other ways of representing the problem. It will often turn out that the new representation captures some subtle aspects of the domain, that were ignored by the naive representation. (We believe that this is the case with our alternative formulation of reasoning about actions.) We return to this issue a number of times below.

8.1.2 Representation dependence

As we saw above, random worlds suffers from a problem of representation dependence: causal information is treated correctly only if it is represented appropriately. This shows that choosing the "right" representation of our knowledge is important in the context of the random-worlds approach.

In some ways, this representation dependence is a serious problem because, in practice, how can we know whether we have chosen a good representation or not? Before addressing this, we note that the situation with random worlds is actually not as bad as it might be. As we pointed out in Section 4.1, the random-worlds approach is not sensitive to merely syntactic changes in the knowledge base: logically equivalent knowledge bases always result in the same degrees of belief. So if a changed representation gives different answers, it can only be because we have changed the semantics: we might be using a different ontology, or the new representation might model the world with a different level of detail and accuracy. The representation dependence exhibited by random worlds concerns more than mere syntax. This gives us some hope that the phenomenon can be understood and, at least in some cases, be seen to be entirely appropriate.

Unfortunately, it does seem as if random worlds really is too sensitive; minor and seemingly irrelevant changes can affect things. Perhaps the most disturbing examples concern *language dependence*, or sensitivity to definitional changes. For instance, suppose the only predicate in our language is *White*, and we take *KB* to be *true*, then $\Pr_\infty(\mathit{White}(c)|\mathit{KB}) = 1/2$. On the other hand, if we refine $\neg\mathit{White}$ by adding *Red* and *Blue* to our language and having KB' assert that $\neg\mathit{White}$ is their disjoint union, then $\Pr_\infty(\mathit{White}(c)|\mathit{KB}') = 1/3$. The fact that simply expanding the language and giving a definition of an old notion ($\neg\mathit{White}$) in terms of the new notions (*Red* and *Blue*) can affect the degree of belief seems to be a serious problem. There are several approaches to dealing with this issue.

The first is to declare that representation dependence is justified, i.e., that the choice of an appropriate vocabulary is indeed a significant one, which does encode some of the information at our disposal. In our example above, we can view the choice of vocabulary as reflecting the bias of the reasoner with respect to the partition of the world into colors. Researchers in machine learning and the philosophy of induction have long realized that bias is an inevitable component of effective inductive reasoning. So we should not be completely surprised if it turns that the related area, of finding degrees of belief, should also depend on the bias. Of course, if this is the case we would hope to have a good intuitive understanding of how the degrees of belief depend on the bias. In particular, we would like to give the knowledge base designer some guidelines to selecting the "appropriate" representation. This is an important and seemingly difficult problem in the context of random worlds.

A very different response to the problem of representation dependence is to search for a method of computing degrees of belief that does not suffer from it. To do this, it is important to have a formal definition of representation independence. Once we have such a definition, we can investigate whether there are nontrivial approaches to generating degrees of belief that are representation independent. It can be shown (under a few very weak assumptions) that *any* approach that gives point-valued degrees of belief that act like probabilities cannot be representation independent. This result suggests that we might generalize our concept of "degrees of belief". In fact, there are other reasons to consider doing this as well. In particular, there has been considerable debate about whether the extreme precision forced by point-valued probabilities is reasonable. One frequent suggestion to avoid this involves looking at intervals in $[0,1]$ rather than points. We suspect that interval-valued degrees of belief, if defined appropriately, might in fact be representation independent in many more circumstances than, say, random worlds. We are currently investigating this possibility.

A third response to the problem is to prove representation independence with respect to a large class of queries. To understand this approach, consider another example. Suppose that we know that only about half of birds can fly, Tweety is a bird, and Opus is some other individual (who may or may not be a bird). One obvious way to represent this information is to have a language with predicates *Bird* and *Fly*, and take the *KB* to consist of the statements $\|Fly(x)|Bird(x)\|_x \approx 0.5$ and $Bird(Tweety)$. It is easy to see that $\Pr_\infty(Fly(Tweety)|Bird) = 0.5$ and $\Pr_\infty(Bird(Opus)|KB) = 0.5$. But suppose that we had chosen to use a different language, that uses the basic predicates *Bird* and *FlyingBird*. We would then take KB' to consist of the statements $\|FlyingBird(x)|Bird(x)\|_x \approx 0.5$, $Bird(Tweety)$, and $\forall x(FlyingBird(x) \Rightarrow Bird(x))$. We now get $\Pr_\infty(FlyingBird(Tweety)|KB') = 0.5$ and $\Pr_\infty(Bird(Opus)|KB') = 2/3$. Note that our degree of belief that Tweety flies is 0.5 in both cases. In fact, we can prove that this degree of belief will not be affected by reasonable representational changes. On the other hand, our degree of belief that Opus is a bird differs in the two representations. Arguably, our degree of belief that Opus is a bird *should* be language dependent, since our knowledge base does not contain sufficient information to assign it a "justified" value. This suggests that it would be useful to characterize those queries that are language independent, while recognizing that not all queries will be.

8.1.3 Acceptance and learning

The most fundamental assumption in this thesis is that we are given a knowledge base KB, and wish to calculate degrees of belief relative this knowledge. We have not considered how one comes to know KB in the first place. That is, when do we *accept* information as knowledge? We do not have a good answer to this question. This is unfortunate, since it seems plausible that the processes of gaining knowledge and computing degrees of belief should be interrelated. In particular, Kyburg [Kyb88] has argued that perhaps we might accept assertions that are believed sufficiently strongly. For example, suppose we observe a block b that appears to be white. But it could be that we are is not entirely sure that the block is indeed white; it might be some other light color. Nevertheless, if our confidence in $White(b)$ exceeds some threshold, we might accept it (and so include it in KB).

The problem of acceptance in such examples, concerned with what we learn directly from the senses, is well-known in philosophy [Jef68]. But the problem of acceptance we face is even more difficult than usual, because of our statistical language. Under what circumstances is a statement such as $\|Fly(x)|Bird(x)\|_x \approx 0.9$ accepted as knowledge? Although we regard this as an objective statement about the world, it is unrealistic to suppose that anyone could examine all the birds in the world and count how many of them fly. In practice, it seems that this statistical statement would appear in KB if someone inspects a (presumably large) sample of birds and about 90% of the birds in this sample fly. Then a leap is made: the sample is assumed to be typical, and we then conclude that 90% of all birds fly. This would be in the spirit of Kyburg's suggestion so long as we believe that, with high confidence, the full population has statistics similar to those of the sample.

Unfortunately, the random-worlds method by itself does not support this leap, at least not if we represent the sampling in the most obvious way. That is, suppose we represent our sample using a predicate S. We could then represent the fact that 90% of a sample of birds fly as $\|Fly(x)|Bird(x) \land S(x)\|_x \approx 0.9$. If the KB consists of this fact and $Bird(Tweety)$, we might hope that $\Pr_\infty(Fly(Tweety)|KB) = .9$, but it is not. In fact, random worlds treats the birds in S and those outside S as two unrelated populaions; it maintains the default degree of belief (1/2) that a bird not in S will fly. A related observation, that random worlds cannot do learning (although in a somewhat different sense), was made by Carnap [Car52], who apparently lost a lot of interest in (his version of) random worlds for precisely this reason.

Of course, the failure of the obvious approach does not imply that random worlds is incapable of learning statistics. As was the case for causal reasoning, the solution may be to find an appropriate representation. Perhaps we need a representation reflecting the fact that different individuals do not acquire their properties completely independently of each other. If we see that an animal is tall, it may tell us something about its genetic structure and so, by this mechanism, hint at properties of other animals. But clearly this issue is subtle. If we see a giraffe, this tells us much less about the height of animals in general than it does about other giraffes, and a good representation should reflect this.

Another related approach for dealing with the learning problem is to use a variant of random worlds presented in [BGHK92] called the *random-propensities* approach. This approach is based on the observation that random worlds has an extremely strong bias towards believing that exactly half the domain has any given property, and this is not always reasonable. Why should it be more likely that half of all birds fly than that a third of them do? Roughly speaking, the random-propensities approach postulates the existence of a parameter denoting the "propensity" of a bird to fly. This parameter is shared by the entire population of birds. Initially, all propensities are equally likely. Intuitively, this joint parameter forms the basis for learning: observing a flying bird gives us information about the propensity of birds to fly, and hence about the flying ability of other birds. As shown in [BGHK92], the random propensities method does, indeed, learn from samples. Unfortunately, random propensities has its own problems. In particular, it learns "too often", i.e., even from arbitrary subsets that are not representative samples. Given the assertion "All giraffes are tall", random propensities would conclude that almost everything is tall.

While we still hope to find ways of doing sampling within random worlds, we can also look for other ways of coping with the problem of learning. One idea is to include statements about degrees of belief in the knowledge base. Thus, if 20% of animals in a sample are tall, and we believe that the sample is random, then we might add a statement such as $\Pr(||\mathit{Tall}(x)||_x \approx_1 0.2) \geq 0.9$ to the KB. Although this does not "automate" the sampling procedure, it allows us to assert that we a sample is likely to be representative, without committing absolutely to this fact. In particular, this representation allows further evidence to convince the agent that a sample is, in fact, biased. Adding degrees of belief would also let us deal with the problem of acceptance, mentioned at the beginning of this subsection. If we believe that block b is white, but are not certain, we could write $\Pr(\mathit{White}(b)) \geq 0.9$. We then do not have to fix an (arbitrary?) threshold for acceptance.

However, adding degree of belief statements to a knowledge base is a nontrivial step. Up to now, all the assertions we allowed in a knowledge base were either true or false in a given world. This is not the case for a degree of belief statement. Indeed, our semantics for degrees of belief involve looking at sets of possible worlds. Thus, in order to handle such a statement appropriately, we would need to ensure that our probability distribution over the possible worlds satisfies the associated constraint. While we have some ideas on how this could be done, it is still an open question whether we can handle such statements in a reasonable way.

8.1.4 Computational issues

Our goal in this research has been to understand some of the fundamental issues involved in first-order probabilistic and default reasoning. Until such issues are understood, it is perhaps reasonable to ignore or downplay concerns about computation. If an ideal normative theory turns out to be impractical for computational reasons, we can still use it as guidance in a search for approximations and heuristics.

As we have shown in Chapter 5, computing degrees of belief according to random worlds is, indeed, intractable in general. This is not surprising: our language extends first-order logic, for which validity is undecidable.[1] Although unfortunate, we do not view this as an insurmountable

[1] Although, in fact, finding degrees of belief using random worlds is even *more* intractable than the problem

problem. Note that, in spite of its undecidability, first-order logic is nevertheless viewed as a powerful and useful tool. We believe that the situation with random worlds is analogous. Random worlds is not just a computational tool; it is inherently interesting because of what it can tell us about probabilistic reasoning,

But even in terms of computation, the situation with random worlds is not as bleak as it might seem. We have presented one class of tractable knowledge bases: those using only unary predicates and constants. We showed in Chapter 7 that, in this case, we can often use maximum entropy as a computational tool in deriving degrees of belief. While computing maximum entropy is also hard in general, there are many heuristic techniques that work efficiently in practical cases (see [Gol87] and the references therein). As we have already claimed, this class of problems is an important one. In general, many properties of interest can be expressed using unary predicates, since they express properties of individuals. For example, in physics applications we are interested in such predicates as quantum state (see [DD85]). Similarly, AI applications and expert systems typically use only unary predicates ([Che83]) such as symptoms and diseases. In fact, a good case can be made that statisticians tend to reformulate all problems in terms of unary predicates, since an event in a sample space can be identified with a unary predicate [Sha]. Indeed, most cases where statistics are used, we have a basic unit in mind (an individual, a family, a household, etc.), and the properties (predicates) we consider are typically relative to a single unit (i.e., unary predicates). Thus, results concerning computing degrees of belief for unary knowledge bases are quite significant in practice.

Even for non-unary knowledge bases, there is hope. The intractability proofs given in Chapter 5 use knowledge bases that force the possible worlds to mimic a Turing machine computation. In real life, typical knowledge bases do not usually encode Turing machines! There may therefore be many cases in which computation is practical. In particular, specific domains typically impose additional structure, which may simplify computation. This seems to be the case, for instance, in certain problems that involve reasoning about action.

Furthermore, as we have seen, we *can* compute degrees of belief in many interesting cases. In particular, we have presented a number of theorems that tell us what the degrees of belief are for certain important classes of knowledge bases and queries. Most of these theorems hold for our language in its full generality, including non-unary predicates. We believe that many more such results could be found. Particularly interesting would be more "irrelevance" results that tell us when large parts of the knowledge base can be ignored. Such results could then be used to reduce apparently complex problems to simpler forms, to which other techniques apply. We have already seen that combining different results can often let us compute degrees of belief in cases where no single result suffices.

8.2 Summary

The random-worlds approach for probabilistic reasoning is derived from two very intuitive ideas: possible worlds and the principle of indifference. In spite of its simple semantics, it has many

of deciding validity in first-order logic.

attractive features:

- It can deal with very rich knowledge bases, that involve quantitative information in the form of statistics, qualitative information in the form of defaults, and first-order information. We have had trouble finding realistic problems for which this language is too weak; even fairly esoteric demands such as nested defaults are dealt with naturally.

- Random worlds uses a simple and well-motivated statistical interpretation for defaults. The corresponding semantics allow us to examine the reasonableness of a default with respect to our entire knowledge base, including other default rules.

- It possesses many desirable properties, like preference for more specific information, the ability to ignore irrelevant information, a default assumption of unique names, the ability to combine different pieces of evidence, and more. These are not the result of *ad hoc* assumptions but instead arise naturally from the semantics, as derived theorems.

- It avoids many of the problems that have plagued systems of reference-class reasoning (such as the disjunctive reference class problem) and many of the problems that have plagued systems of non-monotonic reasoning (such as exceptional-subclass inheritance or the lottery paradox). Many systems have been forced to work hard to avoid problems which, in fact, never arose for us at all.

- The random-worlds approach subsumes several important reasoning systems, and generalizes them to the case of first-order logic. In particular, it encompasses deductive reasoning, probabilistic reasoning, certain theories of nonmonotonic inference, the principle of maximum entropy, some rules of evidence combination, and more. But it is far more powerful than any single one of these.

As we saw in Section 8.1, there are certainly some problems with the random-worlds method. We believe that these problems are far from insuperable. But, even conceding these problems for the moment, the substantial success of random-worlds supports a few general conclusions.

One conclusion concerns the role of statistics and degrees of belief. The difference between these, and the problem of relating the two, is at the heart of our work. People have long realized that degrees of belief provide a powerful model for understanding rational behavior (for instance, through decision theory). The random-worlds approach shows that it is possible to assign degrees of belief, using a principled technique, in almost any circumstance. The ideal situation, in which we have complete statistical knowledge concerning a domain, is, of course, dealt with appropriately by random worlds. But more realistically, even a few statistics (which need not even be precise) are still utilized by random worlds to give useful answers. Likewise, completely non-numeric data, which may include defaults and/or a rich first-order theory of some application domain, can be used. Probabilistic reasoning need not make unrealistic demands of the user's knowledge base. Indeed, in a sense it makes less demands that any other reasoning paradigm we know of.

This leads to our next, more general conclusion, which is that many seemingly disparate forms of representation and reasoning *can* (and, we believe, *should*) be unified. The first two points listed above suggest that we can take a large step towards this goal by simply finding a powerful language (with clear semantics) that subsumes specialized representations. The advantages we have found (such as a clear and general way of using nested defaults, or combining defaults and statistics) apply even if one rejects the random-worlds reasoning method itself. But the language is only part of the answer. Can diverse types of reasoning really be seen as aspects of a single more general system? Clearly this is not always possible; for instance, there are surely some interpretations of "defaults" which have no interesting connection to statistics whatsoever. However, we think that our work demonstrates that the alleged gap between probabilistic reasoning and default reasoning is much narrower than previously thought. In fact, the success of random worlds encourages us to hope that a synthesis between different knowledge representation paradigms is possible in most of the interesting domains.

Appendix A

Proofs for Chapter 4

Theorem 4.1.5: *Assume that $KB \mathrel{|\!\!\sim}_{rw} \varphi$ and $KB \mathrel{|\!\!\not\sim}_{rw} \neg\theta$. Then $KB \wedge \theta \mathrel{|\!\!\sim}_{rw} \varphi$ provided that $\Pr_\infty(\varphi|KB \wedge \theta)$ exists. Moreover, a sufficient condition for $\Pr_\infty(\varphi|KB \wedge \theta)$ to exist is that $\Pr_\infty(\theta|KB)$ exists.*

Proof: Since $KB \mathrel{|\!\!\not\sim}_{rw} \neg\theta$, it is not the case that $\Pr_\infty(\neg\theta|KB) = 1$, so $\Pr_\infty(\theta|KB) \neq 0$. Therefore, there exists some $\epsilon > 0$ for which we can construct a sequence of pairs $N^i, \vec{\tau}^i$ as follows: N^i is an increasing sequence of domain sizes, $\vec{\tau}^i$ is a decreasing sequence of tolerance vectors, and $\Pr^{\vec{\tau}^i}_{N^i}(\theta|KB) > \epsilon$. For these pairs $N^i, \vec{\tau}^i$ we can conclude that

$$\Pr^{\vec{\tau}^i}_{N^i}(\neg\varphi|KB \wedge \theta) = \frac{\Pr^{\vec{\tau}^i}_{N^i}(\neg\varphi \wedge \theta|KB)}{\Pr^{\vec{\tau}^i}_{N^i}(\theta|KB)} \leq \frac{\Pr^{\vec{\tau}^i}_{N^i}(\neg\varphi|KB)}{\Pr^{\vec{\tau}^i}_{N^i}(\theta|KB)}.$$

Since $\Pr_\infty(\neg\varphi|KB) = 0$, it is also the case that $\lim_{i\to\infty} \Pr^{\vec{\tau}^i}_{N^i}(\neg\varphi|KB) = 0$. Moreover, we know that for all i, $\Pr^{\vec{\tau}^i}_{N^i}(\theta|KB) > \epsilon > 0$. We can therefore take the limit as $i \to \infty$, and conclude that $\lim_{i\to\infty} \Pr^{\vec{\tau}^i}_{N^i}(\neg\varphi|KB \wedge \theta) = 0$. Thus, if $\Pr_\infty(\varphi|KB \wedge \theta)$ exists, it must be 1.

For the second half of the theorem, suppose $\Pr_\infty(\theta|KB)$ exists. Since $KB \mathrel{|\!\!\not\sim}_{rw} \theta$, we must have that $\Pr_\infty(\theta|KB) = p > 0$. Therefore, for all $\vec{\tau}$ sufficiently small and all N sufficiently large (where "sufficiently large" may depend on $\vec{\tau}$), we can assume that $\Pr^{\vec{\tau}}_N(\theta|KB) > \epsilon > 0$. But now, for any such pair $N, \vec{\tau}$ we can again prove that

$$\Pr^{\vec{\tau}}_N(\neg\varphi|KB \wedge \theta) \leq \frac{\Pr^{\vec{\tau}}_N(\neg\varphi|KB)}{\Pr^{\vec{\tau}}_N(\theta|KB)}.$$

Taking the limit, we obtain that $\Pr_\infty(\neg\varphi|KB \wedge \theta) = 0$, as desired. ∎

Theorem 4.2.11: *Let c be a constant and let KB be a knowledge base satisfying the following conditions:*

(a) $KB \models \psi_0(c)$,

(b) *for any expression of the form* $\|\varphi(x)|\psi(x)\|_x$ *in* KB*, it is the case that either* $KB \models \psi_0 \Rightarrow \psi$ *or that* $KB \models \psi_0 \Rightarrow \neg\psi$*,*

(c) *the (predicate and constant) symbols in* $\varphi(x)$ *appear in* KB *only on the left-hand side of the conditional in the proportion expressions described in condition (b),*

(d) *the constant* c *does not appear in the formula* $\varphi(x)$*.*

Assume that for all sufficiently small tolerance vectors $\vec{\tau}$*:*

$$KB[\vec{\tau}] \models \|\varphi(x)|\psi_0(x)\|_x \in [\alpha, \beta].$$

Then $\Pr_\infty(\varphi(c)|KB) \in [\alpha, \beta]$*.*

Proof: This theorem is proved with the same general strategy we used for Theorem 4.2.1. That is, for each domain size N and tolerance vector $\vec{\tau}$, we partition the worlds of size N satisfying $KB[\vec{\tau}]$ into clusters and prove that, within each cluster, the probability of $\varphi(c)$ is in the interval $[\alpha, \beta]$. As before, this suffices to prove the result. However the clusters are defined quite differently in this theorem.

We define the clusters as maximal sets satisfying the following three conditions:

1. All worlds in a cluster must agree on the denotation of every vocabulary symbol except, possibly, for those appearing in $\varphi(x)$. Note that, in particular, they agree on the denotation of the constant c. They must also agree as to which elements satisfy $\psi_0(x)$; let this set be A_0.

2. The denotation of symbols in φ must also be constant, except, possibly, when a member of A_0 is involved. More precisely, let $\overline{A_0}$ be the set of domain elements $\{1, \ldots, N\} \Leftrightarrow A_0$. Then for any predicate symbol R of arity r appearing in $\varphi(x)$, and for all worlds W' in the cluster, if $d_1, \ldots, d_r \in \overline{A_0}$ then $R(d_1, \ldots, d_r)$ holds in W' iff it holds in W, Similarly, for any constant symbol c' appearing in $\varphi(x)$, if it denotes $d' \in \overline{A_0}$ in W, then it must denote d' in W'.

3. All worlds in the cluster are isomorphic with respect to the vocabulary symbols in φ. More precisely, if W and W' are two worlds in the cluster, then there exists some permutation π of the domain such that for any predicate symbol R as above, and any domain elements $d_1, \ldots, d_r \in \{1, \ldots, N\}$, $R(d_1, \ldots, d_r)$ holds in W iff $R(\pi(d_1), \ldots, \pi(d_r))$ holds in W'. Similarly, for any constant symbol c' appearing in $\varphi(x)$, if it denotes d' in W, then it denotes $\pi(d')$ in W'.

It should be clear that clusters so defined are mutually exclusive and exhaustive.

We now want to prove that each cluster is, in a precise sense, symmetric with respect to the elements in A_0. That is, let π be any permutation of the domain which is the identity on any element outside of A_0 (i.e., for any $d \notin A_0$, $\pi(d) = d$). Let W be any world in our cluster, and let W' be the world where all the symbols not appearing in φ get the same interpretation as they do in W, while the interpretation of the symbols appearing in φ is obtained from their interpretation in W by applying π as described above. We want to prove that W' is also in the cluster. Condition (1) is an immediate consequence of the definition of W'; the restriction on the choice of π implies condition (2); condition (3) holds by definition. It remains only to prove that $W' \models KB[\vec{\tau}]$. Because of condition (c) in the statement of the theorem, and the fact that vocabulary symbols not in φ have the same denotation in W and in W', this can only happen if some expression $\|\varphi(x)|\psi(x)\|_x$ has different values in W and in W'. We show that this is impossible. Let A be the set of domain elements satisfying $\psi(x)$ for worlds in this cluster. By condition (b) there are only two cases. Either $\psi_0(x) \Rightarrow \neg\psi(x)$, in which case A_0 and A are disjoint, or $KB \models \psi_0(x) \Rightarrow \neg\psi(x)$, so that $A_0 \subseteq A$. In the first case, since π is the identity on A, exactly the same elements of A satisfy $\varphi(x)$ in W' and in W. In the second case, the set of elements satisfying $\varphi(x)$ can change. But because $A_0 \subseteq A$, π is a permutation of A into itself, so the actual number of elements satisfying $\varphi(x)$ cannot change. We conclude that W' does satisfy $KB[\vec{\tau}]$, and is therefore also in the cluster. Since we restricted the cluster to consist only of worlds that are isomorphic to W in the above sense, and we now proved that all worlds formed in this way are in the cluster, the cluster contains precisely all such worlds.

Having defined the clusters, we want to show that the degree of belief of $\varphi(c)$ is in the range $[\alpha, \beta]$ when we look at any single cluster. By assumption, $KB[\vec{\tau}] \models \|\varphi(x)|\psi_0(x)\|_x \in [\alpha, \beta]$. Therefore, for each world in the cluster, the subset of the elements of A_0 that satisfy $\varphi(x)$ is in the interval $[\alpha, \beta]$. Moreover, by condition (a), KB also entails the assertion $\psi_0(c)$. Therefore, the denotation of c is some domain element d in A_0. Condition (d) says that c does not appear in φ, and so the denotation of c is the same for all worlds in the cluster. Now consider a world W in the cluster, and let B be the subset of A_0 whose members satisfy $\varphi(x)$ in W. We have shown that every permutation of the elements in A_0 (leaving the remaining elements constant) has a corresponding world in the cluster. In particular, all possible subsets B' of size $|B|$ are possible denotations for $\varphi(x)$ in worlds in the cluster. Furthermore, because of symmetry, they are all equally likely. It follows that the fixed element d satisfies $\varphi(x)$ in precisely $|B|/|A_0|$ of the worlds in the cluster. Since $|B|/|A_0| \in [\alpha, \beta]$, the probability of $\varphi(c)$ in any one cluster is in this range also.

As in Theorem 4.2.1, the truth of this fact for each cluster implies its truth in general and at the limit. In particular, since $KB[\vec{\tau}] \models \|\varphi(x)|\psi_0(x)\|_x \in [\alpha, \beta]$ for every sufficiently small $\vec{\tau}$, we conclude that $\Pr_\infty(\varphi(c)|KB) \in [\alpha, \beta]$, if the limit exists. ▮

Theorem 4.3.1: *Suppose KB has the form*

$$\bigwedge_{i=1}^{m} (\alpha_i \preceq_{\ell_i} \|\varphi(x)|\psi_i(x)\|_x \preceq_{r_i} \beta_i) \ \wedge \ \psi_1(c) \ \wedge \ KB',$$

and, for all i, KB $\models \forall x\ (\psi_i(x) \Rightarrow \psi_{i+1}(x)) \wedge \neg(||\psi_1(x)||_x \approx_1 0)$*. Assume also that no symbol appearing* $\varphi(x)$ *appears in KB′ or in any* $\psi_i(c)$*. Further suppose that, for some j,* $[\alpha_j, \beta_j]$ *is the tightest interval. That is, for all* $i \neq j$, $\alpha_i < \alpha_j < \beta_j < \beta_i$*. Then*

$$\Pr_\infty(\varphi(c)|KB) \in [\alpha_j, \beta_j].$$

Proof: The proof of the theorem is based on the following result. Consider any KB of the form

$$\neg(||\psi'(x)||_x \approx_1 0) \;\wedge\; \forall x\; (\psi'(x) \Rightarrow \psi(x)) \;\wedge\; \alpha \preceq_\ell \|\varphi(x)|\psi(x)\|_x \preceq_r \beta \;\wedge\; KB'$$

where none of KB', $\psi(x), \psi'(x)$ mention any symbol appearing in $\varphi(x)$. Then, for any $\epsilon > 0$,

$$\Pr_\infty(\alpha \Leftrightarrow \epsilon \leq \|\varphi(x)|\psi'(x)\|_x \leq \beta + \epsilon \mid KB) = 1.$$

Note that this is quite similar in spirit to Theorem 4.2.11. There, we proved that (under certain conditions) an individual c satisfying $\psi(c)$ "inherits" the statistics over $\psi(x)$; that is, the degree of belief is derived from these statistics. This result allows us perform the same type of reasoning for a larger subset of the class defined by $\psi(x)$. Not surprisingly, the proof of the new result is similar to that of Theorem 4.2.11, and we refer the reader to that proof for many of the details.

We begin by clustering worlds exactly as in the earlier proof, with $\psi(x)$ playing the role of the earlier $\psi_0(x)$. Now consider any particular cluster and let A be the corresponding denotation of $\psi(x)$. In the cluster, the proportion of A that satisfies $\varphi(x)$ is some γ such that $\alpha \Leftrightarrow \tau_\ell \leq \gamma \leq \beta + \tau_r$. (Recall that τ_ℓ and τ_r are the tolerances associated with the approximate comparisons $\approx_\ell$ and $\approx_r$ in KB). In this cluster, the denotation of $\varphi(x)$ in A ranges over subsets of A of size $\gamma|A|$. From the proof of Theorem 4.2.11, we know that there is, in fact, an equal number of worlds in the cluster corresponding to every such subset.

Now let A' be the denotation of $\psi'(x)$ in the cluster (recall that it follows from the construction of the clusters that all worlds in a cluster have the same dentoation for $\psi'(x)$). For a proportion $\gamma' \in [0,1]$, we are interested in computing the fraction of worlds in the cluster such that the proportion of $\varphi(x)$ withing A' is γ'. From our discussion above, it follows that this is a purely combinatorial question: given a set A of size n and a subset A' of size n', how many ways are there of choosing γn elements (representing the elements for which $\varphi(x)$ holds) so that $\gamma' n'$ elements come from A'? We estimate this using the observation that the distribution of $\gamma' n'$ is derived from a process ofsampling without replacement.[1] Hence, it behaves according to the well-known hypergeometric distribution (see, for example, [JL81]). We can thus conclude that γ' is distributed with mean γ and variance

$$\frac{\gamma(1 \Leftrightarrow \gamma)(n \Leftrightarrow n')}{(n \Leftrightarrow 1)n'} \leq \frac{\gamma(1 \Leftrightarrow \gamma)}{n'} \leq \frac{1}{4n'} \,.$$

Now, based on our assumption that $KB \models \neg(||\psi'(x)||_x \approx_1 0)$, we know that $n' = |A'| \geq \tau_1 N$. Thus, this variance tends to 0 as N grows large. Now, consider the event: "a world in the cluster has a proportion of $\varphi(x)$ within A' which is not in the interval $[\gamma \Leftrightarrow \epsilon, \gamma + \epsilon]$". By Chebychev's inequality, this is bounded from above by some small probability p_N which depends only on $\tau_1 N$. That is, the fraction of worlds in each cluster that have the "wrong" proportion is at

[1] There are, in fact, a number of ways to solve this problem. One alternative is to use an entropy-based technique (see Chapter 7). We can do this because, at this point in the proof, it no longer matters whether KB uses non-unary predicates or not; we can therefore safely apply techniques that usually only work in the unary case.

most p_N. Since this is the case for every cluster, it is also true in general. More precisely, the fraction of overall worlds for which $\|\varphi(x)|\psi'(x)\|_x \notin [\alpha \Leftrightarrow \tau_\ell \Leftrightarrow \epsilon, \beta + \tau_r + \epsilon]$ is at most p_N. But this probability goes to 0 as N tends to infinity. Therefore,

$$\Pr^{\vec{\tau}}_{\infty}(\alpha \Leftrightarrow \tau_\ell \Leftrightarrow \epsilon \leq \|\varphi(x)|\psi'(x)\|_x \leq \beta + \tau_r + \epsilon \mid KB) = 1.$$

As $\vec{\tau} \rightarrow \vec{0}$ we can simply omit τ_ℓ and τ_r, proving the required result.

It is now a simple matter to prove the theorem itself. Consider the following modification KB'' of the KB given in the statement of the theorem:

$$\bigwedge_{i=j}^{m} (\alpha_i \preceq_{\ell_i} \|\varphi(x)|\psi_i(x)\|_x \preceq_{r_i} \beta_i) \ \wedge\ \psi_1(c) \ \wedge\ KB',$$

where we eliminate the statistics for the reference classes that are contained in ψ_j (the more specific reference classes). From Theorem 4.2.11 we can conclude that $\Pr_\infty(\varphi(c)|KB'') \in [\alpha_j, \beta_j]$ (the conditions of that theorem are clearly satisfied). But we also know, from the result above, that for each ψ_i, for $i < j$:

$$\Pr_\infty(\alpha_j \Leftrightarrow \epsilon \leq \|\varphi(x)|\psi'(x)\|_x \leq \beta_j + \epsilon \mid KB'') = 1.$$

For sufficiently small $\epsilon > 0$, the assertion that

$$\alpha_j \Leftrightarrow \epsilon \leq \|\varphi(x)|\psi'(x)\|_x \leq \beta_j + \epsilon$$

logically implies that

$$\alpha_i \preceq_{\ell_i} \|\varphi(x)|\psi'(x)\|_x \preceq_{r_i} \beta_i,$$

so that this latter assertion also has probability 1 given KB''. We therefore also have probability 1 (given KB'') in the finite conjunction

$$\bigwedge_{i=1}^{j} (\alpha_i \preceq_{\ell_i} \|\varphi(x)|\psi'(x)\|_x \preceq_{r_i} \beta_i).$$

We can now apply Theorem 4.1.2 to conclude that we can add this finite conjunction to KB'' without affecting any of the degrees of belief. But the knowledge base resulting from adding this conjunction to KB'' is precisely the original KB. We conclude that

$$\Pr_\infty(\varphi(c)|KB) = \Pr_\infty(\varphi(c)|KB'') \in [\alpha_j, \beta_j],$$

as required. ▮

Theorem 4.3.4: *Let P be a unary predicate, and consider a knowledge base KB of the following form:*

$$\bigwedge_{i=1}^{m} (\|P(x)|\psi_i(x)\|_x \approx_i \alpha_i \wedge \psi_i(c)) \ \wedge\ \bigwedge_{\substack{i,j=1\\ i\neq j}}^{m} \exists! x\ (\psi_i(x) \wedge \psi_j(x))\ ,$$

where $0 < \alpha_j < 1$*, for* $j = 1, \ldots, m$*. Then, if neither* P *nor* c *appear anywhere in the formulas* $\psi_i(x)$*, then*

$$\Pr_\infty(P(c)|KB) = \delta(\alpha_1, \ldots, \alpha_m).$$

Proof: As in previous theorems, we prove the result by dividing the worlds into clusters. More precisely, consider any $\vec{\tau}$ such that $\alpha_i \Leftrightarrow \tau_i > 0$ and $\alpha_i + \tau_i < 1$. For any such $\vec{\tau}$ and any domain size N, we divide the worlds of size N satisfying $KB[\vec{\tau}]$ into clusters, and prove that, within each cluster, the probability of $\varphi(c)$ is in the interval $[\delta(\alpha_1 \Leftrightarrow \tau_1, \ldots, \alpha_m \Leftrightarrow \tau_m), \delta(\alpha_1 + \tau_1, \ldots, \alpha_m + \tau_m)]$. Since δ is a continuous function at these points, this suffices to prove the theorem.

We partition the worlds satisfying $KB[\vec{\tau}]$ into maximal clusters that satisfy the following three conditions:

1. All worlds in a cluster must agree on the denotation of every vocabulary symbol except for P. In particular, the denotations of $\psi_1(x), \ldots, \psi_m(x)$ is fixed. For $i = 1, \ldots, m$, let A_i denote the denotation of $\psi_i(x)$ in the cluster, and let n_i denote $|A_i|$.
2. All worlds in a cluster must have the same denotation of P for elements in $\overline{A} = \{1, \ldots, N\} \Leftrightarrow \cup_{i=1}^m A_i$.
3. For all $i = 1, \ldots, m$, all worlds in the cluster must have the same number of elements r_i satisfying P within each set A_i. Note that, since all worlds in the cluster satisfy $KB[\vec{\tau}]$, it follows that $r_i/n_i \in [\alpha_i \Leftrightarrow \tau_i, \alpha_i + \tau_i]$ for $i = 1, \ldots, m$.

Now, consider a cluster as defined above. The assumptions of the theorem imply that, besides the proportion constraints defined by the numbers r_i, there are no other constraints on the denotation of P within the sets $A_1, \ldots, A_m$. Therefore, all possible denotations of P satisfying these constraints are possible. There are two types of worlds in the cluster, those that satisfy $P(c)$ and those that do not. Let d be the denotation of c in this cluster. Our assumptions guarantee that d is the only member of $A_i \cap A_j$. Hence, the number of elements of A_i for which P has not yet been chosen is $n_i \Leftrightarrow 1$. In worlds that satisfy $P(c)$, precisely $r_i \Leftrightarrow 1$ of these elements must satisfy P. Since the A_i are disjoint except for d, the choice of P within each A_i can be made independently of the other choices. Therefore, the number of such worlds in the cluster is:

$$\prod_{i=1}^m \binom{n_i \Leftrightarrow 1}{r_i \Leftrightarrow 1}.$$

Similarly, the number of worlds in the cluster for which $P(c)$ does not hold is

$$\prod_{i=1}^m \binom{n_i \Leftrightarrow 1}{r_i}.$$

Therefore, the fraction of worlds in the cluster satisfying $P(c)$ is:

$$\frac{\prod_{i=1}^m \binom{n_i-1}{r_i-1}}{\prod_{i=1}^m \binom{n_i-1}{r_i-1} + \prod_{i=1}^m \binom{n_i-1}{r_i}} = \frac{\prod_{i=1}^m r_i}{\prod_{i=1}^m r_i + \prod_{i=1}^m (n_i \Leftrightarrow r_i)}$$

$$\begin{aligned} &= \frac{\prod_{i=1}^{m} r_i/n_i}{\prod_{i=1}^{m} r_i/n_i + \prod_{i=1}^{m}(n_i \Leftrightarrow r_i)/n_i} \\ &= \delta(r_1/n_1, \ldots, r_m/n_m) \ . \end{aligned}$$

Since δ is easily seen to be monotonically increasing in each of its arguments and $r_i/n_i \in [\alpha_i \Leftrightarrow \tau_i, \alpha_i + \tau_i]$, we must have that $\delta(r_1/n_1, \ldots, r_m/n_m)$ is in the interval $[\delta(\alpha_1 \Leftrightarrow \tau_1, \ldots, \alpha_m \Leftrightarrow \tau_m), \delta(\alpha_1 + \tau_1, \ldots, \alpha_m + \tau_m)]$. Using the same argument as in the previous theorems and the continuity of δ, we deduce the desired result. ▮

Theorem 4.4.1: *Let Φ_1 and Φ_2 be two vocabularies disjoint except for the constant c. Consider $KB_1, \varphi_1 \in \mathcal{L}(\Phi_1)$ and $KB_2, \varphi_2 \in \mathcal{L}(\Phi_2)$. Then*

$$\Pr_\infty(\varphi_1 \wedge \varphi_2 | KB_1 \wedge KB_2) = \Pr_\infty(\varphi_1 | KB_1) \cdot \Pr_\infty(\varphi_2 | KB_2).$$

Proof: As in previous proofs, we first fix N and $\vec{\tau}$. Consider the set of worlds over the vocabulary $\Phi_1 \cup \Phi_2 \cup \{c\}$, and divide these worlds into N clusters, corresponding to the different denotations of the constant c. The worlds in cluster d, for $d = 1, \ldots, N$, are precisely those where the denotation of c is d. It should be clear that all the clusters are isomorphic: for any formula ξ in the language, the number of worlds satisfying ξ is the same in all clusters. Therefore, we can restrict attention to a single cluster. Each cluster defines a denotation for each symbol in Φ_1 and each symbol in Φ_2. Thus, there is a simple mapping between worlds W in the cluster and pairs of worlds (W_1, W_2) where W_1 is a world over $\Phi_1 \cup \{c\}$ and W_2 a world over $\Phi_2 \cup \{c\}$. Moreover, it is clear that $W \models KB_1$ (resp., $W \models \varphi_1$) iff $W_1 \models KB_1$ (resp., $W_1 \models KB_1$), and similarly for W_2. Therefore, the number of worlds W satisfying $KB_1 \wedge KB_2$ is precisely the product of the number of worlds W_1 satisfying KB_1 and the number of worlds W_2 satisfying KB_2. A similar analysis holds for worlds satisfying $(\varphi_1 \wedge KB_1) \wedge (\varphi_2 \wedge KB_2)$. Therefore, the probability in this cluster of $\varphi_1 \wedge \varphi_2$ given $KB_1 \wedge KB_2$ is the product of the respective probabilities of φ_i given KB_i over worlds where the denotation of c is d. Since we have claimed that restricting to a single cluster does not affect the probabilities, we conclude that

$$\Pr_N^{\vec{\tau}}(\varphi_1 \wedge \varphi_2 | KB_1 \wedge KB_2) = \Pr_N^{\vec{\tau}}(\varphi_1 | KB_1) \cdot \Pr_N^{\vec{\tau}}(\varphi_2 | KB_2).$$

By taking limits, we obtain the desired result. ▮

Appendix B

Proofs for Chapter 5

B.1 Simulating Turing machines

The following definition will turn out to be useful.

Definition B.1.1: Let ξ be a formula, and let $\omega(x)$ be a formula with a single free variable x. We define ξ *restricted to* ω to be the formula $\xi' \wedge \xi_\omega$, where ξ' is a conjunction of formulas $\omega(z)$ for any constant or free variable z appearing in ξ, and ξ_ω is defined by induction on the structure of formulas as follows:

- $\xi_\omega = \xi$ for any atomic formula ξ,
- $(\neg\xi)_\omega = \neg\xi_\omega$,
- $(\xi \wedge \xi')_\omega = \xi_\omega \wedge \xi'_\omega$,
- $(\forall y\, \xi(y))_\omega = \forall y(\omega(y) \Rightarrow \xi_\omega(y))$. ▮

Intuitively, ξ restricted to ω holds if ξ holds on the submodel consisting of the set of elements which satisfy ω.

Given a deterministic Turing machine $\mathbf{M}$, we construct $KB_{\mathbf{M}}$ as follows. Think of the computation of $\mathbf{M}$ as consisting of a sequence of instantaneous descriptions (IDs), which specify the head position, state, and the contents of (at least) that part of the tape which has been read or written so far. Without loss of generality, we can assume that the jth ID contains exactly the first j symbols on the tape (padding it with blanks if necessary). The construction uses two binary predicate symbols, H and V, to impose a matching "layered" structure on the elements of a finite domain (see Figure B.1).

More specifically, we force the domain to look like a sequence of n layers for some n, where there are exactly j elements in the jth layer for $1 \leq j < n$, but the last layer may be "incomplete", and have less than n elements. (This ensures that such a partition of domain elements into layers is possible for any domain size.) We construct each layer separately, by

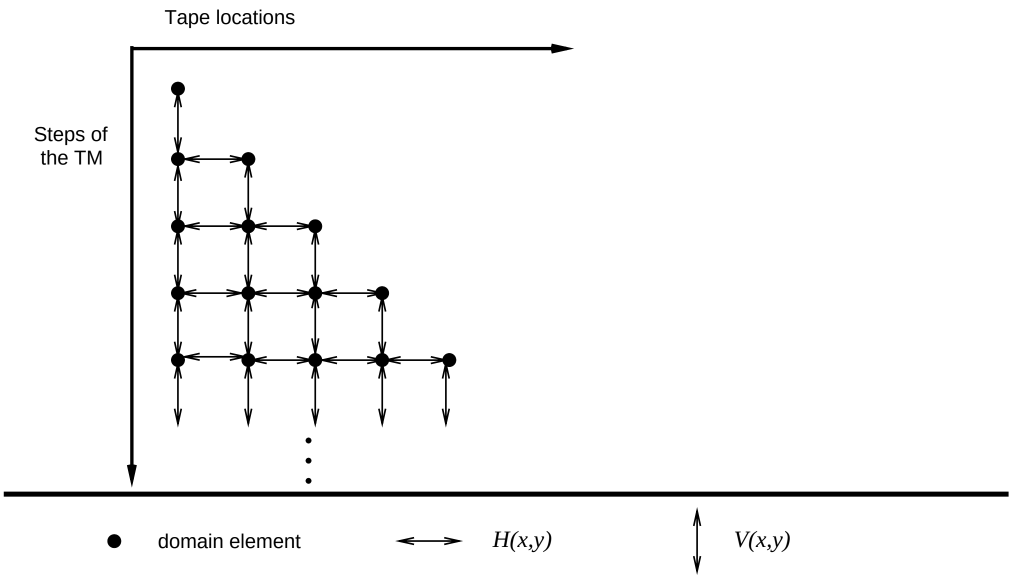

Figure B.1: The structure forced by $KB_{\mathbf{M}}$

assigning each element a *horizontal successor*. The horizontal successor of the ith element in the jth layer is the $(i+1)$st element in the jth layer. This successor must exist except when i is the last element in the layer $(i = j)$, or j is the last (and possibly incomplete) layer $(j = n)$. We connect one layer to the next by assigning each element a *vertical successor*. The vertical successor of the ith element in the jth layer is the ith element in the $(j+1)$st layer. This successor must exist except if j is the last layer $(j = n)$, and possibly if j is the next-to-last layer $(j = n \Leftrightarrow 1)$. These two types of successor relationship are captured using H and V: $H(x, y)$ holds iff y is the horizontal successor of x, and $V(x, y)$ holds iff y is the vertical successor of x. Straightforward assertions in first-order logic can be used to constrain H and V to have the right properties.

We use the jth layer to encode the jth ID, using unary predicates to encode the contents of each cell in the ID and the state of the machine $\mathbf{M}$. It is straightforward to write a sentence $KB_{\mathbf{M}}$ that ensures that this simulation of the Turing machine starts correctly, and continues according to the rules of $\mathbf{M}$. It follows that there is an exact one-to-one correspondence between finite models of $KB_{\mathbf{M}}$ and finite prefixes of computations of $\mathbf{M}$, as required.

We have assumed that two binary and several unary predicate symbols are available. In fact, it is possible to do all the necessary encoding using only a single binary (or any non-unary) predicate symbol. Because this observation will be important later, we sketch how the extra predicate and constant symbols can be eliminated. First, note that the predicates H and V can be encoded using a single predicate R. Since H holds only between elements on the same layer, and V only between elements on two consecutive layers, we can define $R(x, y)$ to mean $H(x, y)$ in the first case, and $V(x, y)$ in the second (we can construct the sentences so that it is easy to tell whether two elements are on the same layer). Any unary predicate P used in the construction can be eliminated by replacing $P(x)$ with $R(c, x)$ for some special constant symbol c. We then replace $KB_{\mathbf{M}}$ with $KB_{\mathbf{M}}$ restricted to $x \neq c$, as in Definition B.1.1, thus making the denotation of c a distinguished element which does not participate in the construction of the Turing machine. Finally, it is possible to eliminate the use of constant symbols by using additional variables quantified with "exists unique"; we omit details. However, note for future reference that for every constant we eliminate, we increase the quantifier depth of the formula by one.

This construction has another very useful property. First, note that the layered structure imposed by H and V ensures that every domain element plays a unique role (i.e., for each element we can find a first-order formula with one free variable which holds of that element and no other). So if we (nontrivially) permute the domain elements in one model, we obtain a different (although isomorphic) model. This property has been called *rigidity*. Rigidity implies that, if the domain size is N, every isomorphism class of worlds satisfying $KB_{\mathbf{M}}$ contains exactly $N!$ worlds. This implies that any two size N models of $KB_{\mathbf{M}}$ are isomorphic (because the machine $\mathbf{M}$ is assumed to be deterministic and thus has a unique computation path when started on the empty input). From this observation and rigidity, we conclude that the number of size N models of $KB_{\mathbf{M}}$ is exactly $N!$; this fact will also be useful later.

B.2 Nonexistence proof

Theorem 5.3.2: *Let A be any computable regular matrix transform, and let Φ be a vocabulary containing at least one non-unary predicate symbol. There exist $\varphi, KB \in \mathcal{L}(\Phi)$ such that the A-transform of the sequence $\Pr_N(\varphi|KB)$ exists, but does not converge.*

Proof: In the following, let U be a rational number within 0.01 of $\limsup_{i\to\infty} \sum_{j=1}^{\infty} |a_{ij}|$, i.e., $|U \Leftrightarrow \limsup_{i\to\infty} \sum_{j=1}^{\infty} |a_{ij}| < 0.01$. We will use U as a parameter to the algorithm we are about to construct. Notice that although the existence of an appropriate U is guaranteed by R3, we may not be able to compute its value. Thus, the proof we are about to give is not necessarily constructive. On the other hand, this is the only nonconstructive aspect of our algorithm. A value for U is computable in many cases of interest (for example, if a_{ij} is nonnegative for all i and j, then we can take $U = 1$); in these cases, our proof becomes constructive. Let $i_{\min}$ be such that whenever $i \geq i_{\min}$, we have $\sum_{j=1}^{\infty} |a_{ij}| < U + 0.01$. Such an $i_{\min}$ must exist (because of the way U is defined); it is not necessarily computable either, but the following does not actually depend on its value (i.e., we only refer to $i_{\min}$ when proving that the constructed machine works as required).

We use the value of U in the construction of a three-tape four-head Turing machine $\mathbf{M}$. Tape 2 of $\mathbf{M}$ will always (after the first step) contain an alternating sequence of 0's and 1's. The sentence $KB_{\mathbf{M}}$ is constructed so that finite models of KB encode partial computations of $\mathbf{M}$, exactly as outlined in Section B.1. The sentence φ is chosen to be true only in models of KB where the last element written on tape 2 is 1. Note that, as usual, we can assume that $\varphi, KB_{\mathbf{M}} \in \mathcal{L}(\{R\})$ for a binary predicate symbol R.

The idea of the proof is as follows. Suppose b_j is the truth value of φ (either 0 or 1) in a domain of size j, and let $c_i = \sum_{j=1}^{\infty} a_{ij} b_j$. Obviously, the sequence (b_j) is determined by the times at which $\mathbf{M}$ writes a new symbol to tape 2. We construct $\mathbf{M}$ to guarantee that the sequence (b_j) has appropriately spaced runs of zeros and ones, so that there are infinitely many i where c_i is greater than 0.9 and infinitely many i where c_i is less than 0.1. This ensures that the sequence (c_i) does not converge.

As we have said, $\mathbf{M}$ is a three-tape four-head Turing machine. Heads 1a and 1b read tape 1, head 2 reads tape 2, and head 3 reads tape 3. We assume that any subset of heads can move in the same step. Tape 1 is used for keeping track, in unary, of the number of steps that $\mathbf{M}$ has taken so far. Tape 2 contains an alternating sequence of 0's and 1's. As we have indicated, the goal of the rest of the construction will be to ensure that tape 2 is updated at appropriate intervals. Finally, tape 3 is a work tape, used for all necessary calculations.

Every fourth step, head 1a writes a 1 at the right end of tape 1, and then moves one step to the right. This is done independently of the operation of the rest of the machine. Thus, if we represent the number written on tape 1 at a certain point as m, the actual number of steps taken by $\mathbf{M}$ up to that point is between $4m$ and $4m+3$. Moreover, if we assume (as we do without loss of generality) that the size of the ith ID of the computation of $\mathbf{M}$ is i, then to encode the first i steps of the computation we need a domain of size $i(i+1)/2+C$, where C is a constant independent of i. In particular, the size of the domain required to encode the prefix of the computation at the point where m is the number on tape 1 is roughly $2m(4m+1)$, and is certainly bounded above by $9m^2$ and below by $7m^2$ for all sufficiently large m. We will use these estimates in describing $\mathbf{M}$.

The machine $\mathbf{M}$ proceeds in phases; each phase ends by writing a symbol on tape 2. At the completion of phase k, for all k large enough, there will exist some number i_k such that $c_{i_k} < 0.1$ if k is even, and $c_{i_k} > 0.9$ if k is odd. Since we will also show that $i_{k+1} > i_k$, this will prove the theorem.

The first phase consists of one step; at this step, $\mathbf{M}$ writes 0 on tape 2, and head 2 moves to the right. Suppose the kth phase ends with writing a 1 on tape 2. We now describe the $(k+1)$st phase. (The description if the kth phase ends with writing a 0 is almost identical, and left to the reader.)

Let n_l be the size of the domain required to encode the prefix of the computation up to the end of phase l. Since the value at the end of tape 2 changes only at the end of every phase, and b_j is 1 if and only if the last element on tape 2 is 1, b_j is 0 for $n_1 \leq j < n_2$, b_j is 1 for $n_2 \leq j < n_3$, and so on. $\mathbf{M}$ begins the $(k+1)$st phase by copying the number m on tape 1 to tape 3 (the work tape). The copying is done using head 1b (head 1a continues to update the number every fourth step). Suppose the number eventually copied is m_k. Clearly, m_k will be greater than the number that was on tape 1 in the computation prefix that was encoded by domain size n_k. Therefore, $n_k < 9m_k^2$ for k sufficiently large.

We now get to the heart of the construction, which is the computation of when to next write a value on tape 2. (Note that this value will be a 0, since we want the values to alternate.) Notice that by R1, R2, and R3 there must be a pair (i_*, j_*) such that:

(a) $i_* > m_k$,

(b) $\sum_{j=1}^{9m_k^2} |a_{i_*j}| < 0.01$,

(c) $\sum_{j=1}^{j_*} a_{i_*j} > 0.99$, and

(d) $\sum_{j=1}^{j_*} |a_{i_*j}| > U \Leftrightarrow 0.01$.

Moreover, since a_{ij} is computable for all i and j, $\mathbf{M}$ can effectively find such a pair by appropriate dovetailing. Suppose that in fact $i_* > i_{\min}$. (Since $i_* > m_k$ by part (a), this will be true once k is large enough.) Then we claim that, no matter what the values of $b_0, \ldots, b_{n_k}$ and $b_{j_*+1}, b_{j_*+2}, \ldots$, if $b_{n_k+1} = \cdots = b_{j_*} = 1$, then $c_{i_*} > 0.9$. To see this, note that if $i_* > i_{\min}$, then (by definition of $i_{\min}$) $\sum_{j=1}^{\infty} |a_{i_*j}| < U + 0.01$. Thus, by part (d) above it follows that $\sum_{j=j_*+1}^{\infty} |a_{i_*j}| < 0.02$. Using part (b) and the fact that $n_k < 9m_k^2$, it follows that $\sum_{j=1}^{n_k} |a_{i_*j}| < 0.01$. Now from part (c) we get that $\sum_{j=n_k+1}^{j_*} a_{i_*j} > 0.98$. If $b_{n_{k+1}} = \cdots = b_{j_*} = 1$, then

$$\begin{aligned}
c_{i_*} &= \textstyle\sum_{j=1}^{\infty} a_{i_*j} b_j \\
&= \textstyle\sum_{j=1}^{n_k} a_{i_*j} b_j + \sum_{j=n_k+1}^{j_*} a_{i_*j} b_j + \sum_{j=j_*+1}^{\infty} a_{i_*j} b_j \\
&\geq \textstyle\sum_{j=n_k+1}^{j_*} a_{i_*j} \Leftrightarrow \sum_{j=1}^{n_k} |a_{i_*j}| \Leftrightarrow \sum_{j=j_*+1}^{\infty} |a_{i_*j}| \\
&\geq 0.98 \Leftrightarrow 0.01 \Leftrightarrow 0.02 \\
&> 0.9.
\end{aligned}$$

Thus, it suffices for $\mathbf{M}$ to add the next 0 to tape 2 so as to guarantee that $n_{k+1} > j_*$, since our choice of φ will then guarantee that $b_{n_k+1} = \cdots = b_{j_*} = 1$. This can be done by waiting to add the 0, until after the number m on tape 1 is such that $7m^2 > j_*$. As we observed above, the size of the domain required to encode the prefix of the computation up to this point is at least $7m^2$. Since this domain size is n_{k+1} by definition, it follows that $n_{k+1} \geq j_*$, as desired.

This completes the description of the $(k+1)$st phase. We can then take $i_{k+1} = i_*$, and guarantee that $c_{i_{k+1}} > 0.9$, as desired. Note that, for every k, $i_{k+1} > m_k$, and (m_k) is a strictly increasing sequence. Thus, we obtain infinitely many indices i at which $c_i > 0.9$ and infinitely many at which $c_i < 0.1$, as desired.

Since $\#worlds_N^{\{R\}}(KB) \neq 0$ for all sufficiently large N, this shows that both $\Box\Diamond\Pr_\infty(\varphi|KB)$ and $\Diamond\Box\Pr_\infty(\varphi|KB)$ are well-defined, but their A-transform does not converge. ▮

B.3 Undecidability proofs

Theorem 5.4.1: *Let Φ be a vocabulary containing at least one non-unary predicate symbol.*

(a) The problem of deciding whether a sentence in $\mathcal{L}(\Phi)$ is satisfiable for infinitely many domain sizes is Π_2^0-complete.

(b) The problem of deciding whether a sentence in $\mathcal{L}(\Phi)$ is satisfiable for all but finitely many domain sizes is Σ_2^0-complete.

Proof: We start with the upper bounds. First observe that the problem of deciding whether a first-order sentence ξ is satisfiable in some model with domain size N, for some fixed N, is recursive (and with the help of some suitable encoding of formulas as natural numbers, we can encode this problem in the language of arithmetic). Given this, deciding if ξ is satisfiable in infinitely many domain sizes can be encoded using a Π^0_2 block: for all N, there exists $N' > N$ such that ξ holds in some model of domain size N'. Similarly, deciding if ξ is satisfiable for all but finitely many domain sizes can clearly be encoded using a Σ^0_2 block: there exists N such that for all $N' > N$, ξ holds in some model with domain size N'. This proves the upper bounds.

It is well known that the following problem is Π^2_0-complete [Rog67]: "Given a Turing machine $\mathbf{M}$, does $\mathbf{M}$ halt on infinitely many inputs?" It is also well known that the following (dual) problem is Σ^2_0-complete: "Given a Turing machine $\mathbf{M}$, does $\mathbf{M}$ halt on only finitely many inputs?" We prove the two lower bounds by reducing these problems to intermittent and persistent well-definedness, respectively. First, given an arbitrary Turing machine $\mathbf{M}$, we effectively construct another Turing machine $\mathbf{M}'$ that, when started on empty input, starts simulating the computations of $\mathbf{M}$ on all inputs by dovetailing, and enters a special state q_s once for each input on which $\mathbf{M}$ halts. (We leave details of this construction to the reader.) Let $KB_{\mathbf{M}'}$ be the sentence that forces its models to encode prefixes of the computation of $\mathbf{M}'$ on empty input, as described in Section B.1, and let φ be the sentence that says, with respect to this encoding, that the last layer is complete, and that $\mathbf{M}'$ is in state q_s in the ID encoded in this last layer. Clearly $\varphi \wedge KB_{\mathbf{M}'}$ is satisfiable for infinitely many domain sizes N iff $\mathbf{M}$ halts on infinitely many inputs, while $\neg\varphi \wedge KB_{\mathbf{M}'}$ is satisfiable for all but finitely many domain sizes N iff $\mathbf{M}$ halts on only finitely many inputs. This proves the lower bounds. ▮

We prove Theorem 5.4.3 by first showing that the problem of deciding whether an r.e. sequence of rationals converges to 0 is Π^0_3-complete.

Theorem B.3.1: *The problem of deciding whether a recursively enumerable infinite sequence of rational numbers converges to zero is Π^0_3-complete.*

Proof: The following problem is known to be Π^0_3-complete: "Does each of the Turing machines in a given r.e. set of Turing machines diverge on all but finitely many inputs?", where the input to this problem is itself a Turing machine (that generates the encodings for the collection of Turing machines we are asking about). See [Rog67] for details. For our purposes it is slightly better to consider a variant of this problem, namely "Does each of the Turing machines in a given r.e. set of Turing machines enter some distinguished state, say q_s, only finitely many times when started on the empty input?" The two problems are easily seen to be equivalent, in that either one can be effectively reduced to the other.

The lower-bound is proved by reducing this problem to the question of whether a sequence converges to zero. We assume, without loss of generality, that our Turing machine generator $\mathbf{G}$ computes a total function, whose values are encodings of other Turing machines. That is, on input i, it is guaranteed to terminate and produce the ith machine (note that the machines produced by $\mathbf{G}$ on different inputs are not necessarily distinct). We now define H_{ij} to have value 1 if the ith machine generated by $\mathbf{G}$ is in state q_s on its jth step after being started on empty input, and value 0 otherwise. Note that H_{ij} is a computable function of i, j, and the encoding of $\mathbf{G}$, because we can simulate $\mathbf{G}$ to obtain the encoding of the ith machine, then simulate this machine for j steps.

We use the numbers H_{ij} to define an r.e. sequence $s_1, s_2, \ldots$ of rational numbers in [0,1], where s_k is defined as $0.H_{1k}\,H_{2k}\,\ldots H_{kk}$. The computability of H_{ij} guarantees that this sequence is recursively enumerable. Clearly the sequence $s_1, s_2, \ldots$ converges to 0 iff, for all i, the sequence $H_{i1}, H_{i2}, \ldots$ is eventually 0, i.e., there exists n_i such that $H_{ij} = 0$ for all $j > n_i$. But the sequence $H_{i1}, H_{i2}, \ldots$ is eventually 0 iff the ith Turing machine reaches q_s only finitely often. This proves the lower bound.

For the upper bound, note that the question of whether the limit of $s_1, s_2, \ldots$ exists and equals 0 can be written: "For all M, does there exist N_0 such that for all $N > N_0$, $|s_N| < 1/M$?" The unquantified part of this question is clearly recursive and can be formulated in the language of arithmetic, while the quantifier block is a Π_3^0 prefix. The result follows. ∎

Theorem 5.4.3: *Let Φ be a vocabulary containing at least one non-unary predicate symbol. For sentences $\varphi, KB \in \mathcal{L}(\Phi)$, the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) exists is Π_3^0-complete. The lower bound holds even if we have an oracle that tells us whether the limit is well-defined and its value if it exists.*

Proof: To prove the lower bound, we reduce the problem of deciding if an r.e. sequence of rationals converges to 0 to that of deciding if a particular asymptotic conditional probability exists. Suppose $\mathbf{S}$ is a machine that generates an infinite sequence of rational numbers, $s_1, s_2, \ldots$. Without loss of generality, we can assume that the numbers are in $[0,1]$; if necessary, a new machine $\mathbf{S}'$ such that $s_i' = max(1, |s_i|)$ is easily constructed which clearly has the same properties with respect to convergence to zero. We also assume that the output is encoded in a special form: a rational value a/b is output on the tape as a sequence of a 1's, followed by $(b \Leftrightarrow a)$ 0's, suitably delimited.

Let R be a binary predicate symbol. (Of course, any non-unary predicate will suffice.) We begin by constructing $KB_{\mathbf{S}} \in \mathcal{L}(\{R\})$ such that finite models of $KB_{\mathbf{S}}$ correspond naturally to prefixes of computations of $\mathbf{S}$, as described in Section B.1. Let c be a constant. Let $KB'_{\mathbf{S}} \in \mathcal{L}(\{c, R\})$ be the conjunction of $KB_{\mathbf{S}}$ and sentences asserting that, in the computation-prefix of $\mathbf{S}$ encoded by the domain, the denotation of c corresponds to a cell in that section of the last complete ID that represents the output. Note that for any fixed domain size, $KB'_{\mathbf{S}}$ has $a + (b \Leftrightarrow a) = b$ times as many models over $\{c, R\}$ as $KB_{\mathbf{S}}$ does over $\{R\}$, where a/b is the most recent sequence value generated by $\mathbf{S}$ in the computation simulated so far. According to our discussion at the end of Section B.1, $\#worlds_N^{\{R\}}(KB_{\mathbf{S}}) = N!$, so $\#worlds_N^{\{c,R\}}(KB'_{\mathbf{S}}) = b \cdot N!$.

To complete the reduction, consider a sentence φ that says that the simulated computation has just finished writing another sequence element, and the denotation of c corresponds to a cell in that output containing the symbol 1. Assume that the last sequence element written in the prefix corresponding to domain size N is a/b. Note that if there are models of $\varphi \wedge KB'_{\mathbf{S}}$ of domain size N, then there are in fact $a \cdot N!$ such models over $\{c, R\}$ (corresponding to the a choices for the denotation of c). In this case $\Pr_N(\varphi|KB'_{\mathbf{S}})$ has value a/b. It follows that the sequence $\Pr_N(\varphi|KB'_{\mathbf{S}})$, for increasing N, is precisely the sequence generated by **S** interspersed with zeros at domain sizes corresponding to computations that have not just output a new value. Note that both persistent and intermittent limits are well-defined for this sequence. If this limit exists at all, it must have value zero, and this will be the case just if the sequence generated by **S** has this property. This proves the lower bound. We remark that the use of an extra constant c is not necessary in our proof; it can be eliminated as discussed in Section B.1.

To prove the upper bound, note that the question of existence for $\Box\Diamond\Pr_\infty(\varphi|KB)$ can be stated as: "Is it true that for all integers M, there exist rational numbers $r_1 \leq r_2$ and integers N_0 and $N_1 > M$ such that for all $N \geq N_0$, (1) $\#worlds^\Phi_{N_1}(KB) \neq 0$, (2) if $\#worlds^\Phi_N(KB) \neq 0$, then $\Pr_N(\varphi|KB) \in [r_1, r_2]$, and (3) $r_2 \Leftrightarrow r_1 \leq 1/M$?" The unquantified part is clearly recursive, showing that the problem of deciding whether $\Box\Diamond\Pr_\infty(\varphi|KB)$ exists is in Π^3_0. We can state the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ exists as follows: "Is it true that for all integers M, there exist rational numbers $r_1 \leq r_2$ and an integer N_0 such that for all $N \geq N_0$, (1) $\#worlds^\Phi_N(KB) \neq 0$, (2) $\Pr_N(\varphi|KB) \in [r_1, r_2]$, and (3) $r_2 \Leftrightarrow r_1 \leq 1/M$?" Thus, the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ exists is also in Π^0_3. ▮

Theorem 5.4.4: *Let Φ be a vocabulary containing at least one non-unary predicate symbol, and let $r, r_1, r_2 \in [0,1]$ be rational numbers such that $r_1 \leq r_2$. For sentences $\varphi, KB \in \mathcal{L}(\Phi)$, given an oracle for deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) exists,*

(a) the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB) = r$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB) = r$) is Π^0_2-complete,

(b) if $[r_1, r_2] \neq [0,1]$, then the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB) \in [r_1, r_2]$) is Π^0_2-complete,

(c) if $r_1 \neq r_2$, then the problem of deciding if $\Diamond\Box\Pr_\infty(\varphi|KB) \in (r_1, r_2)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB) \in (r_1, r_2)$) is Σ^0_2-complete.

Proof: We start with part (a). Just as with our earlier results, the upper bound is the easier part. This problem can be stated as "For all M, does there exist an $N > M$ such that $\#worlds^\Phi_N(KB) > 0$, and $|\Pr_N(\varphi|KB) \Leftrightarrow r| < 1/M$?" It is easy to see that this sentence has the appropriate form for Π^0_2. Furthermore, it is true just if there is some subsequence of domain sizes such that the asymptotic probability, when restricted to these sizes, has value r. If the sequence as a whole has any limit at all (and we can check this with the oracle) then this limit must also be r.

To prove the lower bound, we proceed just as in the proof of Theorem 5.4.1 by reducing the problem "Does a Turing machine reach a specified state q_s infinitely often?" to the problem of deciding whether the asymptotic probability is r. Let **M** be an arbitrary Turing machine. As discussed in Section B.1, we can find a sentence $KB_{\mathbf{M}} \in \mathcal{L}(\{R\})$ such that finite models of $KB_{\mathbf{M}}$ correspond naturally to prefixes of computations of **M**.

Our next step is to construct sentences φ_r and KB_r such that $\Pr_N(\varphi_r|KB_r) = r$, for all N. Suppose $r = a/b$, and choose k such that $2^k > b$. We can easily construct propositional formulas α_r and β_r using k primitive propositions $p_1, \ldots, p_k$ such that β_r has exactly b satisfying assignments and $\alpha_r \wedge \beta_r$ has exactly a satisfying assignments. Let φ_r and KB_r be the sentences that result by replacing occurrences of the primitive proposition p_i in α_r or β_r by $P_i(c)$, where P_i is a unary predicate symbol, and c is a constant symbol. It is easy to see that $\Pr_N(\varphi_r | KB_r) = r$ for all N.

Let Q be a unary predicate not among $\{P_1, \ldots, P_k\}$, and let KB' be a sentence asserting that there is exactly one domain element satisfying Q, and that this element corresponds to one of the tape cells representing the head position when the machine is in state q_s. Define KB to be $KB_{\mathbf{M}} \wedge KB_r \wedge (KB' \vee \forall x\, Q(x))$. For any domain size N, let t_N denote the number of times the machine has reached q_s in the computation so far. The sentence KB has $t_N + 1$ times as many models over $\{R, P_1, \ldots, P_k, Q, c\}$ as the sentence $KB_{\mathbf{M}} \wedge KB_r$ has over $\{R, P_1, \ldots, P_k, c\}$. We now consider two cases: $r < 1$ and $r = 1$. If $r < 1$, let φ be simply $\varphi_r \vee (\neg\varphi_r \wedge \forall x\, Q(x))$. It is easy to see that $\Pr_N(\varphi | KB)$ is $r + (1 \Leftrightarrow r)/(t_N + 1)$. If $\mathbf{M}$ reaches q_s finitely often, say t' times, the limit as $N \rightarrow \infty$ is $r + (1 \Leftrightarrow r)/(t' + 1)$, otherwise the limit is r. The limit always exists, so our oracle is not helpful. This proves the required lower bound if $r < 1$. If $r = 1$, then we can take KB to be $KB_{\mathbf{M}} \wedge (KB' \vee \forall x\, Q(x))$ and φ to be $\neg \forall x\, Q(x)$. In this case, $\Pr_N(\varphi | KB)$ is $t_N/(t_N + 1)$; therefore, the limit is 1 if $\mathbf{M}$ reaches q_s infinitely often, and strictly less than 1 otherwise. Again, the lower bound follows. Note that, as discussed in Section B.1, we can avoid actually using new unary predicates and constants by encoding them with the binary predicate R.

For part (b), the upper bound follows using much the same arguments as the upper bound for part (a). For the lower bound, we also proceed much as in part (a). Suppose we are given an interval $[r_1, r_2]$ with $r_2 < 1$, and a Turing machine $\mathbf{M}$. Using the techniques of part (a), we can construct sentences φ and KB such that $\Box\Diamond\Pr_\infty(\varphi | KB)$ and $\Diamond\Box\Pr_\infty(\varphi | KB)$ are both well-defined, and such the asymptotic probability is r_2 if $\mathbf{M}$ reaches state q_s infinitely often, and strictly greater than r_2 otherwise. This proves the lower bound in this case. If $r_2 = 1$, we use similar arguments to construct sentences φ and KB such that the asymptotic conditional probability is r_1 if $\mathbf{M}$ reaches state q_s infinitely often, and is strictly less than r_1 otherwise. Again, the lower bound follows.

Finally, for part (c), observe that the asymptotic probability is in (r_1, r_2) iff it is not in $[0, r_1] \cup [r_2, 1]$. The arguments of part (b) showing that checking whether the asymptotic probability is in a closed interval is Π_2^0-complete can be extended without difficulty to dealing with the union of two closed intervals. Thus, the problem of deciding whether the asymptotic probability is in an open interval is Σ_2^0-complete. ∎

B.4 Eliminating equality

Theorem 5.5.1: *Suppose G and E are binary predicate symbols not appearing in Φ, and $\varphi, KB \in \mathcal{L}(\Phi)$ are such that $\#worlds_N^{\Phi}(KB)$ is a non-decreasing function of N. Then we can* *find sentences* $\varphi', KB' \in \mathcal{L}^{-}(\Phi \cup \{G, E\})$ *such that*

$$\lim_{N\to\infty} (\Pr_N(\varphi|KB) \Leftrightarrow \Pr_N(\varphi'|KB')) = 0 \ .$$

Proof: The idea of the proof is somewhat similar to that used in [KV90] to eliminate equality. Let φ and KB be as in the hypotheses of the theorem. Define KB^E to be the result of replacing all subformulas of KB of the form $t_1 = t_2$ by $E(t_1, t_2)$; we define φ^E similarly. Thus, we are using E to represent equality. Let η be a conjunction of formulas that force E to be an equivalence relation, as well as a congruence on G and on all symbols in Φ. Thus, a typical conjunct of η (which in fact forces E to be a congruence on G) has the form:

$$\forall x\, y\, z(E(x,y) \Rightarrow ((G(x,z) \Leftrightarrow G(y,z)) \wedge (G(z,x) \Leftrightarrow G(z,y)))).$$

Let KB' be $KB^E \wedge \eta$, and φ' be φ^E.

As we now show, there are many more models of KB' of size N where E is true equality than there are where E is some equivalence relation other than equality. To simplify the notation, we write w_N instead of $\#\mathit{worlds}^{\Phi}_N(KB)$. It is easy to see that there are precisely $w_N \cdot 2^{N^2}$ models of size N of KB' over $\Phi \cup \{G, E\}$ where E is equality: for every model of size N of KB over Φ, there are 2^{N^2} models of KB', because the choice of G is unrestricted.

Now we must get an estimate on the number of models of KB' where E is an equivalence relation, but not equality. It turns out that the crucial factor is the number of equivalence classes into which E partitions the domain. Let $\genfrac\{\}{0pt}{}{N}{k}$ be the number of ways of partitioning N elements into exactly k equivalence classes. ($\genfrac\{\}{0pt}{}{N}{k}$ is known as a *Stirling number of the second kind*; see [GKP89]). It is easy to see that there are $w_k \cdot \genfrac\{\}{0pt}{}{N}{k} \cdot 2^{k^2}$ models of KB' where E partitions the domain into k equivalence classes, since for each such way, there are 2^{k^2} choices for G, and w_k choices for the denotations of the predicates in Φ that make KB^E true. Thus, our goal is to show that $(\sum_{k=1}^{N-1} w_k \cdot \genfrac\{\}{0pt}{}{N}{k} \cdot 2^{k^2})/w_N \cdot 2^{N^2}$ asymptotically converges to 0.

To do this, we need a good estimate on $\genfrac\{\}{0pt}{}{N}{k}$. We begin by showing that $\binom{N}{k}N!$ is an overestimate for $\genfrac\{\}{0pt}{}{N}{k}$. To see this, consider any partition, order the equivalence classes by the minimal elements appearing in them, and order the elements in an equivalence class in increasing order. This gives us an ordering of the N elements in the domain. Suppose the equivalence classes (listed in this order) have size $n_1, \ldots, n_k$. This corresponds to choosing elements $n_1, n_1 + n_2, \ldots, n_1 + \cdots + n_k$ from the domain. Thus, with each partition into k equivalence classes, we can associate a unique pair consisting of a permutation and a choice of k elements out of N.

This estimate suffices for values of k which are relatively small compared to N. We use a finer estimate for $\genfrac\{\}{0pt}{}{N}{k}$ if $k \geq N \Leftrightarrow \log N$. In this case, at least $k \Leftrightarrow \log N$ equivalence classes must have size 1. The remaining $\log N$ equivalence classes partition at most $N \Leftrightarrow (k \Leftrightarrow \log N) \leq 2 \log N$ elements. Thus, a bound on $\genfrac\{\}{0pt}{}{N}{k}$ in this case is given by

$$\binom{N}{k \Leftrightarrow \log N} \left\{ {N \Leftrightarrow (k \Leftrightarrow \log N)} \atop {\log N} \right\} \leq \binom{N}{N \Leftrightarrow 2 \log N} \left\{ {2 \log N} \atop {\log N} \right\}$$

$$\begin{aligned}
&\le \binom{N}{N \Leftrightarrow 2\log N}\binom{2\log N}{\log N}(2\log N)! \\
&\le \binom{N}{N \Leftrightarrow 2\log N} 2^{2\log N}\,(2\log N)! \\
&= \frac{N!}{(N \Leftrightarrow 2\log N)!}\,2^{2\log N} \\
&\le N^{2\log N}\,2^{2\log N} \\
&= 2^{2\log^2 N+2\log N}\ .
\end{aligned}$$

Thus, we have that

$$\begin{aligned}
\sum_{k=1}^{N-1}\begin{Bmatrix}N\\k\end{Bmatrix}\cdot 2^{k^2} &= \sum_{k=1}^{N-\log N}\begin{Bmatrix}N\\k\end{Bmatrix}\cdot 2^{k^2} + \sum_{k=N-\log N+1}^{N-1}\begin{Bmatrix}N\\k\end{Bmatrix}\cdot 2^{k^2} \\
&\le N!\,2^{(N-\log N)^2}\left(\sum_{k=1}^{N-\log N}\binom{N}{k}\right) + 2^{2\log^2 N+2\log N}\sum_{k=N-\log N+1}^{N-1} 2^{k^2} \\
&\le 2^{N\log N}\,2^{(N-\log N)^2}2^{N} + 2^{2\log^2 N+2\log N}2^{(N-1)^2+1} \\
&\le 2^{N^2-N\log N+N+\log^2 N} + 2^{N^2-2N+2\log^2 N+2\log N+2} \\
&\le 2^{N^2-\Omega(N)}\ .
\end{aligned}$$

Let σ be the formula $E(x,y) \Leftrightarrow x = y$, which says that E is true equality. (Note that σ is not in $\mathcal{L}^-(\Phi\cup\{G,E\})$, since it mentions $=$, but that is not relevant to the discussion below.) It now easily follows that for any $\epsilon > 0$, we can choose N_0 large enough, so that for any $N > N_0$,

$$\begin{aligned}
\Pr_N(\neg\sigma|KB') &\le \frac{\sum_{k=1}^{N-1} w_k\cdot\left\{{N \atop k}\right\}\cdot 2^{k^2}}{w_N\cdot 2^{N^2}} \\
&\le \frac{w_N\sum_{k=1}^{N-1}\left\{{N \atop k}\right\}\cdot 2^{k^2}}{w_N\cdot 2^{N^2}} \\
&\le \frac{2^{N^2-\Omega(N)}}{2^{N^2}} = 2^{-\Omega(N)} < \epsilon/2\ .
\end{aligned}$$

Therefore, since $\Pr_N(\varphi'|KB'\wedge\sigma) = \Pr_N(\varphi|KB)$, it follows that

$$\begin{aligned}
&|\Pr_N(\varphi'|KB') \Leftrightarrow \Pr_N(\varphi|KB)| \\
&\quad= |[\Pr_N(\varphi'|KB'\wedge\sigma)\cdot\Pr_N(\sigma|KB') + \Pr_N(\varphi'|KB'\wedge\neg\sigma)\cdot\Pr_N(\neg\sigma|KB')] \Leftrightarrow \Pr_N(\varphi|KB)| \\
&\quad\le |\Pr_N(\varphi|KB)(1 \Leftrightarrow \Pr_N(\sigma|KB'))| + |\Pr_N(\neg\sigma|KB')| \\
&\quad\le \epsilon/2 + \epsilon/2 = \epsilon \quad \blacksquare
\end{aligned}$$

Theorem 5.5.3: *Let Φ be a vocabulary containing at least two non-unary predicate symbols. For $KB \in \mathcal{L}^-(\Phi)$, the problem of deciding if $\Box\Diamond\Pr_\infty(*|KB)$ (resp., $\Diamond\Box\Pr_\infty(*|KB)$) is well-defined is r.e.-complete.*

Proof: We can state the problem of deciding whether $\Box\Diamond\Pr_\infty(*|KB)$ is well-defined as follows: Does there exist an $N > 0$ for which $\#worlds_N^\Phi(KB) > 0$. The unquantified part is clearly recursive, thus proving the upper bound. For the lower bound, we proceed as before. For a given Turing machine $\mathbf{M}$, we let $KB_{\mathbf{M}}$ encode a prefix of the computation of $\mathbf{M}$ on empty input which is a complete prefix currently in an accepting state. Let $KB_{\mathbf{M}}^E$ be the same formula, but with equality replaced by the binary predicate E, as in the proof of Theorem 5.5.1. Let η be the formula forcing E to be an equivalence relation and a congruence on R. The sentence $KB_{\mathbf{M}}^E \wedge \eta$ is satisfiable in infinitely many domain sizes iff it is satisfiable for some domain size iff $\mathbf{M}$ halts. Note that we did not need the additional predicate G in this proof. ▮

We now formally state and prove the theorems asserting that the remaining complexity results do carry over for a language without equality. It is clear that all our upper bounds hold trivially for the language without equality. We consider the lower bounds, one by one.

Theorem B.4.1: *Let Φ be a vocabulary containing at least three non-unary predicate symbols. For sentences $\varphi, KB \in \mathcal{L}^-(\Phi)$, the problem of deciding if $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) exists is Π_3^0-complete. The lower bound holds even if we have an oracle that tells us whether the limit is well-defined.*

Proof: The sentence $KB'_{\mathbf{S}}$ in Theorem 5.4.3 does not satisfy the requirement of Theorem 5.5.1, since $\#worlds_N^\Phi(KB'_{\mathbf{S}}) = N! \cdot b$, where a/b is the is the most recent sequence value generated by $\mathbf{S}$ in the computation so far. The values of b do not necessarily form a non-decreasing sequence. However, it is easy to transform $\mathbf{S}$ to an equivalent Turing machine $\mathbf{S}'$, that outputs the rationals in a non-reduced form satisfying the constraint. Using this transformation, the result follows from Theorem 5.5.1. ▮

Theorem B.4.2: *Let Φ be a vocabulary containing at least three binary predicates, and let $r, r_1, r_2 \in [0,1]$ be rational numbers such that $r_1 \leq r_2$. For sentences $\varphi, KB \in \mathcal{L}^-(\Phi)$, given an oracle for deciding if $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) exists,*

(a) the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB) = r$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB) = r$) is Π_2^0-complete,

(b) if $[r_1, r_2] \neq [0,1]$, then the problem of deciding whether $\Diamond\Box\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB) \in [r_1, r_2]$) is Π_2^0-complete,

(c) if $r_1 \neq r_2$, then the problem of deciding if $\Diamond\Box\Pr_\infty(\varphi|KB) \in (r_1, r_2)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB) \in (r_1, r_2)$) is Σ_2^0-complete.

Proof: It can be verified that the sentences constructed in the proof of Theorem 5.4.4 satisfy the constraints of Theorem 5.5.1. ▮

B.5 Decidability for a finite language

Theorem 5.6.2: *For all d, there exists a Turing machine $\mathbf{M}_d$ such that for all $\varphi, KB \in \mathcal{L}_d(\Phi)$, $\mathbf{M}_d$ decides in time linear in the length of φ and KB whether $\Diamond\Box\Pr_\infty(\varphi|KB)$ (resp., $\Box\Diamond\Pr_\infty(\varphi|KB)$) is well-defined, if so whether it exists, and if it exists computes an arbitrarily good rational approximation to its value.*

Proof: Let $\mathcal{L}_i^d(\Phi)$ consist of all formulas (not necessarily sentences) of quantification depth at most i that mention only the variables $x_1, \ldots, x_d$. Notice that there is an algorithm that runs in linear time that effectively converts a sentence in $\mathcal{L}_d(\Phi)$ to a sentence in $\mathcal{L}_d^d(\Phi)$. We now show that (a) we can effectively find a finite set Σ_i^d of formulas such that every formula in $\mathcal{L}_i^d(\Phi)$ is equivalent to a formula in Σ_i^d, and (b) there is a linear time algorithm that effectively converts a formula in $\mathcal{L}_i^d(\Phi)$ to an equivalent formula in Σ_i^d. This is sufficient to show that any problem—including all those relating to conditional probabilities—whose inputs are formulas in $\mathcal{L}_i^d(\Phi)$ and whose outputs only depend on the semantics of formulas, is solvable in linear time. This is because there exists a constant time algorithm—essentially a lookup table—that, given a formula in Σ_i^d, outputs the correct response. So, given any formula, we can find the equivalent formula in Σ_i^d, and use this algorithm to obtain the appropriate output. Note that we cannot necessarily give an effective construction that produces the lookup table.

We first prove the existence of Σ_i^d for each fixed d by induction on i. For the base case $i = 0$, observe that our assumptions imply that there are only finitely many distinct "literals" consisting of a predicate symbol, followed by the appropriate number of arguments drawn from the constants in Φ and $x_1, \ldots x_d$. (For the purpose of this proof, we treat equality just like any other binary predicate.) Every formula in $\mathcal{L}_0^d(\Phi)$ is a Boolean combination of these literals, and there are only finitely many non-equivalent Boolean combinations of formulas in a finite set. We can effectively construct a set Σ_0^d consisting of one representative of each equivalence class of equivalent formulas. For later ease of exposition, we assume that if the equivalence class includes a literal, then that is the representative chosen to be in Σ_0^d.

For the inductive step, suppose that we have constructed Σ_i^d. Every formula in $\mathcal{L}_{i+1}^d(\Phi)$ is equivalent to a Boolean combination of formulas of the form $Qx_j\,\psi$, where $j \leq d$, ψ has depth at most i, and Q is either $\exists, \forall$, or is absent altogether. By the inductive hypothesis, we can replace ψ by an equivalent formula $\sigma_\psi \in \Sigma_i^d$. Therefore, every formula in $\mathcal{L}_{i+1}^d(\Phi)$ is equivalent to a Boolean combination of formulas of the form $Qx_j\,\sigma_\psi$, where $j \leq d$ and $\sigma_\psi \in \Sigma_i^d$. Since Σ_i^d is finite and $j \leq d$, this is a Boolean combination of formulas in a finite set. Using the fact that there are only finitely many inequivalent Boolean combinations of formulas in a finite set, we can again construct a finite set Σ_{i+1}^d extending Σ_i^d for which the result follows.

To complete the proof, we need to show how to determine the appropriate $\sigma \in \Sigma_i^d$ given a sentence $\xi \in \mathcal{L}_i^d(\Phi)$. We assume that ξ is fully parenthesized. First, it is clear that there exists a constant time algorithm (a lookup table) such that: given a formula of the form $\sigma_1 \wedge \sigma_2$, $\neg\sigma_1$, or $\exists x_j\,\sigma_1$, for $\sigma_1, \sigma_2 \in \Sigma_i^d$, it finds an equivalent formula in Σ_i^d. This is easy to see because, as Σ_i^d is finite, there are only a finite number of possible inputs. The final algorithm is presented in Figure B.2.

It is straightforward to prove by induction that if σ_1 and σ_2 are popped off the stack, then

```
Read $\xi$ from left to right, doing the following:
    If symbol read is a literal, a Boolean connective, or a quantifier then
        Push symbol onto the stack
    If symbol read is a right parenthesis, then
        Pop immediately preceding symbols off the stack, so that:
            We obtain a subformula $\xi$ of the form $\sigma_1 \wedge \sigma_2$, $\neg\sigma_1$, or $\exists x_j\, \sigma_1$
        Find the formula $\sigma \in \Sigma_i^d$ which is equivalent to $\xi$
        Push $\sigma$ onto the stack
Print the contents of the stack.
```

Figure B.2: An algorithm for constructing $\sigma \in \Sigma_i^d$.

they are both in Σ_i^d. The base case follows by our assumption about Σ_i^d containing all literals. The inductive step follows by the construction of the lookup table algorithm. Moreover, the subformula σ pushed onto the stack in the last step in the loop is logically equivalent to the formula it replaces. It follows that after ξ is read, there is exactly one formula on the stack, which is equivalent to ξ.

Given Φ and d, it is easy to construct Σ_i^d and a Turing machine that, for each pair of formulas $\varphi, KB \in \mathcal{L}_i^d(\Phi)$, finds the equivalent formulas $\sigma_\varphi, \sigma_{KB} \in \Sigma_i^d$. Given that, it remains only to construct a lookup table that tells us, for any formulas $\sigma_\varphi, \sigma_{KB} \in \Sigma_i^d$, the behavior of $\Diamond\Box\Pr_\infty(\varphi|KB)$ ($\Box\Diamond\Pr_\infty(\varphi|KB)$). We can easily construct a finite set of linear-time Turing machines, corresponding to the different possible lookup tables. One of these will allow us to correctly determine the behavior of the asymptotic probability (well-definedness, existence, and value of limit). ∎

Appendix C

Proofs for Chapter 6

C.1 Unary expressivity

Theorem 6.2.7: *If ξ is a formula in $\mathcal{L}(\Psi)$ whose free variables are contained in $\mathcal{X}$, and $M \geq d(\xi) + |\mathcal{C}| + |\mathcal{X}|$,[1] then there exists a set of atomic descriptions $\mathcal{A}^{\Psi}_{\xi} \subseteq \mathcal{A}^{\Psi}_{M,\mathcal{X}}$ such that ξ is equivalent to $\bigvee_{\psi \in \mathcal{A}^{\Psi}_{\xi}} \psi$.*

Proof: We proceed by a straightforward induction on the structure of ξ. We assume without loss of generality that ξ is constructed from atomic formulas using only the operators $\wedge$, $\neg$, and $\exists$.

First suppose that ξ is an atomic formula. That is, ξ is either of the form $P(z)$ or of the form $z = z'$, for $z, z' \in \mathcal{C} \cup \mathcal{X}$. In this case, either the formula ξ or its negation appears as a conjunct in each atomic description $\psi \in \mathcal{A}^{\Psi}_{M,\xi}$. Let $\mathcal{A}^{\Psi}_{\xi}$ be those atomic descriptions in which ξ appears as a conjunct. Clearly, ξ is inconsistent with the remaining atomic descriptions. Since the disjunction of the atomic descriptions in $\mathcal{A}^{\Psi}_{M,\mathcal{X}}$ is valid, we obtain that ξ is equivalent to $\bigvee_{\psi \in \mathcal{A}^{\Psi}_{\xi}} \psi$.

If ξ is of the form $\xi_1 \wedge \xi_2$, then by the induction hypothesis, ξ_i is equivalent to the disjunction of a set $\mathcal{A}^{\Psi}_{\xi_i} \subseteq \mathcal{A}^{\Psi}_{M,\mathcal{X}}$, for $i = 1, 2$. Clearly ξ is equivalent to the disjunction of the atomic descriptions in $\mathcal{A}^{\Psi}_{\xi_1} \cap \mathcal{A}^{\Psi}_{\xi_2}$. (Recall that the empty disjunction is equivalent to the formula *false*.)

If ξ is of the form $\neg\xi'$ then, by the induction hypothesis, ξ' is equivalent to the disjunction of the atomic descriptions in $\mathcal{A}^{\Psi}_{\xi'}$. It is easy to see that ξ is the disjunction of the atomic descriptions in $\mathcal{A}^{\Psi}_{\neg\xi'} = \mathcal{A}^{\Psi}_{M,\mathcal{X}} \Leftrightarrow \mathcal{A}^{\Psi}_{\xi'}$.

Finally, we consider the case that ξ is of the form $\exists y\, \xi'$. Recall that $M \geq d(\xi) + |\mathcal{C}| + |\mathcal{X}|$. Since $d(\xi') = d(\xi) \Leftrightarrow 1$, it is also the case that $M \geq d(\xi') + |\mathcal{C}| + |\mathcal{X} \cup \{y\}|$. By the induction hypothesis, ξ' is therefore equivalent to the disjunction of the atomic descriptions in $\mathcal{A}^{\Psi}_{\xi'}$. Clearly ξ is equivalent to $\exists y \vee_{\psi \in \mathcal{A}^{\Psi}_{\xi'}} \psi$, and standard first-order reasoning shows that $\exists y \vee_{\psi \in \mathcal{A}^{\Psi}_{\xi'}} \psi$ is

[1] Recall that $d(\xi)$ denotes the depth of quantifier nesting of ξ. See Definition 5.6.1.

equivalent to $\vee_{\psi\in\mathcal{A}^{\Psi}_{\xi'}}\exists y\,\psi$. Since $\mathcal{A}^{\Psi}_{\xi'}\subseteq\mathcal{A}^{\Psi}_{M,\mathcal{X}\cup\{y\}}$, it suffices to show that for each atomic description $\psi\in\mathcal{A}^{\Psi}_{M,\mathcal{X}\cup\{y\}}$, $\exists y\,\psi$ is equivalent to an atomic description in $\mathcal{A}^{\Psi}_{M,\mathcal{X}}$.

Consider some $\psi\in\mathcal{A}^{\Psi}_{M,\mathcal{X}\cup\{y\}}$; we can clearly pull out of the scope of $\exists y$ all the conjuncts in ψ that do not involve y. It follows that $\exists y\,\psi$ is equivalent to $\psi'\wedge\exists y\,\psi''$, where ψ'' is a conjunction of $A(y)$, where A is an atom over $\mathcal{P}$, and formulas of the form $y=z$ and $y\neq z$. It is easy to see that ψ' is a consistent atomic description over Ψ and $\mathcal{X}$ of size M. To complete the proof, we now show that $\psi'\wedge\exists y\,\psi''$ is equivalent to ψ'. There are two cases to consider. First suppose that ψ'' contains a conjunct of the form $y=z$. Let $\psi''[y/z]$ be the result of replacing all free occurrences of y in ψ'' by z. Standard first-order reasoning (using the fact that $\psi''[y/z]$ has no free occurrences of y) shows that $\psi''[y/z]$ is equivalent to $\exists y\,\psi''[y/z]$, which is equivalent to $\exists y\,\psi''$. Since ψ is a complete atomic description which is consistent with ψ'', it follows that each conjunct of $\psi''[y/z]$ (except $z=z$) must be a conjunct of ψ', so ψ' implies $\psi''[y/z]$. It immediately follows that ψ' is equivalent to $\psi'\wedge\exists y\,\psi''$ in this case. Now suppose that there is no conjunct of the form $y=z$ in ψ''. In this case, $\exists y\,\psi''$ is certainly true if there exists a domain element satisfying atom A different from the denotations of all the symbols in $\mathcal{X}\cup\mathcal{C}$. Notice that ψ implies that there exists such an element, namely, the denotation of y. However, ψ' must already imply the existence of such an element since ψ' must force there to be enough elements satisfying A to guarantee the existence of such an element. (We remark that it is crucial for this last part of the argument that $M\geq|\mathcal{X}|+1+|\mathcal{C}|$.) Thus, we again have that ψ' is equivalent to $\psi'\wedge\exists y\,\psi''$. It follows that $\exists y\,\psi$ is equivalent to a consistent atomic description in $\mathcal{A}^{\Psi}_{M,\mathcal{X}}$, namely ψ', as required. ∎

C.2 A conditional 0-1 law

Proposition C.2.1: *The theory T is complete.*

Proof: The proof is based on a result of Łoś and Vaught [Vau54] which says that any first-order theory with no finite models, such that all of its countable models are isomorphic, is complete. The theory T obviously has no finite models. The fact that all of its countable models are isomorphic follows by a standard "back and forth" argument. That is, let $\mathcal{U}$ and $\mathcal{U}'$ be countable models of T. Without loss of generality, assume that both models have the same domain $\mathcal{D}=\{1,2,3,\ldots\}$. We must find a mapping $F:\mathcal{D}\rightarrow\mathcal{D}$ which is an isomorphism between $\mathcal{U}$ and $\mathcal{U}'$ with respect to Φ.

We first map the named elements in both models to each other, in the appropriate way. Recall that T contains the assertion $\exists x_1,\ldots,x_n\,D_{\mathcal{V}}$. Since $\mathcal{U}\models T$, there must exist domain elements $d_1,\ldots,d_n\in\mathcal{D}$ that satisfy $D_{\mathcal{V}}$ under the model $\mathcal{U}$. Similarly, there must exist corresponding elements $d'_1,\ldots,d'_n\in\mathcal{D}$ that satisfy $D_{\mathcal{V}}$ under the model $\mathcal{U}'$. We define the mapping F so $F(d_i)=d'_i$ for $i=1,\ldots,n$. Since $D_{\mathcal{V}}$ is a complete description over these elements, and the two substructures both satisfy $D_{\mathcal{V}}$, they are necessarily isomorphic.

In the general case, assume we have already defined F over some j elements $\{d_1,d_2,\ldots,d_j\}\in\mathcal{D}$ so that the substructure of $\mathcal{U}$ over $\{d_1,\ldots,d_j\}$ is isomorphic to the substructure of $\mathcal{U}'$ over

$\left\{d'_1, \ldots, d'_j\right\}$, where $d'_i = F(d_i)$ for $i = 1, \ldots, j$. Because both substructures are isomorphic there must be a description D that is satisfied by both. Since we began by creating a mapping between the named elements, we can assume that D extends $D_{\mathcal{V}}$. We would like to extend the mapping F so that it eventually exhausts both domains. We accomplish this by using the even rounds of the construction (the rounds where j is even) to ensure that $\mathcal{U}$ is covered, and the odd rounds to ensure that $\mathcal{U}'$ is covered. More precisely, if j is even, let d be the first element in $\mathcal{D}$ which is not in $\{d_1, \ldots, d_j\}$. There is a model description D' extending D that is satisfied by $d_1, \ldots, d_j, d$ in $\mathcal{U}$. Consider the extension axiom in T asserting that any j elements satisfying D can be extended to $j+1$ elements satisfying D'. Since $\mathcal{U}'$ satisfies this axiom, there exists an element d' in $\mathcal{U}'$ such that $d'_1, \ldots, d'_j, d'$ satisfy D'. We define $F(d) = d'$. It is clear that the substructure of $\mathcal{U}$ over $\{d_1, \ldots, d_j, d\}$ is isomorphic to the substructure of $\mathcal{U}'$ over $\left\{d'_1, \ldots, d'_j, d'\right\}$. If j is odd, we follow the same procedure, except that we find a counterpart to the first domain element (in $\mathcal{U}'$) which does not yet have a pre-image in $\mathcal{U}$. It is is easy to see that the final mapping F is an isomorphism between $\mathcal{U}$ and $\mathcal{U}'$. ▮

Proposition C.2.2: *For any* $\varphi \in \mathcal{L}(\Phi)$, *if* $T \models \varphi$ *then* $\Pr_\infty(\varphi | \psi \wedge \mathcal{V}) = 1$.

Proof: We begin by proving the claim for a sentence $\xi \in T$. By the construction of T, ξ is either $\psi \wedge \exists x_1, \ldots, x_n\, D_{\mathcal{V}}$ or an extension axiom. In the first case, Proposition 6.2.21 trivially implies that $\Pr_\infty(\xi | \psi \wedge \mathcal{V}) = 1$. The proof for the case that ξ is an extension axiom is based on a straightforward combinatorial argument, which we briefly sketch. Recall that one of the conjuncts of ψ is a size description σ. The sentence σ includes two types of conjuncts: those of the form $\exists^m x\, A(x)$ and those of the form $\exists^{\geq M} x\, A(x)$. Let σ' be σ with the conjuncts of the second type removed. Let ψ' be the same as ψ except that σ' replaces σ. It is easy to show that $\Pr_\infty(\exists^{\geq M} x\, A(x) | \psi' \wedge \mathcal{V}) = 1$ for any active atom A, and so $\Pr_\infty(\psi | \psi' \wedge \mathcal{V}) = 1$. Since $\psi \Rightarrow \psi'$, by straightforward probabilistic arguments, it suffices to show that $\Pr_\infty(\xi | \psi' \wedge \mathcal{V}) = 1$.

Suppose ξ is an extension axiom involving D and D', where D is a complete description over $\mathcal{X} = \{x_1, \ldots, x_j\}$ and D' is a description over $\mathcal{X} \cup \{x_{j+1}\}$ that extends D. Fix a domain size N, and some particular j domain elements, $d_1, \ldots, d_j$ that satisfy D. Observe that, since D extends $D_{\mathcal{V}}$, all the named elements are among $d_1, \ldots, d_j$. For a given $d \notin \{d_1, \ldots, d_j\}$, let $B(d)$ denote the event that $d_1, \ldots, d_j, d$ satisfies D', conditioned on $\psi' \wedge \mathcal{V}$. The probability of $B(d)$, given that $d_1, \ldots, d_j$ satisfies D, is typically very small but is bounded away from 0 by some β independent of N. To see this, note that D' is consistent with $\psi \wedge \mathcal{V}$ (because D' is part of an extension axiom) and so there is a consistent way choosing how d is related to $d_1, \ldots, d_j$ so as to satisfy D'. Then observe that the total number of possible ways to choose d's properties (as they relate to $d_1, \ldots, d_j$) is independent of N. Since D extends $D_{\mathcal{V}}$, the model fragment defined over the elements $d_1, \ldots, d_j$ satisfies $\psi' \wedge \mathcal{V}$. (Note that it does not necessarily satisfy ψ, which is why we replaced ψ with ψ'.) Since the properties of an element d and its relation to $d_1, \ldots, d_j$ can be chosen independently of the properties of a different element d', the different events $B(d), B(d'), \ldots$ are all independent. Therefore, the probability that there is no domain element at all that, together with $d_1, .., d_j$, satisfies D' is at most $(1 \Leftrightarrow \beta)^{N-j}$. This bounds the probability of the extension axiom being false, relative to fixed $d_1, \ldots, d_j$. There are exactly $\binom{N}{j}$ ways of choosing j elements, so the probability of the axiom being false anywhere in a model is at most $\binom{N}{j}(1 \Leftrightarrow \beta)^{N-j}$. However, this tends to 0 as N goes to infinity. Therefore, the axiom $\forall x_1, \ldots, x_j\, (D \Rightarrow \exists x_{j+1}\, D')$ has asymptotic probability 1 given $\psi' \wedge \mathcal{V}$, and therefore also given $\psi \wedge \mathcal{V}$.

It remains to deal only with the case of a general formula $\varphi \in \mathcal{L}(\Phi)$ such that $T \models \varphi$. By the Compactness Theorem for first-order logic, if $T \models \varphi$ then there is some finite conjunction of assertions $\xi_1, \ldots, \xi_m \in T$ such that $\wedge_{i=1}^{m} \xi_i \models \varphi$. By the previous case, each such ξ_i has asymptotic probability 1, and therefore so does this finite conjunction. Hence, the asymptotic probability $\Pr_\infty(\varphi | \psi \wedge \mathcal{V})$ is also 1. ∎

C.3 Assigning weights to model descriptions

Lemma 6.3.5: *Let ψ be a consistent atomic description of size $M \geq |\mathcal{C}|$ over Ψ, and let $(\psi \wedge \mathcal{V}) \in \mathcal{M}^{\Phi}(\psi)$.*

(a) If $\alpha(\psi) = 0$ and $N > \nu(\psi)$, then $\#worlds_N^{\Psi}(\psi) = 0$. In particular, this holds for all $N > 2^{|\mathcal{P}|}M$.

(b) If $\alpha(\psi) > 0$, then

$$\#worlds_N^{\Phi}(\psi \wedge \mathcal{V}) \sim \binom{N}{n} a^{N-n} 2^{\sum_{i \geq 2} b_i (N^i - n^i)},$$

where $a = \alpha(\psi)$, $n = \nu(\psi)$, and b_i is the number of predicates of arity i in Φ.

Proof: Suppose that $C_\psi = \langle (f_1, g_1), \ldots, (f_{2^{|\mathcal{P}|}}, g_{2^{|\mathcal{P}|}}) \rangle$ is the characteristic of ψ. Let W be a model of cardinality N, and let N_i be the number of domain elements in W satisfying atom A_i. In this case, we say that the *profile of* W is $\langle N_1, \ldots, N_{2^{|\mathcal{P}|}} \rangle$. Clearly we must have $N_1 + \cdots + N_{2^{|\mathcal{P}|}} = N$. We say that the profile $\langle N_1, \ldots, N_{2^{|\mathcal{P}|}} \rangle$ is *consistent* with C_ψ if $f_i \neq *$ implies that $N_i = f_i$, while $f_i = *$ implies that $N_i \geq M$. Notice that if W is a model of ψ, then the profile of W must be consistent with C_ψ.

For part (a), observe that if $\alpha(\psi) = 0$ and $N > \sum_i f_i$, then there can be no models of cardinality N whose profile is consistent with C_ψ. However, if $\alpha(\psi) = 0$, then $\sum_i f_i$ is precisely $\nu(\psi)$. Hence there can be no models of ψ of cardinality N if $N > \nu(\psi)$. Moreover, since $\nu(\psi) \leq 2^{|\mathcal{P}|}M$, the result holds for any $N > 2^{|\mathcal{P}|}$. This proves part (a).

For part (b), let us first consider how many ways there are of choosing a world satisfying $\psi \wedge \mathcal{V}$ with cardinality N and profile $\langle N_1, \ldots, N_{2^{|\mathcal{P}|}} \rangle$. To do the count, we first choose which elements are to be the named elements in the domain. Clearly, there are $\binom{N}{n}$ ways in which this can be done. Once we choose the named elements, their properties are completely determined by $\mathcal{V}$.

It remains to specify the rest of the properties of the world. Let R be a non-unary predicate of arity $i \geq 2$. To completely describe the behavior of R in a world, we need to specify which of the N^i i-tuples over the domain are in the denotation of R. We have already specified this for those i-tuples all of whose components are named elements. There are n^i such i-tuples. Therefore, we have $N^i \Leftrightarrow n^i$ i-tuples left to specify. Since each subset is a possible denotation, we have $2^{N^i - n^i}$ possibilities for the denotation of R. The overall number of choices for the denotations of all non-unary predicates in the vocabulary is therefore $2^{\sum_{i\geq 2} b_i(N^i - n^i)}$.

It remains only to choose the denotations of the unary predicates for the $N' = N \Leftrightarrow n$ domain elements that are not named. Let $i_1, \ldots, i_a$ be the active atoms in ψ, and let $h_j = N_{i_j} \Leftrightarrow g_{i_j}$ for $j = 1, \ldots, a$. Thus, we need to compute all the ways of partitioning the remaining N' elements so that there are h_j elements satisfying atom A_{i_j}; there are $\binom{N'}{h_1\ h_2\ \ldots\ h_a}$ ways of doing this.

We now need to sum over all possible profiles, i.e., those consistent with $\psi \wedge \mathcal{V}$. If $i_j \in A(\psi)$, then there must be at least M domain elements satisfying A_{i_j}. Therefore $N_{i_j} \geq M$, and $h_j = N_{i_j} \Leftrightarrow g_{i_j} \geq M \Leftrightarrow g_{i_j}$. This is the only constraint on h_j. Thus, it follows that

$$\#worlds_N^{\Phi}(\psi \wedge \mathcal{V}) \sim \sum_{\{h_1,\ldots,h_a:\ h_1+\cdots+h_a=N',\ \forall j\ h_j \geq M - g_{i_j}\}} \binom{N}{n} 2^{\sum_{i\geq 2} b_i(N^i - n^i)} \binom{N'}{h_1\ \ldots\ h_a}.$$

This is equal to

$$\binom{N}{n} 2^{\sum_{i\geq 2} b_i(N^i - n^i)} S$$

for

$$S = \sum_{\{h_1,\ldots,h_a:\ h_1+\cdots+h_a=N',\ \forall j\ h_j \geq M - g_{i_j}\}} \binom{N'}{h_1\ \ldots\ h_a}.$$

It remains to get a good asymptotic estimate for S. Notice that

$$\sum_{\{h_1,\ldots,h_a:\ h_1+\cdots+h_a=N'\}} \binom{N'}{h_1\ \ldots\ h_a} = a^{N'},$$

since the sum can be viewed as describing all possible ways to assign one of a possible atoms to each of N' elements. Our goal is to show that $a^{N'}$ is actually a good approximation for S as well. Clearly $S < a^{N'}$. Let $S_j = \sum_{\{h_1,\ldots,h_a:\ h_j<M, h_1+\cdots+h_a=N'\}} \binom{N'}{h_1\ \ldots\ h_a}$. Straightforward computation shows that

$$\begin{aligned}
S_1 &= \sum_{\{h_1,\ldots,h_a:\ h_1<M,\ h_1+\cdots+h_a=N'\}} \binom{N'}{h_1\ \ldots\ h_a} \\
&= \sum_{h_1=0}^{M-1} \sum_{\{h_2,\ldots,h_a:\ h_2+\cdots+h_a=N'-h_1\}} \binom{N'}{h_1} \binom{N' \Leftrightarrow h_1}{h_2\ \ldots\ h_a} \\
&\leq \sum_{h_1=0}^{M-1} \frac{(N')^{h_1}}{h_1!} (a \Leftrightarrow 1)^{N'-h_1} \\
&< MN^M (a \Leftrightarrow 1)^{N'}.
\end{aligned}$$

Similar arguments show that $S_j < MN^M(a \Leftrightarrow 1)^{N'}$ for all j. It follows that

$$\begin{aligned} S &> \sum_{\{h_1,\ldots,h_a : h_1+\cdots+h_a=N'\}} \binom{N'}{h_1 \; \ldots \; h_a} \Leftrightarrow (S_1 + \cdots + S_a) \\ &> a^{N'} \Leftrightarrow aMN^M(a \Leftrightarrow 1)^{N'} . \end{aligned}$$

Therefore,

$$S \sim a^{N'} = a^{N-n},$$

thus concluding the proof. ∎

Lemma 6.3.7: *Suppose that $KB \in \mathcal{L}(\Psi)$, and $M = d(KB) + |\mathcal{C}_{KB}|$. Then the following conditions are equivalent:*

(a) KB is satisfied in some model of cardinality greater than $2^{|\mathcal{P}|}M$,

(b) $\alpha^{\Psi}(KB) > 0$,

(c) for all $N > 2^{|\mathcal{P}|}M$, KB is satisfiable in some model of cardinality N,

(d) $\Pr_\infty(*|KB)$ is well-defined.

Proof: By definition, KB is satisfiable in some model of cardinality N iff $\#worlds_N^{\Psi}(KB) > 0$. We first show that (a) implies (b). If KB is satisfied in some model of cardinality N greater than $2^{|\mathcal{P}|}M$, then there is some atomic description $\psi \in \mathcal{A}_{KB}^{\Psi}$ such that ψ is satisfied in some model of cardinality N. Using part (a) of Lemma 6.3.5, we deduce that $\alpha(\psi) > 0$ and therefore that $\alpha^{\Psi}(KB) > 0$. That (b) entails (c) can be verified by examining the proof of Lemma 6.3.5. That (c) implies (d) and (d) implies (a) is immediate from the definition of well-definedness. ∎

Theorem 6.3.10: *Let $KB \in \mathcal{L}(\Psi)$ and $\Delta^{\Psi}(KB) = \delta$. Let ψ be an atomic description in $\mathcal{A}_{KB}^{\Psi}$, and let $\psi \wedge \mathcal{V} \in \mathcal{M}^{\Phi}(\psi)$.*

(a) If $\Delta(\psi) < \delta$ then $\Pr_\infty(\psi \wedge \mathcal{V}|KB) = 0$.

(b) If $\Delta(\psi) = \delta$ then $\Pr_\infty(\psi \wedge \mathcal{V}|KB) = 1/|\mathcal{M}^{\Phi}(\mathcal{A}_{KB}^{\Psi,\delta})|$.

Proof: We begin with part (a). Since $\Delta^{\Psi}(KB) = \delta = (a, n)$, there must exist some atomic description $\psi' \in \mathcal{A}_{KB}^{\Psi}$ with $\Delta(\psi') = \delta$. Let $\psi' \wedge \mathcal{V}'$ be some model description in $\mathcal{M}(\psi')$.

$$\begin{aligned} \Pr_N(\psi \wedge \mathcal{V}|KB) &= \frac{\#worlds_N^{\Phi}(\psi \wedge \mathcal{V})}{\#worlds_N^{\Phi}(KB)} \\ &\leq \frac{\#worlds_N^{\Phi}(\psi \wedge \mathcal{V})}{\#worlds_N^{\Phi}(\psi' \wedge \mathcal{V}')} \\ &\sim \frac{\binom{N}{\nu(\psi)}(\alpha(\psi))^{N-\nu(\psi)} 2^{\sum_{i\geq 2} b_i(N^i - \nu(\psi)^i)}}{\binom{N}{n} a^{N-n} 2^{\sum_{i\geq 2} b_i(N^i - n^i)}} \\ &= O(N^{\nu(\psi)-n}(\alpha(\psi)/a)^N). \end{aligned}$$

The last step uses the fact that n and $\nu(\psi)$ can be considered to be constant, and that for any constant k, $\binom{N}{k} \sim N^k/k!$. Since $\Delta(\psi) < \delta = (a, n)$, either $\alpha(\psi) < a$ or $\alpha(\psi) = a$ and $\nu(\psi) < n$. In either case, it is easy to see that $N^{\nu(\psi)-n}(\alpha(\psi)/a)^N$ tends to 0 as $N \to \infty$, giving us our result.

To prove part (b), we first observe that, due to part (a), we can essentially ignore all model descriptions of low degree. That is:

$$\#\mathit{worlds}^{\Phi}_{N}(\mathit{KB}) \sim \sum_{(\psi'\wedge\mathcal{V}')\in\mathcal{M}(\mathcal{A}^{\Psi,\delta}_{\mathit{KB}})} \#\mathit{worlds}^{\Phi}_{N}(\psi'\wedge\mathcal{V}').$$

Therefore,

$$\begin{aligned}
\Pr_N(\psi\wedge\mathcal{V}|\mathit{KB}) &= \frac{\#\mathit{worlds}^{\Phi}_{N}(\psi\wedge\mathcal{V})}{\sum_{(\psi'\wedge\mathcal{V}')\in\mathcal{M}(\mathcal{A}^{\Psi,\delta}_{\mathit{KB}})}\#\mathit{worlds}^{\Phi}_{N}(\psi'\wedge\mathcal{V}')}\\
&\sim \frac{\binom{N}{n}a^{N-n}2^{\sum_{i\geq 2} b_i(N^i-n^i)}}{\sum_{(\psi'\wedge\mathcal{V}')\in\mathcal{M}(\mathcal{A}^{\Psi,\delta}_{\mathit{KB}})}\binom{N}{n}a^{N-n}2^{\sum_{i\geq 2} b_i(N^i-n^i)}}\\
&= \frac{1}{|\mathcal{M}(\mathcal{A}^{\Psi,\delta}_{\mathit{KB}})|},
\end{aligned}$$

as desired. ▮

C.4 Computing the 0-1 probabilities

The proof that `Compute01` is correct is based on the following proposition, which can easily be proved using the same techniques as Proposition C.2.1.

Proposition C.4.1: *If D is a complete description over Φ and $\mathcal{X}$ and $\xi \in \mathcal{L}(\Phi)$ is a formula all of whose free variables are in $\mathcal{X}$, then either $T \models D \Rightarrow \xi$ or $T \models D \Rightarrow \neg\xi$.*

Proof: We know that T has no finite models. By the Löwenheim-Skolem Theorem [End72, page 141], we can, without loss of generality, restrict attention to countably infinite models of T.

Suppose $\mathcal{X} = \{x_1, x_2, \ldots, x_j\}$ and that $T \not\models D \Rightarrow \xi$. Then there is some countable model $\mathcal{U}$ of T, and j domain elements $\{d_1, \ldots, d_j\}$ in the domain of $\mathcal{U}$, which satisfy $D \wedge \neg\xi$. Consider another model $\mathcal{U}'$ of T, and any $\left\{d'_1, \ldots, d'_j\right\}$ in the domain of U' that satisfy D. Because D is a complete description, the substructures over $\{d_1, \ldots, d_j\}$ and $\left\{d'_1, \ldots, d'_j\right\}$ are isomorphic. We can use the back and forth construction of Proposition C.2.1 to extend this to an isomorphism between $\mathcal{U}$ and $\mathcal{U}'$. But then it follows that $\left\{d'_1, \ldots, d'_j\right\}$ must also satisfy $\neg\xi$. Since $\mathcal{U}$ was arbitrary, $T \models D \Rightarrow \neg\xi$. The result follows. ▮

The following result shows that the algorithm above gives a sound and complete procedure for determining whether $T \models D_{\mathcal{V}} \Rightarrow \varphi$.

Theorem C.4.2: *Each of the equivalences used in steps (1)–(5) of* `Compute01` *is true.*

Proof: The equivalences for steps (1)–(3) are easy to show, using Proposition C.4.1. To prove (4), consider some formula $D \Rightarrow \exists y\, \xi'$, where D is a complete description over $x_1, \ldots, x_j$ and the free variables of ξ are contained in $\{x_1, \ldots, x_j\}$. Let $\mathcal{U}$ be some countable model of T, and let $d_1, \ldots, d_j$ be elements in $\mathcal{U}$ that satisfy D. If $\mathcal{U}$ satisfies $D \Rightarrow \exists y\, \xi'$ then there must exist some other element d_{j+1} that, together with $d_1, \ldots, d_j$, satisfies ξ. Consider the description D' over $x_1, \ldots, x_{j+1}$ that extends D and is satisfied by $d_1, \ldots, d_{j+1}$. Clearly $T \not\models D' \Rightarrow \neg\xi'[y/x_{j+1}]$ because this is false in $\mathcal{U}$. So, by Proposition C.4.1, $T \models D' \Rightarrow \xi'[y/x_{j+1}]$ as required.

For the other direction, suppose that $T \models D' \Rightarrow \xi'[y/x_{j+1}]$ for some D' extending D. It follows that $T \models \exists x_{j+1} D' \Rightarrow \exists x_{j+1} \xi'[y/x_{j+1}]$. The result follows from the observation that T contains the extension axiom $\forall x_1, \ldots, x_j (D \Rightarrow \exists x_{j+1} D')$.

The proof for case (5) is similar to case (4), and is omitted. ■

Theorem 6.4.1: *There exists an alternating Turing machine that takes as input a finite vocabulary Φ, a model description $\psi \wedge \mathcal{V}$ over Φ, and a formula $\varphi \in \mathcal{L}(\Phi)$, and decides whether $\Pr_\infty(\varphi | \psi \wedge \mathcal{V})$ is 0 or 1. The machine uses time $O(|\Phi| 2^{|\mathcal{P}|}(\nu(\psi) + |\varphi|)^\rho)$ and $O(|\varphi|)$ alternations, where ρ is the maximum arity of predicates in Φ. If $\rho > 1$, the number of branches is $2^{O(|\Phi|(\nu(\psi)+|\varphi|)^\rho)}$. If $\rho = 1$, the number of branches is $O((2^{|\Phi|} + \nu(\psi))^{|\varphi|})$.*

Proof: `Compute01`, described in Figure 6.2, can easily be implemented on an ATM. Each inductive step corresponding to a disjunction or an existential quantifier can be implemented using a sequence of existential guesses. Similarly, each step corresponding to a conjunction or a universal quantifier can be implemented using a sequence of universal guesses. Note that the number of alternations is at most $|\varphi|$. We must analyze the time and branching complexity of this ATM. Given $\psi \wedge \mathcal{V}$, each computation branch of this ATM can be regarded as doing the following. It:

(a) constructs a complete description D over the variables $x_1, \ldots, x_{n+k}$ that extends $D_{\mathcal{V}}$ and is consistent with ψ, where $n = \nu(\psi)$ and $k \leq |\varphi|/2$ is the number of variables appearing in φ,

(b) chooses a formula ξ or $\neg\xi$, where ξ is an atomic subformula of φ (with free variables renamed appropriately so that they are included in $\{x_1, \ldots, x_{n+k}\}$), and

(c) checks whether $T \models D \Rightarrow \xi$.

Generating a complete description D requires time $|D|$, and if we construct D by adding conjuncts to $D_{\mathcal{V}}$ then it is necessarily the case that D extends $D_{\mathcal{V}}$. To check whether D is consistent with ψ, we must verify that D does not assert the existence of any new element in any finite atom. Under an appropriate representation of ψ (outlined after Corollary 6.4.2 below), this check can be done in time $O(|D| 2^{|\mathcal{P}|})$. Choosing an atomic subformula ξ of φ can takes time

$O(|\varphi|)$. Finally, checking whether $T \models D \Rightarrow \xi$ can be accomplished by simply scanning $|D|$. It is easy to see that we can do this without backtracking over $|D|$. Since $|D| > |\xi|$, it can be done in time $O(|D|)$. Combining all these estimates, we conclude that the length of each branch is $O(|D|2^{|\mathcal{P}|} + |\varphi|)$.

Let D be any complete description over Φ and $\mathcal{X}$. Without loss of generality, we assume that each constant in Φ is equal to (at least) one of the variables in $\mathcal{X}$. To fully describe D we must specify, for each predicate R of arity i, which of the i-tuples of variables used in D satisfy R. Thus, the number of choices needed to specify the denotation of R is bounded by $|\mathcal{X}|^\rho$ where ρ is the maximum arity of a predicate in Φ. Therefore, $|D|$ is $O(|\Phi||\mathcal{X}|^\rho)$. In the case of the description D generated by the algorithm, $\mathcal{X}$ is $\{x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+k}\}$, and $n + k$ is less than $n + |\varphi|$. Thus, the length of such a description D is $O(|\Phi|(n + |\varphi|)^\rho)$.

Using this expression, and our analysis above, we see that the computation time is certainly $O(|\Phi|2^{|\mathcal{P}|}(n + |\varphi|)^\rho)$. In general, the number of branches of the ATM is at most the number of complete descriptions multiplied by the number of atomic formulas in φ. The first of these terms can be exponential in the length of each description. Therefore the number of branches is $O(|\varphi|2^{|\Phi|(n+|\varphi|)^\rho}) = 2^{O(|\Phi|(n+|\varphi|)^\rho)}$. We can, however, get a better bound on the number of branches if all predicates in Φ are unary (i.e., if $\rho = 1$). In this case, ψ already specifies all the properties of the named elements. Therefore, a complete description D is determined when we decide, for each of the at most k variables in D not corresponding to named elements, whether it is equal to a named element and, if not, which atom it satisfies. It follows that there are at most $(2^{|\Phi|} + n)^k$ complete descriptions in this case, and so at most $|\varphi|(2^{|\Phi|} + n)^k$ branches. Since $k \leq |\varphi|/2$, the number of branches is certainly $O((2^{|\Phi|} + n)^{|\varphi|})$ if $\rho = 1$. ▮

Corollary 6.4.2: *There exists a deterministic Turing machine that takes as input a finite vocabulary Φ, a model description $\psi \wedge \mathcal{V}$ over Φ, and a formula $\varphi \in \mathcal{L}(\Phi)$, and decides whether $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$ is 0 or 1. If $\rho > 1$ the machine uses time $2^{O(|\Phi|(\nu(\psi)+|\varphi|)^\rho)}$ and space $O(|\Phi|(\nu(\psi) + |\varphi|)^\rho)$. If $\rho = 1$ the machine uses time $2^{O(|\varphi||\Phi|\log(\nu(\psi)+1))}$ and space $O(|\varphi||\Phi|\log(\nu(\psi) + 1))$.*

Proof: An ATM can be simulated by a deterministic Turing machine which traverses all possible branches of the ATM, while keeping track of the intermediate results necessary to determine whether the ATM accepts or rejects. The time taken by the deterministic simulation is linear in the product of the number of branches of the ATM and the time taken by each branch. The space required is the logarithm of the number of branches plus the space required for each branch. In this case, both these terms are $O(|D| + |\varphi|)$, where D is the description generated by the machine. ▮

C.5 Complexity analysis

Examining `Compute-Pr`$_\infty$, we see that its complexity is dominated by two major quantities: the time required to generate all model descriptions, and the time required to compute the 0-1 probability given each one. The complexity analysis of the latter is given above in Theorem 6.4.1 and Corollary 6.4.2. The following proposition analyzes the length of a model description; the time required to generate all model descriptions is exponential in this length.

Proposition C.5.1: *If $M > |\mathcal{C}|$ then the length of a model description of size M over Φ is*

$$O(|\Phi|(2^{|\mathcal{P}|}M)^{\rho}).$$

Proof: Consider a model description over Φ of size $M = d(KB) + |\mathcal{C}|$. Such a model description consists of two parts: an atomic description ψ over Ψ and a model fragment $\mathcal{V}$ over Φ which is in $\mathcal{M}(\psi)$. To specify an atomic description ψ, we need to specify the unary properties of the named elements; furthermore, for each atom, we need to say whether it has any elements beyond the named elements (i.e., whether it is active). Using this representation, the size of an atomic description ψ is $O(|\Psi|\nu(\psi)+2^{|\mathcal{P}|})$. As we have already observed, the length of a complete description D over Φ and $\mathcal{X}$ is $O(|\Phi||\mathcal{X}|^{\rho})$. In the case of a description $D_{\mathcal{V}}$ for $\mathcal{V} \in \mathcal{M}(\psi)$, this is $O(|\Phi|\nu(\psi)^{\rho})$. Using $\nu(\psi) \leq 2^{|\mathcal{P}|}M$, we obtain the desired result. ∎

C.5.1 Finite vocabulary

Theorem 6.4.3: *Fix a finite vocabulary Φ with at least one unary predicate symbol. For $KB \in \mathcal{L}(\Psi)$, the problem of deciding whether $\Pr_\infty(*|KB)$ is well-defined is PSPACE-complete. The lower bound holds even if $KB \in \mathcal{L}^-(\{P\})$.*

Proof: It follows from Lemma 6.3.7 that $\Pr_\infty(*|KB)$ is well-defined iff $\alpha^\Psi(KB) > 0$. This is true iff there is some atomic description $\psi \in \mathcal{A}^\Psi_{KB}$ such that $\alpha(\psi) > 0$. This holds iff there exists an atomic description ψ of size $M = d(KB) + |\mathcal{C}|$ over Ψ and some model fragment $\mathcal{V} \in \mathcal{M}^\Psi(\psi)$ such that $\alpha(\psi) > 0$ and $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$. Since we are working within Ψ, we can take $\rho = 1$ and $|\mathcal{P}|$ to be a constant, independent of KB. Thus, the length of a model description $\psi \wedge \mathcal{V}$ as given in Proposition C.5.1 is polynomial in $|KB|$. It is therefore possible to generate model descriptions in PSPACE. Using Corollary 6.4.2, we can check, in polynomial space, for a model description $\psi \wedge \mathcal{V}$ whether $\Pr_\infty(KB|\psi \wedge \mathcal{V})$ is 1. Therefore, the entire procedure can be done in polynomial space.

For the lower bound, we use a reduction from the problem of checking the truth of quantified Boolean formulas (QBF), a problem well known to be PSPACE complete [Sto77]. The reduction is similar to that used to show that checking whether a first-order sentence is true in a given finite structure is PSPACE-hard [CM77]. Given a quantified Boolean formula β, we define a first-order sentence $\xi_\beta \in \mathcal{L}^-(\{P\})$ as follows. The structure of ξ_β is identical to that of β, except that any reference to a propositional variable x, except in the quantifier, is replaced by $P(x)$. For example, if β is $\forall x\, \exists y\, (x \wedge y)$, ξ_β will be $\forall x\, \exists y\, (P(x) \wedge P(y))$. Let KB be $\xi_\beta \wedge \exists x\, P(x) \wedge \exists x\, \neg P(x)$. Clearly, $\Pr_\infty(*|KB)$ is well-defined exactly if β is true. ∎

Theorem 6.4.4: *Fix a finite vocabulary Φ. For $\varphi \in \mathcal{L}(\Phi)$ and $KB \in \mathcal{L}(\Psi)$, the problem of computing $\Pr_\infty(\varphi|KB)$ is PSPACE-complete. Indeed, deciding if $\Pr_\infty(\varphi|true) = 1$ is PSPACE-hard even if $\varphi \in \mathcal{L}^-(\{P\})$ for some unary predicate symbol P.*

Proof: The upper bound is obtained directly from `Compute-Pr`$_\infty$ in Figure 6.3. The algorithm generates model descriptions one by one. Using the assumption that Φ is fixed and finite, each model description has polynomial length, so that this can be done in PSPACE. Corollary 6.4.2 implies that, for a fixed finite vocabulary, the 0-1 probabilities for each model description can also be computed in polynomial space. While $\mathit{count}(KB)$ and $\mathit{count}(\varphi)$ can be exponential (as large as the number of model descriptions), only polynomial space is required for their binary representation. Thus, `Compute-Pr`$_\infty$ works in PSPACE under the assumption of a fixed finite vocabulary.

For the lower bound, we provide a reduction from QBF much like that used in Theorem 6.4.3. Given a quantified Boolean formula ξ and a unary predicate symbol P, we construct a sentence $\xi_\beta \in \mathcal{L}^-(\{P\})$ just as in the proof of Theorem 6.4.3. It is easy to see that $\Pr_\infty(\xi_\beta|\mathit{true}) = 1$ iff β is true. (By the unconditional 0-1 law, $\Pr_\infty(\xi_\beta|\mathit{true})$ is necessarily either 0 or 1.) ▮

Theorem 6.4.5: *Fix a finite vocabulary Φ that contains at least two unary predicates and rational numbers $0 \le r_1 \le r_2 \le 1$ such that $[r_1, r_2] \ne [0,1]$. For $\varphi, KB \in \mathcal{L}(\mathcal{P})$, the problem of deciding whether $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ is PSPACE-hard, even given an oracle that tells us whether the limit is well-defined.*

Proof: We first show that, for any rational number r with $0 < r < 1$, we can construct φ_r, KB_r such that $\Pr_\infty(\varphi_r|KB_r) = r$. Suppose $r = q/p$. We assume, without loss of generality, that $\Phi = \{P, Q\}$. Let KB_r be the sentence

$$\exists^{p-1}x\, P(x) \wedge \Big(\exists^{q-1}x\,(P(x)\wedge Q(x)) \vee \exists^{q}x\,(P(x)\wedge Q(x))\Big) \wedge \exists^{0}x\,(\neg P(x)\wedge\neg Q(x)).$$

That is, no elements satisfy the atom $\neg P \wedge \neg Q$, either q or $q \Leftrightarrow 1$ elements satisfy the atom $P \wedge Q$, and $p \Leftrightarrow 1$ elements satisfy P. Thus, there are exactly two atomic descriptions consistent with KB_r. In one of them, ψ_1, there are $q \Leftrightarrow 1$ elements satisfying $P \wedge Q$ and $p \Leftrightarrow q$ elements satisfying $P \wedge \neg Q$ (all the remaining elements satisfy $\neg P \wedge Q$). In the other, ψ_2, there are q elements satisfying $P \wedge Q$ and $p \Leftrightarrow q \Leftrightarrow 1$ elements satisfying $P \wedge \neg Q$. Clearly, the degree of ψ_1 is the same as that of ψ_2, so that neither one dominates. In particular, both define $p \Leftrightarrow 1$ named elements. The number of model fragments for ψ_1 is $\binom{p-1}{q-1} = \frac{(p-1)!}{(q-1)!(p-q)!}$. The number of model fragments for ψ_2 is $\binom{p-1}{q} = \frac{(p-1)!}{q!(p-q-1)!}$. Let φ_r be ψ_1. Clearly

$$\begin{aligned}
\Pr_\infty(\varphi_r|KB_r) &= \frac{|\mathcal{M}(\psi_1)|}{|\mathcal{M}(\psi_1)| + |\mathcal{M}(\psi_2)|} \\
&= \frac{(p \Leftrightarrow 1)!/((q \Leftrightarrow 1)!(p \Leftrightarrow q)!)}{(p \Leftrightarrow 1)!/((q \Leftrightarrow 1)!(p \Leftrightarrow q)!) + (p \Leftrightarrow 1)!/(q!(p \Leftrightarrow q \Leftrightarrow 1)!)} \\
&= \frac{q}{q + (p \Leftrightarrow q)} = \frac{q}{p} = r\ .
\end{aligned}$$

Now, assume we are given $r_1 \le r_2$. We prove the result by reduction from QBF, as in the proof of Theorem 6.4.3. If $r_1 = 0$ then the result follows immediately from Theorem 6.4.4. If $0 < r_1 = q/p$, let β be a QBF, and consider $\Pr_\infty(\xi_\beta \wedge \varphi_{r_1}|KB_{r_1} \wedge \exists x\, \neg P(x))$. Note that, since $p \geq 2$, KB_{r_1} implies $\exists x\, P(x)$. It is therefore easy to see that this probability is 0 if β is false and $\Pr_\infty(\varphi_{r_1}|KB_{r_1}) = r_1$ otherwise. Thus, we can check if β is true by deciding whether $\Pr_\infty(\xi_\beta \wedge \varphi_{r_1}|KB_{r_1} \wedge \exists x\, \neg P(x)) \in [r_1, r_2]$. This proves PSPACE-hardness.[2] ∎

Theorem 6.4.6: *Fix $d \geq 0$. For $\varphi \in \mathcal{L}(\Phi)$, $KB \in \mathcal{L}(\Psi)$ such that $d(\varphi), d(KB) \leq d$, we can effectively construct a linear time algorithm that decides if $\Pr_\infty(\varphi|KB)$ is well-defined and computes it if it is.*

Proof: The proof of the Theorem 5.6.2 shows that if there is a bound d on the quantification depth of formulas and a finite vocabulary, then there is a finite set Σ_i^d of formulas such that every formula ξ of depth at most d is equivalent to a formula in Σ_i^d. Moreover we can construct an algorithm that, given such a formula ξ, will in linear time find some formula equivalent to ξ in Σ_i^d. (We say "some" rather than "the", because it is necessary for the algorithm's constructability that there will generally be several formulas equivalent to ξ in Σ_i^d.) Give this, the problem reduces to constructing a lookup table for the asymptotic conditional probabilities for all formulas in Σ_d. In general, there is no effective technique for constructing this table. However, if we allow conditioning only on unary formulas, it follows from Theorem 6.4.4 that there is. The result now follows. ∎

C.5.2 Infinite vocabulary—restricted cases

The following theorem, due to Lewis, is the key to proving most of the lower bounds in this section.

Theorem C.5.2: [Lew80] *The problem of deciding whether a sentence $\xi \in \mathcal{L}^-(\mathcal{Q})$ is satisfiable is NEXPTIME-complete. Moreover, the lower bound holds even for formulas ξ of depth 2.*

Lewis proves this as follows: For any nondeterministic Turing machine $\mathbf{M}$ that runs in exponential time, and any input w, he constructs a sentence $\xi \in \mathcal{L}^-(\mathcal{Q})$ of quantifier depth 2 and whose length is polynomial in the size of $\mathbf{M}$ and w, such that ξ is satisfiable iff there is an accepting computation of $\mathbf{M}$ on w.

Theorem 6.4.7: *For $KB \in \mathcal{L}(\Upsilon)$, the problem of deciding if $\Pr_\infty(*|KB)$ is well-defined is NEXPTIME-complete. The NEXPTIME lower bound holds even for $KB \in \mathcal{L}^-(\mathcal{Q})$ where $d(KB) \leq 2$.*

Proof: For the upper bound, we proceed much as in Theorem 6.4.3. Let $\Psi = \Upsilon_{KB}$ and let $\mathcal{C} = \mathcal{D}_{KB}$. We know that $\Pr_\infty(*|KB)$ is well-defined iff there exists an atomic description ψ of size $M = d(KB) + |\mathcal{C}|$ over Ψ and some model fragment $\mathcal{V} \in \mathcal{M}^\Psi(\psi)$ such that $\alpha(\psi) > 0$ and $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$. Since all the predicates in Ψ have arity 1, it follows from Proposition C.5.1 that the size of a model description $\psi \wedge \mathcal{V}$ over Ψ is $O(|\Psi|2^{|\mathcal{P}|}M)$. Since $|\Psi| < |KB|$, this implies that model descriptions have exponential length, and can be generated by a nondeterministic

[2] In this construction, it is important to note that although φ_{r_1} and KB_{r_1} can be long sentences, their length depends only on r_1, which is treated as being fixed. Therefore, the constructed asymptotic probability expression does have length polynomial in $|\beta|$. This is also the case in similar constructions later on.

exponential time Turing machine. Because we can assume that $\rho = 1$ here when applying Corollary 6.4.2, we can also deduce that we can check whether $\Pr_\infty(KB|\psi \wedge \mathcal{V})$ is 0 or 1 using a deterministic Turing machine in time $2^{O(|KB||\Psi|\log(\nu(\psi)+1))}$. Since $|\Psi| \leq |KB|$, and $\nu(\psi)$ is at most exponential in $|KB|$, it follows that we can decide if $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$ in deterministic time exponential in $|KB|$. Thus, to check if $\Pr_\infty(*|KB)$ is well-defined we nondeterministically guess a model description $\psi \wedge \mathcal{V}$ of the right type, and check that $\alpha(\psi) > 0$ and that $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$. The entire procedure can be executed in nondeterministic exponential time.

For the lower bound, observe that if a formula ξ in $\mathcal{L}^-(\Phi)$ is satisfied in some model with domain $\{1, \ldots, N\}$ then it is satisfiable in some model of every domain size larger than N. Therefore, $\xi \in \mathcal{L}^-(\mathcal{Q})$ is satisfiable if and only if the limit $\Pr_\infty(*|\xi)$ is well-defined. The result now follows from Theorem C.5.2. ∎

Theorem 6.4.9: *If either (a)* $\varphi, KB \in \mathcal{L}(\Upsilon)$ *or (b)* $\varphi \in \mathcal{L}(\Omega)$ *and* $KB \in \mathcal{L}^-(\Upsilon)$*, then computing* $\Pr_\infty(\varphi|KB)$ *is #EXP-easy.*

Proof: Let $\Phi = \Omega_{\varphi\wedge KB}$, let $\Psi = \Upsilon_{\varphi\wedge KB}$, and let $\mathcal{P}$ and $\mathcal{C}$ be the appropriate subsets of Ψ. Let $\delta_{KB} = \Delta^\Psi(KB)$. Recall from the proof of Theorem 6.4.4 that we would like to generate the model descriptions $\psi \wedge \mathcal{V}$ of degree δ_{KB}, consider the ones for which $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$, and compute the fraction of those for which $\Pr_\infty(\varphi|\psi \wedge \mathcal{V})$. More precisely, consider the set of model descriptions of size $M = d(\varphi \wedge KB) + |\mathcal{C}|$. For a degree δ, let $\mathit{count}^\delta(KB)$ denote the number of those model descriptions for which $\Pr_\infty(KB|\psi \wedge \mathcal{V}) = 1$. Similarly, let $\mathit{count}^\delta(\varphi)$ denote the number for which $\Pr_\infty(\varphi \wedge KB|\psi \wedge \mathcal{V}) = 1$. We are interested in the value of the fraction $\mathit{count}^{\delta_{KB}}(\varphi)/\mathit{count}^{\delta_{KB}}(KB)$.

We want to show that we can nondeterministically generate model descriptions $\psi \wedge \mathcal{V}$, and check in deterministic exponential time whether $\Pr_\infty(KB|\psi\wedge\mathcal{V})$ (or, similarly, $\Pr_\infty(\varphi\wedge KB|\psi\wedge \mathcal{V})$) is 0 or 1. We begin by showing the second part: that the 0-1 probabilities can be computed in deterministic exponential time. There are two cases to consider. In case (a), φ and KB are both unary, allowing us to assume that $\rho = 1$ for the purposes of Corollary 6.4.2. In this case, the 0-1 computations can be done in time $2^{O(|\varphi\wedge KB||\Psi|\log(\nu(\psi)+1))}$, where $\Psi = \Upsilon_{\varphi\wedge KB}$. As in Theorem 6.4.7, $|\Psi| \leq |\varphi\wedge KB|$ and $\nu(\psi)$ is at most exponential in $|KB|$, allowing us to carry out the computation in deterministic exponential time. In case (b), $KB \in \mathcal{L}^-(\Upsilon)$, and therefore the only named elements are the constants. In this case, the 0-1 probabilities can be computed in deterministic time $2^{O(|\Phi|(\nu(\psi)+|\varphi\wedge KB|)^\rho)}$, where $\Phi = \Omega_{\varphi\wedge KB}$. However, as we have just discussed, $\nu(\psi) < |\varphi \wedge KB|$, implying that the computation can be completed in exponential time.

Having shown how the 0-1 probabilities can be computed, it remains only to generate model descriptions in the appropriate way. However, we do not want to consider all model descriptions, because we must count only those model descriptions of degree δ_{KB}. The problem is that we do not know δ_{KB} in advance. We will therefore construct a nondeterministic exponential time Turing machine $\mathbf{M}$ such that the number of accepting paths of $\mathbf{M}$ encodes, for each degree δ, both $\mathit{count}^\delta(\varphi)$ and $\mathit{count}^\delta(KB)$. We need to do the encoding in such a way as to be able to isolate the counts for δ_{KB} when we finally know its value. This is done as follows.

Let ψ be an atomic description ψ over Ψ of size M. Recall that the degree $\Delta(\psi)$ is a pair $(\alpha(\psi), \nu(\psi))$ such that $\alpha(\psi) \leq 2^{|\mathcal{P}|}$ and $\nu(\psi) \leq 2^{|\mathcal{P}|}M$. Thus, there are at most $E = 2^{2|\mathcal{P}|}M$ possible degrees. Number the degrees in increasing order: $\delta_1, \ldots, \delta_E$. We want it to be the case that the number of accepting paths of $\mathbf{M}$ written in binary has the form

$$p_{10} \cdots p_{1m} q_{10} \cdots q_{1m} \cdots p_{E0} \cdots p_{Em} q_{E0} \cdots q_{Em},$$

where $p_{i0} \ldots p_{im}$ is the binary representation of $\mathit{count}^{\delta_i}(\varphi)$ and $q_{i0} \ldots q_{im}$ is the binary representation of $\mathit{count}^{\delta_i}(\mathit{KB})$. We choose m to be sufficiently large so that there is no overlap between the different sections of the output. The largest possible value of an expression of the form $\mathit{count}^{\delta_i}(\mathit{KB})$ is the maximum number of model descriptions of degree δ_i over Φ. This is clearly less than the overall number of model descriptions, which we computed in the beginning of this section.

The machine $\mathbf{M}$ proceeds as follows. Let m be the smallest integer such that 2^m is more than the number of possible model descriptions, which, by Proposition C.5.1 is $2^{O(|\Phi|(2^{|\mathcal{P}|}M)^{\rho})}$. Note that m is exponential and that $\mathbf{M}$ can easily compute m from Φ. $\mathbf{M}$ then nondeterministically chooses a degree δ_i, for $i = 1, \ldots, E$. It then executes $E \Leftrightarrow i$ phases, in each of which it nondeterministically branches $2m$ times. This has the effect of giving this branch a weight of $2^{2m(E-i)}$. It then nondeterministically chooses whether to compute $p_{i0} \ldots p_{im}$ or $q_{i0} \ldots q_{im}$. If the former, it again branches m times, separating the results for $\mathit{count}^{\delta_i}(\varphi)$ from those for $\mathit{count}^{\delta_i}(\mathit{KB})$. In either case, it now nondeterministically generates all model descriptions $\psi \wedge \mathcal{V}$ over Φ. It ignores those for which $\Delta(\psi) \neq \delta_i$. For the remaining model descriptions $\psi \wedge \mathcal{V}$, it computes $\Pr_\infty(\varphi \wedge \mathit{KB} | \psi \wedge \mathcal{V})$ in the first case, and $\Pr_\infty(\mathit{KB} | \psi \wedge \mathcal{V})$ in the latter. This is done in exponential time, using the same technique as in Theorem 6.4.7. The machine accepts precisely when this probability is 1.

This procedure is executable in nondeterministic exponential time, and results in the appropriate number of accepting paths. It is now easy to compute $\Pr_\infty(\varphi|\mathit{KB})$ by finding the largest degree δ for which $\mathit{count}^{\delta}(\mathit{KB}) > 0$, and dividing $\mathit{count}^{\delta}(\varphi)$ by $\mathit{count}^{\delta}(\mathit{KB})$. ∎

We now want to prove the matching lower bound. As in Theorem 6.4.7, we make use of Lewis' NEXPTIME-completeness result. A straightforward modification of Lewis' proof shows that, given w and a nondeterministic exponential time Turing machine $\mathbf{M}$, we can construct a depth 2 formula $\xi \in \mathcal{L}^-(\mathcal{Q})$ such that the number of simplified atomic descriptions over $\mathcal{P}_\xi$ consistent with ξ is exactly the number of accepting computations of $\mathbf{M}$ on w. This allows us to prove the following theorem:

Theorem C.5.3: *Given $\xi \in \mathcal{L}^-(\mathcal{Q})$, counting the number of simplified atomic descriptions over $\mathcal{P}_\xi$ consistent with ξ is #EXP-complete. The lower bound holds even for formulas ξ of depth 2.*

This theorem forms the basis for our own hardness result. Just as for Theorem 6.4.7, we show that the lower bound actually holds for $\varphi, \mathit{KB} \in \mathcal{L}^-(\mathcal{Q})$ of quantifier depth 2.

Theorem 6.4.10: *Given $\varphi, \mathit{KB} \in \mathcal{L}^-(\mathcal{Q})$ of depth at least 2, the problem of computing $\Pr_\infty(\varphi|\mathit{KB})$ is #EXP-hard, even given an oracle for deciding whether the limit exists.*

A_1	A_2	A_3	A_4
$*$	0	$*$	0
$*$	$*$	0	$*$

Figure C.1: Two atomic descriptions with different degrees

Proof: Given $\varphi \in \mathcal{L}^-(\mathcal{Q})$, we reduce the problem of counting the number of simplified atomic descriptions over $\mathcal{P}_\varphi$ consistent with φ to that of computing an appropriate asymptotic probability. Recall that, for the language $\mathcal{L}^-(\mathcal{Q})$, model descriptions are equivalent to simplified atomic descriptions. Therefore, computing an asymptotic conditional probability for this language reduces to counting simplified atomic descriptions of maximal degree. Thus, the major difficulty we need to overcome here is the converse of the difficulty that arose in the upper bound. We now want to count *all* simplified atomic descriptions consistent with φ, while using the asymptotic conditional probability in the most obvious way would only let us count those of maximum degree. For example, the two atomic descriptions whose characteristics are represented in Figure C.1 have different degrees; the first one will thus be ignored by a computation of asymptotic conditional probabilities.

Let $\mathcal{P}$ be $\mathcal{P}_\varphi = \{P_1, \ldots, P_k\}$, and let Q be a new unary predicate not in $\mathcal{P}$. We let $A_1, \ldots, A_K$ for $K = 2^k$ be all the atoms over $\mathcal{P}$, and let $A'_1, \ldots, A'_{2K}$ be all the atoms over $\mathcal{P}' = \mathcal{P} \cup \{Q\}$, such that $A'_i = A_i \wedge Q$ and $A'_{K+i} = A_i \wedge \neg Q$ for $i = 1, \ldots, K$.

We define KB' as follows:

$$KB' =_{\text{def}} \forall x, y \left(\left(Q(x) \wedge \bigwedge_{i=1}^{k} (P_i(x) \Leftrightarrow P_i(y)) \right) \Rightarrow Q(y) \right) .$$

The sentence KB' guarantees that the predicate Q is "constant" on the atoms defined by $\mathcal{P}$. That is, if A_i is an atom over $\mathcal{P}$, it is not possible to have $\exists x\,(A_i(x) \wedge Q(x))$ as well as $\exists x\,(A_i(x) \wedge \neg Q(x))$. Therefore, if ψ is a simplified atomic description over $\mathcal{P}'$ which is consistent with KB', then, for each $i \leq K$, at most one of the atoms A'_i and A'_{K+i} can be active, while the other is necessarily empty. It follows that $\alpha(\psi) \leq K$. Since there are clearly atomic descriptions of activity count K consistent with KB', the atomic descriptions of maximal degree are precisely those for which $\alpha(\psi) = K$. Moreover, if $\alpha(\psi) = K$, then A'_i is active iff A'_{K+i} is not. Two atomic descriptions of maximal degree are represented in Figure C.2. Thus, for each set $I \subseteq \{1, \ldots, K\}$, there is precisely one simplified atomic description ψ consistent with KB' of activity count K where A'_i is active in ψ iff $i \in I$. Therefore, there are exactly 2^K simplified atomic descriptions ψ over $\mathcal{P}'$ consistent with KB' for which $\alpha(\psi) = K$.

Let $KB = KB' \wedge \exists x\, Q(x)$. Notice that all simplified atomic descriptions ψ with $\alpha(\psi) = K$ that are consistent with KB' are also consistent with KB, except for the one where no atom in $A'_1, \ldots, A'_K$ is active. Thus, $|\mathcal{A}^{\mathcal{P}',K}_{KB}| = 2^K \Leftrightarrow 1$. For the purposes of this proof, we call a simplified atomic description ψ over $\mathcal{P}'$ consistent with KB for which $\alpha(\psi) = K$ a *maximal atomic description.* Notice that there is an obvious one-to-one correspondence between consistent

	A_1	A_2	A_3	A_4
$\wedge Q$:	$*$	0	$*$	0
$\wedge \neg Q$:	0	$*$	0	$*$
$\wedge Q$:	$*$	$*$	0	$*$
$\wedge \neg Q$:	0	0	$*$	0

Figure C.2: Two maximal atomic descriptions

simplified atomic descriptions over $\mathcal{P}$ and maximal atomic descriptions over $\mathcal{P}'$. A maximal atomic description where A'_i is active iff $i \in I$ (and A'_{K+i} is active for $i \notin I$) corresponds to the simplified atomic description over $\mathcal{P}$ where A_i is active iff $i \in I$. (For example, the two consistent simplified atomic descriptions over $\{P_1, P_2\}$ in Figure C.1 correspond to the two maximal atomic descriptions over $\{P_1, P_2, Q\}$ in Figure C.2.) In fact, the reason we consider KB rather than KB' is precisely because there is no consistent simplified atomic description over $\mathcal{P}$ which corresponds to the maximal atomic description where no atom in $A'_1, \ldots, A'_K$ is active (since there is no consistent atomic description over $\mathcal{P}$ where none of $A_1, \ldots, A_K$ are active). Thus, we have overcome the hurdle discussed above.

We now define φ_Q; intuitively, φ_Q is φ restricted refer only to elements that satisfy Q. Formally, we define ξ_Q for any formula ξ by induction on the structure of the formula:

- $\xi_Q = \xi$ for any atomic formula ξ,
- $(\neg \xi)_Q = \neg \xi_Q$,
- $(\xi \wedge \xi')_Q = \xi_Q \wedge \xi'_Q$,
- $(\forall y\, \xi(y))_Q = \forall y\, (Q(y) \Rightarrow \xi_Q(y))$.

Note that the size of φ_Q is linear in the size of φ. The one-to-one mapping discussed above from simplified atomic descriptions to maximal atomic descriptions gives us a one-to-one mapping from simplified atomic descriptions over $\mathcal{P}$ consistent with φ to maximal atomic descriptions consistent with $\varphi_Q \wedge \exists x\, Q(x)$. This is true because a model satisfies φ_Q iff the same model restricted to elements satisfying Q satisfies φ. Thus, the number of model descriptions over $\mathcal{P}$ consistent with φ is precisely $|\mathcal{A}^{\mathcal{P}',K}_{\varphi_Q \wedge KB}|$.

From Corollary 6.3.13, it follows that

$$\Pr_\infty(\varphi_Q | KB) = \frac{|\mathcal{A}^{\mathcal{P}',K}_{\varphi_Q \wedge KB}|}{|\mathcal{A}^{\mathcal{P}',K}_{KB}|} = \frac{|\mathcal{A}^{\mathcal{P}}_{\varphi}|}{2^K \Leftrightarrow 1}.$$

Thus, the number of simplified atomic descriptions over $\mathcal{P}$ consistent with φ is $(2^K \Leftrightarrow 1)\Pr_\infty(\varphi_Q | KB)$. This proves the lower bound. ∎

Theorem 6.4.11: *Fix rational numbers $0 \leq r_1 \leq r_2 \leq 1$ such that $[r_1, r_2] \neq [0,1]$. For $\varphi, KB \in \mathcal{L}^-(\mathcal{Q})$ of depth at least 2, the problem of deciding whether $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ is both NEXPTIME-hard and co-NEXPTIME-hard, even given an oracle for deciding whether the limit exists.*

Proof: Let us begin with the case where $r_1 = 0$ and $r_2 < 1$. Consider any $\varphi \in \mathcal{L}^-(\mathcal{Q})$ of depth at least 2, and assume without loss of generality that $\mathcal{P} = \mathcal{P}_\varphi = \{P_1, \ldots, P_k\}$. Choose $Q \notin \mathcal{P}$, and let $\mathcal{P}' = \mathcal{P} \cup \{Q\}$, and ξ be $\forall x (P_1(x) \wedge \ldots \wedge P_k(x) \wedge Q(x))$. We consider $\Pr_\infty(\varphi|\varphi \vee \xi)$. Clearly $\varphi \vee \xi$ is satisfiable, so that this asymptotic probability is well-defined. If φ is unsatisfiable, then $\Pr_\infty(\varphi|\varphi \vee \xi) = 0$. On the other hand, if φ is satisfiable, then $\alpha^{\mathcal{P}}(\varphi) = j > 0$ for some j. It is easy to see that $\alpha^{\mathcal{P}'}(\varphi) = \alpha^{\mathcal{P}'}(\varphi \vee \xi) = 2j$. Moreover, φ and $\varphi \vee \xi$ are consistent with precisely the same simplified atomic descriptions with $2j$ active atoms. This is true since $\alpha^{\mathcal{P}'}(\xi) = 1 < 2j$. It follows that if φ is satisfiable, then $\Pr_\infty(\varphi|\varphi \vee \xi) = 1$.

Thus, we have that $\Pr_\infty(\varphi|\varphi \vee \xi)$ is either 0 or 1, depending on whether or not φ is satisfiable. Thus, $\Pr_\infty(\neg\varphi|\varphi \vee \xi)$ is in $[r_1, r_2]$ iff φ is satisfiable; similarly, $\Pr_\infty(\neg\varphi|\neg\varphi \vee \xi)$ is in $[r_1, r_2]$ iff φ is valid. By Theorem C.5.2, it follows that this approximation problem is both NEXPTIME-hard and co-NEXPTIME-hard.

If $r_1 = q/p > 0$, we construct sentences φ_{r_1} and KB_{r_1} of depth 2 in $\mathcal{L}^-(\mathcal{Q})$ such that $\Pr_\infty(\varphi_{r_1}|KB_{r_1}) = r_1$.[3] Choose $\ell = \lceil \log p \rceil$, and let $\mathcal{P}'' = \{Q_1, \ldots, Q_\ell\}$ be a set of predicates such that $\mathcal{P}'' \cap \mathcal{P}' = \emptyset$. Let $A_1, \ldots, A_{2^\ell}$ be the set of atoms over $\mathcal{P}''$. We define KB_{r_1} to be

$$\exists^1 x\, (A_1(x) \vee A_2(x) \vee \ldots \vee A_p(x)).$$

Similarly, φ_{r_1} is defined as

$$\exists^1 x\, (A_1(x) \vee A_2(x) \vee \ldots \vee A_q(x)).$$

Recall from Section 6.2 that the construct "$\exists^1 x$" can be defined in terms of a formula of quantifier depth 2. There are exactly p atomic descriptions of size 2 of maximal degree consistent with KB_{r_1}; each has one element in one of the atoms $A_1, \ldots, A_p$, no elements in the rest of the atoms among $A_1, \ldots, A_p$, with all the remaining atoms (those among $A_{p+1}, \ldots, A_{2^\ell}$) being active. Among these atomic descriptions, q are also consistent with φ_{r_1}. Therefore, $\Pr_\infty(\varphi_{r_1}|KB_{r_1}) = q/p$. Since the predicates occurring in φ_{r_1}, KB_{r_1} are disjoint from $\mathcal{P}'$, it follows that

$$\Pr_\infty(\varphi \wedge \varphi_{r_1}|(\varphi \vee \xi) \wedge KB_{r_1}) = \Pr_\infty(\varphi|\varphi \vee \xi) \cdot \Pr_\infty(\varphi_{r_1}|KB_{r_1}) = \Pr_\infty(\varphi|\varphi \vee \xi) \cdot r_1.$$

This is equal to r_1 (and hence is in $[r_1, r_2]$) if and only if φ is satisfiable, and is 0 otherwise. ▮

C.5.3 Sentences of depth 1

Theorem 6.4.12: *For a quantifier-free sentence $KB \in \mathcal{L}^-(\mathcal{Q} \cup \{c\})$, the problem of deciding whether $\Pr_\infty(*|KB)$ is well-defined is NP-hard.*

[3] The sentences constructed in Theorem 6.4.5 for the same purpose will not serve our purpose in this theorem, since they used unbounded quantifier depth.

Proof: We prove the result by reduction from the problem of satisfiability of propositional formulas (SAT). Let β be a propositional formula that mentions the primitive propositions $p_1, p_2, \ldots, p_k$. Let $\Psi = \{P_1, \ldots, P_k, c\}$, and let KB_β be the sentence in $\mathcal{L}^-(\Psi)$ that is just like β except that each occurrence of p_i in β is replaced by $P_i(c)$. It is simple to verify that β is satisfiable iff KB_β is satisfiable. Moreover, $\Pr_\infty(*|KB_\beta)$ is well-defined iff KB_β is satisfiable. This proves the result. ∎

Theorem 6.4.13: *For quantifier-free sentences $\varphi, KB \in \mathcal{L}^-(\mathcal{Q} \cup \{c\})$, the problem of computing $\Pr_\infty(\varphi|KB)$ is #P-hard.*

Proof: We prove the result by reduction from the #P-complete problem of counting satisfying truth assignments in propositional logic. Let β be a propositional formula that mentions the primitive propositions $p_1, p_2, \ldots, p_k$. Let φ_β be the sentence in $\mathcal{L}^-(\mathcal{Q} \cup \{c\})$ that results by replacing each occurrence of p_i in β by $P_i(c)$. We take KB to be simply *true*. Let $\Psi = \{P_1, \ldots, P_k, c\}$. Since φ_β and KB have quantifier depth 0 and c is the only constant, we can restrict attention to atomic descriptions of size 1. Of these, there are certainly some atomic descriptions consistent with φ_β (and KB) in which every atom is active. That is, $\alpha^\Psi(\varphi_\beta) = 2^k$. We call an atomic description ψ of size 1 over Ψ *maximal* if $\alpha(\psi) = 2^k$. For any atomic description ψ, the properties of all named elements with respect to Ψ are completely determined by ψ (because Ψ is unary), so there is a unique model description augmenting ψ. Thus, atomic descriptions and their augmenting descriptions coincide. Note that for maximal atomic descriptions, the only named element is the constant c.

Next, note that a maximal atomic description over $\mathcal{Q} \cup \{c\}$ is completely determined once we specify which atom is satisfied by c. Thus, there are exactly 2^k maximal atomic descriptions. A truth assignment for the propositional variables $p_1, \ldots, p_k$ corresponds exactly to an atom over $P_1, \ldots, P_k$. Let s be the number of satisfying truth assignments for β. A truth assignment satisfying β exactly corresponds to an atom containing c in a model of φ_β. Thus, there are exactly s maximal model descriptions consistent with φ_β. By Theorem 6.3.11 and the equivalence of model descriptions and atomic descriptions for this vocabulary, $\Pr_\infty(\varphi_\beta|\mathit{true}) = s/2^k$. Thus, $2^k \Pr_\infty(\varphi_\beta|\mathit{true})$ is the number of satisfying assignments to β, and we have the required reduction. ∎

Theorem 6.4.14: *Fix rational numbers $0 \le r_1 \le r_2 \le 1$ such that $[r_1, r_2] \ne [0,1]$. For quantifier-free sentences $\varphi, KB \in \mathcal{L}^-(\mathcal{Q} \cup \{c\})$, deciding whether $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ is both NP-hard and co-NP-hard.*

Proof: We first prove this theorem under the assumption that $r_1 = 0$ (so that $r_2 < 1$). The proof proceeds by reducing both the satisfiability problem and the unsatisfiability (or validity) problem for propositional logic to the problem of approximating asymptotic probabilities. Let β be an arbitrary propositional formula, containing the primitive propositions $p_1, \ldots, p_k$. Let $q_1, \ldots, q_l$ be new primitive propositions not appearing in β, where $2^l > r_2/(1 \Leftrightarrow r_2)$, so that $2^l/(2^l+1) > r_2$. Let β' be the propositional formula

$$\beta \vee (p_1 \wedge \ldots \wedge p_k \wedge q_1 \wedge \ldots \wedge q_l).$$

Let $\Psi = \{P_1, \ldots, P_k, Q_1, \ldots, Q_l, c\}$ and let $\varphi_\beta, \varphi_{\beta'}$ be the sentences in $\mathcal{L}^-(\Psi)$ corresponding to β, β', constructed as in the proof of Theorem 6.4.13. Clearly β' (and hence $\varphi_{\beta'}$) is satisfiable. Moreover, if β is unsatisfiable, then $\Pr_\infty(\varphi_\beta | \varphi_{\beta'}) = 0$.

Recall from the proof of Theorem 6.4.13 that, for this language, model descriptions reduce to atomic descriptions of size 1 over Ψ. Moreover, there are *maximal* atomic descriptions ψ for which $\alpha(\psi) = 2^{k+l}$. Let s be the number of truth assignments over $p_1, \ldots, p_k$ that satisfy β. The number of truth assignments over $p_1, \ldots, p_k, q_1, \ldots, q_l$ that satisfy φ_β is exactly $s2^l$. On the other hand, the number of such truth assignments that satisfy $\varphi_{\beta'}$ is either $s2^l$ or $s2^l + 1$ (depending on whether assigning *true* to all of $p_1, \ldots, p_k$ satisfies β). The proof of Theorem 6.4.13 shows that the number of maximal atomic descriptions over Ψ consistent with φ_β (resp., $\varphi_{\beta'}$) is precisely the number of satisfying assignments for the respective propositional formula. Therefore, if β is satisfiable, then $\Pr_\infty(\varphi_\beta | \varphi_{\beta'})$ is at least $s\,2^l/(s\,2^l + 1)$. If β is satisfiable then $s \geq 1$, so that $s\,2^l/(s\,2^l + 1) \geq 2^l/(2^l + 1) > r_2$. Thus, $\Pr_\infty(\varphi_\beta | \varphi_{\beta'}) \in [0, r_2]$ iff β is not satisfiable. This gives us co-NP-hardness. As in Theorem 6.4.11, we get NP-hardness by performing the same transformation on $\neg\beta$.

If $r_1 = p/q > 0$, using the techniques of Theorem 6.4.13, we first find quantifier-free sentences φ_{r_1} and KB_{r_1} such that $\Pr_\infty(\varphi_{r_1} | KB_{r_1}) = r_1$, which can be done by first finding propositional formulas with the appropriate number of satisfying assignments. We can now complete the proof using the techniques of Theorem 6.4.5. We omit details here. ∎

Theorem 6.4.15: *For $KB \in \mathcal{L}(\Upsilon)$ of quantifier depth 1, the problem of deciding whether $\Pr_\infty(*|KB)$ is well-defined is in NP.*

Proof: In the following, let $q < |KB|$ be the number of distinct quantified subformulas in KB. We claim that KB has arbitrarily large finite models (and so the asymptotic probability is well-defined) if and only if KB has a model of size at most $|\mathcal{D}_{KB}| + q + 1$, in which at least one domain element is not the denotation of any constant symbol in $\mathcal{D}_{KB}$. To show one direction, suppose that W is a model of KB of size greater than $|\mathcal{D}_{KB}| + q + 1$, and look at all the subformulas of KB. For every existential subformula $\exists x\, \xi$ which is true in W, (resp., every universal subformula $\forall x\, \xi'$ which is false in W) choose an element satisfying ξ (resp., an element falsifying ξ'). Finally, choose one other arbitrary domain element from W that is not the denotation of any constant. It is easy to verify by induction that the submodel of W constructed by considering the chosen elements, together with all denotations of constants, also satisfies KB. For the other direction, let W be a model of size at most $|\mathcal{D}_{KB}| + q + 1$, and let d be a domain element in W that is not denoted by any constant. It is easy to see that we can construct models of KB of any larger size by adding new elements that satisfy exactly the same predicates as d.

The nondeterministic polynomial time algorithm for well-definedness is as follows. First, guess a model of size at most $|\mathcal{D}_{KB}| + q + 1$, in which some element is not denoted by any constant. It only takes polynomial space to write down this many elements, together with the predicates they satisfy and the constants that denote them. It therefore takes only polynomial time to generate such a model. Next, check whether this model satisfies KB. The only difficult part of this procedure is checking the truth of quantified subformulas. For those, we must examine each of the domain elements in turn. However, since the model is small and quantifiers cannot be nested, this can still be done in polynomial time. Therefore, the entire procedure can be completed in nondeterministic polynomial time. ▮

As we said in Section 6.4.5, in order to prove a PSPACE upper bound, we need a polynomial length substitute for a model description. This substitute consists of two parts. The first is a *conjunctive sentence*, and the second is an *abbreviated description.*

Definition C.5.4: A (closed) sentence is defined to be *conjunctive* if it is of the form $\forall x\, \chi(x) \wedge \exists x\, \xi_1(x) \ldots \wedge \exists x\, \xi_h(x) \wedge \sigma$, where $\chi, \xi_1, \ldots, \xi_h, \sigma$ are all quantifier-free. ▮

Note that $\forall x\, \chi(x) \wedge \forall x\, \chi'(x)$ is equivalent to $\forall x\, (\chi(x) \wedge \chi'(x))$. Therefore, in the context of sentences that have the form of such a conjunction, the assumption of a single universally quantified subformula can be made without loss of generality. As we shall see, the active atoms in atomic descriptions of maximal degree that are consistent with a conjunctive unary formula are precisely those atoms that are consistent with the universal subformula. Moreover, the only named elements in such atomic descriptions are the constants. We define an *abbreviated description* so that it describes only the relevant properties of the named elements (the constants). Let $\Phi = \Phi_{\varphi \wedge KB}$, and let $\Psi = \mathcal{P} \cup \mathcal{C}$ be the unary fragment of Φ. An abbreviated description over Φ is a subformula of a complete description over Φ (see Definition 6.2.2) that specifies only those properties of the constants that correspond to atomic subformulas that actually appear in $\varphi \wedge KB$. Since φ and KB are depth one formulas, we can assume in the following that only one variable, say x, appears in φ or KB.

Definition C.5.5: Let $\{\beta_1, \ldots, \beta_k\}$ consist of all sentences of the form $\gamma[x/c]$, where c is some constant in $\mathcal{C}$ and γ is an atomic subformula of $\varphi \wedge KB$ that involves a non-unary predicate (i.e., one in $\Phi \Leftrightarrow \Psi$). An *abbreviated description $\widehat{D}$ for $\varphi \wedge KB$* is a conjunction of the form $D \wedge \bigwedge_{i=1}^{k} \beta_i'$, where D is a complete description over Ψ, and β_i' is either β_i or $\neg\beta_i$. ▮

Example C.5.6: Suppose $\mathcal{C} = \{a, b, c\}$, $\Phi \Leftrightarrow \mathcal{C} = \{P, R\}$, and $\varphi \wedge KB$ is

$$\forall x\ (P(x) \vee R(x, c)) \wedge \exists x\ R(x, x).$$

Then one abbreviated description would be

$$a \neq b \wedge b \neq c \wedge a \neq c \wedge P(a) \wedge \neg P(b) \wedge P(c) \wedge \neg R(a, c) \wedge \neg R(b, c) \wedge R(c, c) \wedge R(a, a) \wedge \neg R(b, b).$$

Note that this abbreviated description is not actually consistent with $\varphi \wedge KB$. ▮

As we show below, abbreviated descriptions are a substitute for model fragments. Moreover, if $\widehat{D}$ is an abbreviated description for $\varphi \wedge KB$, then $|\widehat{D}|$ is polynomial in $|\varphi \wedge KB|$. Thus, by replacing model fragments with abbreviated descriptions, we reduce the space requirements (and also time requirements) of the algorithm considerably. We define an *abbreviated model description* to consist of a conjunction $\theta \wedge \widehat{D}$, where θ is a conjunctive sentence and $\widehat{D}$ is an abbreviated description for $\varphi \wedge KB$. We also require that every atomic subformula of θ be a subformula of $\varphi \wedge KB$. It turns out that abbreviated model descriptions are suitable replacements for model descriptions in our context. Our first result towards establishing this is that an abbreviated model description compactly encodes some atomic description.

Lemma C.5.7: *Let $\theta \wedge \widehat{D}$ be a consistent abbreviated model description for $\varphi \wedge KB$, and consider atomic descriptions of size $|\mathcal{C}|+1$ over Φ consistent with $\theta \wedge \widehat{D}$. Of these, there is a unique atomic description $\psi[\theta, \widehat{D}]$ whose degree is maximal. Moreover, the only named elements in $\psi[\theta, \widehat{D}]$ are constants, and we can compute $\Delta(\psi[\theta, \widehat{D}])$ in PSPACE.*

Proof: Let $\forall x\, \chi(x)$ be the universal subformula of θ Let $\mathbf{A}$ be the set of atoms A such that $A(x)$ is consistent with $\widehat{D} \wedge \chi(x) \wedge \bigwedge_{c \in \mathcal{C}} x \neq c$. Clearly, in all worlds satisfying $\theta \wedge \widehat{D}$, only atoms in $\mathbf{A}$ are active. The atoms outside $\mathbf{A}$ can contain only elements that are denotations of constant symbols. For an atom A, let m_A be the number of distinct denotations of constants satisfying A according to $\widehat{D}$. By Theorem 6.2.7, $\theta \wedge \widehat{D}$ is equivalent to a disjunction of model descriptions of size $|\mathcal{C}|+1$ over Φ. As we said, at most the atoms in $\mathbf{A}$ can be active. The unique maximal atomic description is therefore:

$$\bigwedge_{A \in \mathbf{A}} \exists^{\geq |\mathcal{C}|+1} x\, A(x) \wedge \bigwedge_{A \notin \mathbf{A}} \exists^{m_A} x\, A(x) \wedge D.$$

It is straightforward to check that this atomic description is, in fact, consistent with $\theta \wedge \widehat{D}$. Therefore, this atomic description is $\psi[\theta, \widehat{D}]$—the maximal atomic description consistent with $\theta \wedge \widehat{D}$. are either unquantified or existential. In the first Note that the only named elements in $\psi[\theta, \widehat{D}]$ are, in fact, the constants. It remains only to show how we can compute $\Delta(\psi[\theta, \widehat{D}])$ in PSPACE (and therefore, without generating $\psi[\theta, \widehat{D}]$ itself). The number of named elements $\nu(\psi[\theta, \widehat{D}])$ can easily be derived in polynomial time from $\widehat{D}$. The activity count $\alpha(\psi[\theta, \widehat{D}])$ can be computed by enumerating atoms $A(x)$ and checking their consistency with $\widehat{D} \wedge \chi(x) \wedge \bigwedge_{c \in\, \mathcal{C}} x \neq c$.

We remark that, if the universal subformula $\forall x\ \chi(x)$ does not include any constant symbols, then the set $\mathbf{A}$ is independent of $\widehat{D}$. In particular, the influence of $\widehat{D}$ on the degree is limited to the determination of how many distinct elements the constants denote. We will use this observation later. ∎

We can view $\theta \wedge \widehat{D}$ as representing $\psi[\theta, \widehat{D}] \wedge \widehat{D}$. Although this latter formula is still not a model description, it functions as one. As we show, the conditional probability of φ given $\theta \wedge \widehat{D}$ is either 0 or 1. Moreover, we provide a PSPACE algorithm for deciding which is the case. Then we show how to assign weights to the different abbreviated model descriptions. We begin by assuming that φ is also conjunctive and has the form:

$$\forall x\, \chi(x) \wedge \exists x\, \xi_1(x) \ldots \wedge \exists x\, \xi_h(x) \wedge \sigma.$$

Moreover, let χ' be the universal subformula of θ, and let $\widehat{D}$ be an abbreviated description for $\varphi \wedge KB$.

Lemma C.5.8: *Let φ be conjunctive and let $\theta \wedge \widehat{D}$ be an abbreviated model description for $\varphi \wedge KB$, as discussed above. If $\Pr_\infty(\varphi|\theta \wedge \widehat{D})$ is well-defined, then its value is either 0 or 1. Its value is 1 iff the following conditions hold:*

(A) (1) $\widehat{D} \Rightarrow \bigwedge_{c\in\mathcal{C}} \chi(c)$ is valid.
(2) $(\widehat{D} \wedge (\bigwedge_{c\in\mathcal{C}} x \neq c)) \Rightarrow (\chi'(x) \Rightarrow \chi(x))$ is valid.

(E) For $j = 1,\ldots,h$, either:
(1) $\widehat{D} \Rightarrow \bigvee_{c\in\mathcal{C}} \xi_j(c)$ is valid, or
(2) $\widehat{D} \wedge (\bigwedge_{c\in\mathcal{C}} x \neq c) \wedge \chi'(x) \wedge \xi_j(x)$ is consistent.

(QF) $\widehat{D} \Rightarrow \sigma$ is valid.

Proof: Suppose that $\Pr_\infty(\varphi|\theta \wedge \sigma)$ is well defined. Assume that conditions (A), (E), and (QF) hold. Consider any model of $\theta \wedge \widehat{D}$. By (A1), all the constants satisfy χ. By (A2), the remaining domain elements also satisfy χ, since they all satisfy χ'. Therefore, $\forall x\, \chi(x)$ holds in all models of $\theta \wedge \widehat{D}$. By (QF), σ also holds in all such models. Now consider a subformula of the form $\exists x\, \xi_j(x)$. If (E1) holds, then the formula is always satisfied by some constant. If (E2) holds, then there exists a description $D(x)$ of an element x such that $D(x) \wedge \widehat{D} \Rightarrow \chi'(x) \wedge \xi_j(x) \wedge \bigwedge_{c\in\mathcal{C}} x \neq c$. Since $D(x)$ describes an element which is not a named element, it is easy to show that $\Pr_\infty(\exists x\, D(x)|\forall x\, \chi'(x)) = 1$, and therefore $\Pr_\infty(\varphi|\theta \wedge \widehat{D}) = 1$.

Now, assume that one of (A), (E), or (QF) does not hold. If (A1) does not hold, then $\widehat{D} \Rightarrow \bigwedge_{c\in\mathcal{C}} \chi(c)$ is not valid. Thus, by the construction of $\widehat{D}$, $\widehat{D}$ implies $\neg \bigwedge_{c\in\mathcal{C}} \chi(c)$; in this case, $\Pr_\infty(\varphi|\theta \wedge \widehat{D}) = 0$. Similar reasoning goes through for the case where (QF) is false. If (E) does not hold, then for some ξ_j, neither (E1) nor (E2) holds. Since (E1) does not hold, $\widehat{D}$ is consistent with $\wedge_{c\in\mathcal{C}} \neg\xi_j(c)$. By the definition of $\widehat{D}$, it follows that $\widehat{D} \Rightarrow (\wedge_{c\in\mathcal{C}} \neg\xi_j(c))$ is valid. From (E2), it follows that $(\widehat{D} \wedge (\wedge_{c\in\mathcal{C}} x \neq c \wedge \chi'(x)) \Rightarrow \neg\xi_j(x))$ is valid. Thus, in any model of $\widehat{D} \wedge \theta$, $\forall x \neg\xi_j(x)$ holds. Again, $\Pr_\infty(\varphi|\theta \wedge \widehat{D}) = 0$. Finally, assume that condition (A2) does not hold. In this case, there exists a description $D(x)$ consistent with $\chi'(x)$ but inconsistent with $\chi(x) \wedge (\wedge_{c\in\mathcal{C}} x \neq c)$. As in case (E) above, $\exists x\, D(x)$ holds with probability 1. But $\exists x\, D(x) \Rightarrow \neg\forall x\, \chi(x)$ is valid, proving the desired result. ∎

Finally, we compute the relative weights of different abbreviated model descriptions. We know, using Theorem 6.2.7 and the definition of model description, that an abbreviated model description $\theta \wedge \widehat{D}$ is equivalent to a disjunction of model descriptions of size $|\mathcal{C}|+1$. Moreover, by Lemma C.5.7, the model descriptions of maximal degree in this disjunction are precisely those that augment the atomic description $\psi[\theta, \widehat{D}]$. We therefore define the degree of an abbreviated model description $KB \wedge \widehat{D}$ to be the degree of $\psi[\theta, \widehat{D}]$. As suggested earlier, our strategy for computing asymptotic conditional probabilities will be to consider all abbreviated model descriptions for $\varphi \wedge KB$, and for each of these compute the probability of φ. However, we also need to compute the relative weights of these abbreviated model descriptions. As usual, the abbreviated model descriptions that are not of maximal degree are dominated completely. However, it turns out that even abbreviated model descriptions with the same degree can have different relative weight. We know that (full) model descriptions of equal degree do in fact have equal weight, and so the relative weight of an abbreviated model description is the number of model descriptions consistent with it. Since the atomic description $\psi[\theta, \widehat{D}]$ is completely determined by the abbreviated model description, it remains only to count model fragments. Recall that the only named elements are the constants. To specify a model fragment consistent with an atomic description $\psi[\theta, \widehat{D}]$, it is necessary to decide which of the elements in $\{1, \ldots, n\}$, where $n = \nu(\psi[\theta, \widehat{D}])$, denotes which of the constants and, for each predicate R of arity $r(R) > 1$, which r-tuples of elements in $\{1, \ldots, n\}$ satisfy R. (The denotation of the unary predicates are already specified by $\psi[\theta, \widehat{D}]$.) Thus, the overall number of model fragments consistent with $\psi[\theta, \widehat{D}]$ is

$$H = n!\, 2^{\sum_{R \in \Phi - \Psi} n^{arity(R)}}.$$

An abbreviated description $\widehat{D}$ already specifies some of the decisions to be made. For example, if $\widehat{D}$ contains $R(c_1, c_2)$, there is one less decision to be made about the denotation of R. Recall that $\widehat{D}$ has the form $D \wedge \bigwedge_{i=1}^{k} \beta_i'$. However, it is not necessarily the case that all formulas β_i give independent pieces of information. For example, assume that $\widehat{D}$ contains both $R(c_1, c_1)$ and $R(c_2, c_2)$; if D specifies that $c_1 = c_2$, then this decides only a single value of the denotation of R; if D specifies that $c_1 \neq c_2$, then these conjuncts decide two distinct values. We define the weight $\omega(\widehat{D})$ of $\widehat{D}$ to be the number of distinct properties specified by $\widehat{D}$; this is always less than or equal to k. The number of model descriptions consistent with $\psi[\theta, \widehat{D}] \wedge \widehat{D}$ is

$$n!\, 2^{(\sum_{R \in \Phi - \Psi} n^{arity(R)}) - \omega(\widehat{D})} = \frac{H}{2^{\omega(\widehat{D})}} = \frac{H}{2^k} 2^{k - \omega(\widehat{D})}. \tag{C.1}$$

Note that k depends only on $\varphi \wedge KB$, so is fixed. H depends on n, and so on the degree of the abbreviated description being considered. However, we will not be using this expression to compare the relative weight of descriptions with different degrees, and so H can be regarded as a constant for our purposes.

We are now in a position to prove the two main results of this section. The first is a PSPACE upper bound for the general problem of quantifier depth 1.

Theorem 6.4.16: *For sentences $KB \in \mathcal{L}(\Upsilon)$ and $\varphi \in \mathcal{L}(\Omega)$ of quantifier depth 1, the problem of computing $\Pr_\infty(\varphi|KB)$ is in PSPACE.*

Proof: First, consider the case where we have an abbreviated description $\widehat{D}$ for $\varphi \wedge KB$, and where φ and KB are both conjunctive. In this case, by Lemma C.5.8, the asymptotic probability $\Pr_\infty(\varphi|KB \wedge \widehat{D})$ is either 0 or 1. (Note that we are taking $\theta = KB$ here.) Moreover, an examination of the conditions in the lemma shows that they can easily be verified in PSPACE. We use this observation later.

We now consider the case where φ and KB need not be conjunctive, and assume without loss of generality that all negations are pushed inwards, so that only atomic formulas are negated. A depth 1 formula in this form is a combination, using disjunctions and conjunctions, of subformulas that are universally quantified, existentially quantified, or quantifier free. If we think of each such subformula as being a propositional symbol, we can generate "truth assignments"

```
$\delta \leftarrow (0, 0)$
For each abbreviated description $\widehat{D}$ for $\varphi \wedge KB$ do:
    $W \leftarrow 2^{k-\omega(\widehat{D})}$
    Generate next conjunctive $\theta$
    For all conjunctive $\theta'$ generated up to now:
        Compare $\psi[\theta, \widehat{D}]$ with $\psi[\theta', \widehat{D}]$
    If $\psi[\theta, \widehat{D}]$ is different from previously generated formulas then
        If $\Delta(\theta \wedge \widehat{D}) > \delta$ and $\Pr_\infty(\varphi | \theta \wedge \widehat{D})$ is well-defined then
            $\delta \leftarrow \Delta(\theta \wedge \widehat{D})$
            $count(KB) \leftarrow W$
            $count(\varphi) \leftarrow W \cdot \Pr_\infty(\varphi | \theta \wedge \widehat{D})$
        If $\Delta(\theta \wedge \widehat{D}) > \delta$ and $\Pr_\infty(\varphi | \theta \wedge \widehat{D})$ is well-defined then
            $count(KB) \leftarrow count(KB) + W$
            $count(\varphi) \leftarrow count(\varphi) + W \cdot \Pr_\infty(\varphi | \theta \wedge \widehat{D})$
Output "$\Pr_\infty(\varphi | KB) = count(\varphi)/count(KB)$".
```

Figure C.3: PSPACE algorithm for depth 1 formulas

for these subformulas that make the overall formula true. Each such truth assignment is a conjunction of (possibly negated) subformulas of the original sentence, and can easily be rewritten in the form of a conjunctive sentence. The length of such a sentence is linear in the length of the original sentence. We conclude that any depth 1 formula is equivalent to a disjunction of (possibly exponentially many) conjunctive sentences. Furthermore, these sentences can easily be generated (one by one) in polynomial space.

The algorithm is described in Figure C.3. Generally speaking, it proceeds along the lines of our standard algorithm `Compute-Pr`$_\infty$, presented in Figure 6.3. That is, it keeps track of three things: (1) the highest degree δ of an abbreviated model description consistent with KB found so far, (2) the number $count(KB)$ of model descriptions (not abbreviated model descriptions) of degree δ consistent with KB, and (3) among the model descriptions of degree δ consistent with KB, the number $count(\varphi)$ of descriptions for which the probability of φ is 1. This is done as follows. First, it generates an abbreviated description $\widehat{D}$ for $\varphi \wedge KB$. Then, it considers, one by one, the conjunctive sentences whose disjunction is equivalent to KB. Let θ be such a conjunctive formula. It then verifies that $\Pr_\infty(\varphi | \theta \wedge \widehat{D})$ is well-defined (by Theorem 6.4.15, this can be done in PSPACE) and computes its value if it is. Note that if this probability is well-defined, then $\theta \wedge \widehat{D}$ is consistent; in this case, all (full) model descriptions extending $\theta \wedge \widehat{D}$ are consistent with KB and should be added to the relevant count. Observe that the value of the probability $\Pr_\infty(\varphi | \theta \wedge \widehat{D})$ is necessarily either 0 or 1, even though φ is not conjunctive. We can compute this value by generating the conjunctive sentences constructed from φ in the fashion described above, and checking for each of them whether its probability given $\theta \wedge \widehat{D}$ is 1. If the answer is yes for any of these conjunctive sentences, then it is yes overall. Otherwise, the probability is clearly zero. So we can generate the conjunctive components of φ one by one in PSPACE and then, as we have observed, we can also compute the conditional probability for each component in PSPACE as well. If indeed $\Pr_\infty(\varphi|\theta \wedge \widehat{D}) = 1$, then this is also the case for all unabbreviated model descriptions extending $\theta \wedge \widehat{D}$; again, $\mathit{count}(\mathit{KB})$ should be adjusted accordingly.

This procedure faces a problem that did not arise for Theorem 6.4.4. It is possible that $\theta \wedge \widehat{D}$ and $\theta' \wedge \widehat{D}$ (for two conjunctive sentences θ and θ') generate the same atomic description—i.e., $\psi[\theta, \widehat{D}] = \psi[\theta', \widehat{D}]$. To avoid double-counting, as we consider each abbreviated model description $\theta \wedge \widehat{D}$, we must compare $\psi[\theta, \widehat{D}]$ with $\psi[\theta', \widehat{D}]$ for all abbreviated model descriptions $\theta' \wedge \widehat{D}$ considered earlier. Of course, it is impossible to save these abbreviated model descriptions, since that would take too much space. Rather, we must reconstruct all previous θ', one by one. Note that we can test whether $\psi[\theta, \widehat{D}] = \psi[\theta', \widehat{D}]$ in PSPACE: we consider the atoms one by one, and compare the size of the atom according to both atomic descriptions. Finally, observe that the weight of an abbreviated model description (the number of unabbreviated model descriptions extending it) can be quite large. However, an examination of Equation (C.1) shows that the first factor is the same for all abbreviated model descriptions with the same degree, and can therefore be ignored (because only the relative weights matter). ∎

The major difficulty in improving from PSPACE to #P is that we want to consider only abbreviated model descriptions of maximal degree, without knowing in advance what that maximum degree is. We therefore have to compute the counts for each degree separately. However, there are exponentially many possible degrees; in fact, it is possible that each conjunctive sentence θ generated from KB leads to a different degree, even if $\widehat{D}$ is fixed. To obtain a #P algorithm, we have had to restrict the form of φ and KB. We can eliminate some of the problems by requiring that both φ and KB be conjunctive sentences. However, even this does not suffice. The same conjunctive sentence can generate different degrees, depending on the abbreviated description $\widehat{D}$. For example, the universally quantified subformula can be $\forall x\, ((R(c,c) \wedge \chi(x)) \vee (\neg R(c,c) \wedge \chi'(x)))$, which clearly can behave very differently according to whether $R(c,c)$ is a conjunct of $\widehat{D}$ or not. To deal with such problems, we assume that φ and KB are *simplified conjunctive sentences*, where a simplified conjunctive sentence is a conjunctive sentence in which no constant symbol appears in the scope of a quantifier. One consequence of this, which we noted in the proof of Theorem C.5.7, is that the activity count of maximal descriptions is fixed by KB, and so only $|\mathcal{C}|$ degrees need be considered (corresponding to the number of distinct denotations of constant symbols). This latter term can vary with $\widehat{D}$, but it is easily computed and the number of possible values is small (i.e., polynomial).

Theorem C.5.9: *For simplified conjunctive sentences $\mathit{KB} \in \mathcal{L}(\Upsilon)$ and $\varphi \in \mathcal{L}(\Omega)$, the problem of computing $\Pr_\infty(\varphi|\mathit{KB})$ is #P-easy.*

Proof: Consider the conditions of Lemma C.5.8. By the simple conjunctivity assumption, $\forall x\, \chi(x)$ and $\forall x\, \chi'(x)$ cannot differentiate between constants and the other domain elements. Therefore, checking (A2) is the same as checking that $\chi'(x) \Rightarrow \chi(x)$ is valid; moreover, the truth of (A2) together with the consistency of $\mathit{KB} \wedge \widehat{D}$ implies the truth of (A1) in this case. Similarly, (E2) is equivalent to checking that $\chi'(x) \wedge \xi_j(x)$ is consistent. Moreover, if (E1) is true (some constant satisfies $\xi_j(c)$ and necessarily also $\chi'(c)$), then $\chi'(x) \wedge \xi_j(x)$ is obviously consistent. Therefore, it suffices to check (E2) for each ξ_j. Note that (A2) and (E2) are independent of the choice of $\widehat{D}$; moreover, they can be viewed as propositional satisfiability and validity tests, respectively (by treating each atomic subformula as a primitive proposition). Therefore, both types of tests can be performed using a #P computation.

The Turing machine we construct proceeds as follows. It first branches into 4 subcomputations. The first checks whether $\Pr_\infty(\varphi|KB)$ is well-defined, using Theorem 6.4.15. The second checks condition (A2). The third divides into h subcomputations, one for each subformula $\exists x\, \xi_h(x)$ in φ, and checks condition (E2) for each one. The fourth generates abbreviated descriptions, and generates, for each relevant degree δ, the appropriate counts $\mathit{count}^\delta(\varphi)$ and $\mathit{count}^\delta(KB)$ (defined in Theorem 6.4.16), as outlined below. As in Theorem 6.4.9, the output of the different subcomputations is separated using appropriate branching.

We expand somewhat on the fourth subcomputation. It begins by branching according to the guess of an abbreviated description $\widehat{D}$. Combined with the conjunctive sentence KB, this defines an abbreviated model description. The algorithm now gives the appropriate weight to the abbreviated description generated. As we have observed, abbreviated descriptions do not necessarily have the same degree: the number of named elements can differ, depending on the equality relations between the different constants. However, as we observed, the activity count is necessarily the same in all cases; moreover, it is easy to compute the number of named elements directly from the abbreviated description in deterministic polynomial time. The Turing machine executes this computation for the abbreviated description guessed, and branches accordingly so as to separate the output corresponding to the different degrees. The machine branches less for higher degrees, so that the output corresponding to them is in the less significant digits of the overall output. Finally, even abbreviated descriptions with the same degree do not have the same weight. The machine computes $\omega(\widehat{D})$, and branches $k \Leftrightarrow \omega(\widehat{D})$ times; this has the effect of giving an abbreviated description $\widehat{D}$ a relative weight of $2^{k-\omega(\widehat{D})}$, as required.

Finally, as in the algorithm described in Figure C.3, we then check whether $\Pr_\infty(*|KB \wedge \widehat{D})$ is well-defined, and whether its value is 0 or 1; the first computation goes towards computing $\mathit{count}^\delta(KB)$ and the second towards computing $\mathit{count}^\delta(\varphi)$. The machine branches according to which test it intends to perform, with appropriate extra branching to separate the output of the two computations (as in Theorem 6.4.9). Now observe that if $\Pr_\infty(*|KB)$ is well-defined, then $\Pr_\infty(*|KB \wedge \widehat{D})$ is well-defined if and only if $\widehat{D} \Rightarrow \sigma'$ is valid, where σ' is the quantifier-free part of KB. Similarly, $\Pr_\infty(\varphi|KB \wedge \widehat{D}) = 1$ if and only if $\widehat{D} \Rightarrow \sigma$ is valid, where σ is the quantifier-free part of $\varphi \wedge KB$. Since $\widehat{D}$ contains all subformulas of σ and σ', these tests can be executed in deterministic polynomial time.

The output of this machine can be used to deduce $\Pr_\infty(\varphi|KB)$ as follows. First, the output of the first subcomputation is checked to verify that this asymptotic probability is well-defined. Then, if not all the branches of the second subcomputation are accepting, the value of the asymptotic probability is 0. Similarly, if one of the h tests in subcomputation 2 did not have at least one accepting branch, the probability is also 0. Finally, the machine scans the output, checking the counts for different degrees. It chooses the highest degree for which $count(KB)$ is non-zero, and computes $count(\varphi)/count(KB)$. ∎

C.5.4 Infinite vocabulary—the general case

Theorem 6.4.17: *For $\varphi \in \mathcal{L}(\Omega)$ and $KB \in \mathcal{L}(\Upsilon)$, the function $\Pr_\infty(\varphi|KB)$ is in #TA(EXP,LIN).*

Proof: Let $\Phi = \Omega_{\varphi\wedge KB}$, let $\Psi = \Upsilon_{\varphi\wedge KB}$, and let ρ be the maximum arity of a predicate in Φ. The proof proceeds precisely as in Theorem 6.4.9. We compute, for each degree δ, the values $count^\delta(KB)$ and $count^\delta(\varphi)$. This is done by nondeterministically generating model descriptions $\psi \wedge \mathcal{V}$ over Φ, branching according to the degree of ψ, and computing $\Pr_\infty(\varphi \wedge KB|\psi \wedge \mathcal{V})$ and $\Pr_\infty(KB|\psi \wedge \mathcal{V})$ using a TA(EXP,LIN) Turing machine.

To see that this is possible, recall from Proposition C.5.1 that the length of a model description over Φ is $O(|\Phi|(2^{|\mathcal{P}|}M)^\rho)$. This is exponential in $|\Phi|$ and ρ, both of which are at most $|\varphi \wedge KB|$. Therefore, it is possible to guess a model description in exponential time. Similarly, as we saw in the proof of Theorem 6.4.9, only exponentially many nondeterministic guesses are required to separate the output so that counts corresponding to different degrees do not overlap. These guesses form the initial nondeterministic stage of our TA(EXP,LIN) Turing machine. Note that it is necessary to construct the rest of the Turing machine so that a universal state always follows this initial stage, so that each guess corresponds exactly to one initial existential path; however, this is easy to arrange.

For each model description $\psi \wedge \mathcal{V}$ so generated, we compute $\Pr_\infty(KB|\psi \wedge \mathcal{V})$ or $\Pr_\infty(\varphi \wedge KB|\psi \wedge \mathcal{V})$ as appropriate, accepting if the conditional probability is 1. It follows immediately from Theorem 6.4.1 and the fact that there can only be exponentially many named elements in any model description we generate that this computation is in TA(EXP,LIN). Thus, the problem of computing $\Pr_\infty(\varphi|KB)$ is in #TA(EXP,LIN). ∎

The proof of the lower bounds is lengthy, but can be simplified somewhat by some assumptions about the construction of the TA(EXP,LIN) machines we consider. The main idea is that the existential "guesses" being made in the the initial phase should be clearly distinguished from the rest of the computation. To achieve this, we assume that the Turing machine has an additional *guess tape*, and the initial phase of every computation consists of nondeterministically generating a *guess string* γ which is written on the new tape. The machine then proceeds with a standard alternating computation, but with the possibility of reading the bits on the guess tape.

More precisely, from now on we make the following assumptions about an ATM $\mathbf{M}$. Consider any increasing functions $T(n)$ and $A(n)$ (in essence, these correspond to the time complexity and number of alternations), and let w be an input of size n. We assume:

- $\mathbf{M}$ has two tapes and two heads (one for each tape). Both tapes are one-way infinite to the right.

- The first tape is a work tape, which initially contains only the string w.
- $\mathbf{M}$ has an initial nondeterministic phase, during which its only action is to nondeterministically generate a string γ of zeros and ones, and write this string on the second tape (the guess tape). The string γ is always of length $T(n)$. Moreover, at the end of this phase, the work tape is as in the initial configuration, the guess tape contains only γ, the heads are at the beginning of their respective tapes, and the machine is in a distinguished universal state s_0.
- After the initial phase, the guess tape is never changed.
- After the initial phase, $\mathbf{M}$ takes at most $T(n)$ steps on each branch of its computation tree, and makes exactly $A(n) \Leftrightarrow 1$ alternations before entering a terminal (accepting or rejecting) state.
- The state before entering a terminal state is always an existential state (i.e., $A(n)$ is odd).

Let $\mathbf{M}'$ be any (unrestricted) TA(T,A) machine that "computes" an integer function f. It is easy to construct some $\mathbf{M}$ satisfying the restrictions above that also computes f. The machine $\mathbf{M}$ first generates the guess string γ, and then simulates $\mathbf{M}'$. At each nondeterministic branching point in the initial existential phase of $\mathbf{M}'$, $\mathbf{M}$ uses the next bit of the string γ to dictate which choice to take. Observe that this phase is deterministic (given γ), and can thus be folded into the following universal phase. (Deterministic steps can be viewed as universal steps with a single successor.) If not all the bits in γ are used, $\mathbf{M}$ continues the execution of $\mathbf{M}'$, but checks in parallel that the unused bits of γ are all 0's. If not, $\mathbf{M}$ rejects. It is easy to see that on any input w, $\mathbf{M}$ has the same number of accepting paths as $\mathbf{M}'$, and therefore accepts the same function f. Moreover, $\mathbf{M}$ has the same number of alternations as $\mathbf{M}'$, and at most a constant factor blowup in the running time.[4] This shows that it will be sufficient to prove our hardness results for the class #TA(EXP,LIN) by considering only those machines that satisfy these restrictions. For the remainder of this section we will therefore assume that all ATM's are of this type.

Let $\mathbf{M}$ be such an ATM and let w be an input of size n. We would like to encode the computation of $\mathbf{M}$ on w using a pair of formulas φ_w, KB_w. (Of course, these formulas depend on $\mathbf{M}$ as well, but we suppress this dependence.) Our first theorem shows how to encode part of this computation: Given some appropriate string γ of length $T(n)$, we construct formulas that encode the computation of $\mathbf{M}$ immediately following the initial phase of guessing γ. More precisely, we say that $\mathbf{M}$ *accepts w given* γ if, on input w, the initial existential path during which $\mathbf{M}$ writes γ on the guess tape leads to an accepting node. We construct formulas $\varphi_{w,\gamma}$ and $KB_{w,\gamma}$ such that $\Pr_\infty(\varphi_{w,\gamma}|KB_{w,\gamma})$ is either 0 or 1, and is equal to 1 iff $\mathbf{M}$ accepts w given γ.

We do not immediately want to specify the process of guessing γ, so our initial construction will not commit to this. For a predicate R, let $\varphi[R]$ be a formula that uses the predicate R. Let

[4] For ease of presentation, we can and will (somewhat inaccurately, but harmlessly) ignore this constant factor and say that the time complexity of $\mathbf{M}$ is, in fact, $T(n)$.

ξ be another formula that has the same number of free variables as the arity of R. Then $\varphi[\xi]$ is the formula where every occurrence of R is replaced with the formula ξ, with an appropriate substitution of the arguments of R for the free variables in ξ.

Theorem C.5.10: *Let* $\mathbf{M}$ *be a TA(T,A) machine as above, where* $T(n) = 2^{t(n)}$ *for some polynomial* $t(n)$ *and* $A(n) = O(n)$. *Let* w *be an input string of length* n, *and* $\gamma \in \{0,1\}^{T(n)}$ *be a guess string.*

(a) For a unary predicate R, *there exist formulas* $\varphi_w[R], \xi_\gamma \in \mathcal{L}(\Omega)$ *and* $KB_w \in \mathcal{L}(\Upsilon)$ *such that* $\Pr_\infty(\varphi_w[\xi_\gamma]|KB_w)$ *is 1 iff* $\mathbf{M}$ *accepts* w *given* γ *and is 0 otherwise. Moreover,* φ_w *uses only predicates with arity 2 or less.*

(b) For a binary predicate R, *there exist formulas* $\varphi'_w[R], \xi'_\gamma \in \mathcal{L}(\Omega)$ *such that* $\Pr_\infty(\varphi'_w[\xi'_\gamma]|\mathit{true})$ *is* 1 *iff* $\mathbf{M}$ *accepts* w *given* γ *and is 0 otherwise.*

The formulas $\varphi_w[R]$, KB_w, *and* $\varphi'_w[R]$ *are independent of* γ, *and their length is polynomial in the representation of* $\mathbf{M}$ *and* w. *Moreover, none of the formulas constructed use any constant symbols.*

Proof: Let , be the tape alphabet of $\mathbf{M}$ and let S be the set of states of $\mathbf{M}$. We will identify an instantaneous description (ID) of length ℓ of $\mathbf{M}$ with a string Σ^ℓ for $\Sigma = \Sigma_W \times \Sigma_G$, where Σ_W is , $\cup$(, $\times S$) and Σ_G is $(\{0,1\} \cup (\{0,1\} \times \{h\}))$. We think of the Σ_W component of the ith element in a string as describing the contents of the ith location in the work tape and also, if the tape head is at location i, the state of of the Turing machine. The Σ_G component describes the contents of the ith location in the guess tape (whose alphabet is $\{0,1\}$) and whether the guess tape's head is positioned there. Of course, we consider only strings in which exactly one element in , $\times S$ appears in the first component and exactly one element in $\{0,1\} \times \{h\}$ appears in the second component. Since $\mathbf{M}$ halts within $T(n)$ steps (not counting the guessing process, which we treat separately), we need only deal with ID's of length at most $T(n)$. Without loss of generality, assume all ID's have length exactly $T(n)$. (If necessary we can pad shorter ID's with blanks.)

In both parts of the theorem, ID's are encoded using the properties of domain elements. In both cases, the vocabulary contains predicates whose truth value with respect to certain combinations of domain elements represent ID's. The only difference between parts (a) and (b) is in the precise encoding used. We begin by showing the encoding for part (a).

In part (a), we use the sentence KB_w to define $T(n)$ named elements. This is possible since KB_w is allowed to use equality. Each ID of the machine will be represented using a single domain element. The properties of the ID will be encoded using the relations between the domain element representing it and the named elements. More precisely, assume that the vocabulary has $t(n)$ unary predicates $P_1, \ldots, P_{t(n)}$, and one additional unary predicate P^*. The domain is divided into two parts: the elements satisfying P^* are the named elements used in the process of encoding ID's, while the elements satisfying $\neg P^*$ are used to actually represent ID's.

The formula KB_w asserts (using equality) that each of the atoms A over $\{P^*, P_1, \ldots, P_{t(n)}\}$ in which P^* (as opposed to $\neg P^*$) is one of the conjuncts contains precisely one element:

$$\forall x, y\left(\left(P^*(x) \wedge P^*(y) \wedge \bigwedge_{i=1}^{t(n)} (P_i(x) \Leftrightarrow P_i(y))\right) \Rightarrow x = y\right).$$

Note that KB_w has polynomial length and is independent of γ.

We can view an atom A over $\{P^*, P_1, \ldots, P_{t(n)}\}$ in which P^* is one of the conjuncts as encoding a number between 0 and $T(n) \Leftrightarrow 1$, written in binary: if A contains P_j rather than $\neg P_j$, then the jth bit of the encoded number is 1, otherwise it is 0. (Recall that $T(n)$, the running time of $\mathbf{M}$, is $2^{t(n)}$.) In the following, we let A_i, for $i = 0, \ldots, T(n) \Leftrightarrow 1$, denote the atom corresponding to the number i according to this scheme. Let e_i be the unique element in the atom A_i for $i = 0, \ldots, T(n) \Leftrightarrow 1$. When representing an ID using a domain element d (where $\neg P^*(d)$), the relation between d and e_i is used to represent the ith coordinate in the ID represented by d. Assume that the vocabulary has a binary predicate R_σ for each $\sigma \in \Sigma$. Roughly speaking, we say that the domain element d represents the ID $\sigma_0 \ldots \sigma_{T(n)-1}$ if $R_{\sigma_i}(d, e_i)$ holds for $i = 0, \ldots, T(n) \Leftrightarrow 1$. More precisely, we say that d *represents* $\sigma_0 \ldots \sigma_{T(n)-1}$ if

$$\neg P^*(d) \wedge \bigwedge_{i=0}^{T(n)-1} \forall y\left(A_i(y) \Rightarrow \left(R_{\sigma_i}(d, y) \wedge \bigwedge_{\sigma' \in \Sigma - \{\sigma_i\}} \neg R_{\sigma'}(d, y)\right)\right).$$

Note that not every domain element d such that $\neg P^*(d)$ holds encodes a valid ID. However, the question of which ID, if any, is encoded by a domain element d depends only on the relations between d and the finite set of elements $e_0, \ldots, e_{T(n)-1}$. This implies that, with asymptotic probability 1, every ID will be encoded by some domain element. More precisely, let $ID(x) = \sigma_0 \ldots \sigma_{T(n)-1}$ be a formula which is true if x denotes an element that represents $\sigma_0 \ldots \sigma_{T(n)-1}$. (It should be clear that such a formula is indeed expressible in our language.) Then for each ID $\sigma_0 \ldots \sigma_{T(n)-1}$ we have

$$\Pr_\infty(\exists x\, (ID(x) = \sigma_0 \ldots \sigma_{T(n)-1}) | KB_w) = 1.$$

For part (b) of the theorem, we must represent ID's in a different way because we are not allowed to condition on formulas that use equality. Therefore, we cannot create an exponential number of named elements using a polynomial-sized formula. The encoding we use in this case uses two domain elements per ID rather than one. We now assume that the vocabulary Ω contains a $t(n)$-ary predicate R_σ for each symbol $\sigma \in \Sigma$. Note that this uses the assumption that there is no bound on the arity of predicates in Ω. For $i = 0, \ldots, T(n) \Leftrightarrow 1$, let $b^i_{t(n)} \ldots b^i_1$ be the binary encoding of i. We say that the pair (d_0, d_1) of domain element *represents the ID* $\sigma_0 \ldots \sigma_{T(n)-1}$ if

$$d_0 \neq d_1 \wedge \bigwedge_{i=0}^{T(n)-1}\left(R_{\sigma_i}(d_{b^i_1}, \ldots, d_{b^i_{t(n)}}) \wedge \bigwedge_{\sigma' \in \Sigma - \{\sigma_i\}} \neg R_{\sigma'}(d_{b^i_1}, \ldots, d_{b^i_{t(n)}})\right).$$

Again, we can define a formula in our language $\mathit{ID}(x_0, x_1) = \sigma_0 \ldots \sigma_{T(n)-1}$ which is true if x_0, x_1 denote a pair of elements that represent $\sigma_0 \ldots \sigma_{T(n)-1}$. As before, observe that for each ID $\sigma_0 \ldots \sigma_{T(n)-1}$ we have

$$\Pr_\infty(\exists x_0, x_1\, (\mathit{ID}(x_0, x_1) = \sigma_0 \ldots \sigma_{T(n)-1}) | \mathit{true}) = 1.$$

In both case (a) and case (b), we can construct formulas polynomial in the size of $\mathbf{M}$ and w that assert certain properties. For example, in case (a), $\mathit{Rep}(x)$ is true of a domain element d if and only if d encodes an ID. In this case, $\mathit{Rep}(x)$ is the formula

$$\neg P^*(x) \wedge \forall y \left(P^*(y) \Rightarrow \dot{\bigvee} R_\sigma(x, y)\right) \wedge$$
$$\exists! y\, (P^*(y) \wedge \textstyle\bigvee_{\sigma \in ((\Gamma \times S) \times \Sigma_G)} R_\sigma(x, y)) \wedge \exists! y\, (P^*(y) \wedge \bigvee_{\sigma \in (\Sigma_W \times (\{0,1\} \times \{h\}))} R_\sigma(x, y))$$

where $\dot{\bigvee}$ is an abbreviation whose meaning is that precisely one of its disjuncts is true.

In case (b), $\mathit{Rep}(x_0, x_1)$ is true of a pair (d_0, d_1) if and only if it encodes an ID. The construction is similar. For instance, the conjunct of $\mathit{Rep}(x_0, x_1)$ asserting that each tape position has a uniquely defined content is

$$x_0 \neq x_1 \wedge \forall z_1, \ldots, z_{t(n)} \left(\left(\bigwedge_{i=1}^{t(n)} (z_i = x_0 \vee z_i = x_1) \right) \Rightarrow \dot{\bigvee} R_\sigma(z_1, \ldots, z_{t(n)}) \right).$$

Except for this assertion, the construction for the two cases is completely parallel given the encoding of ID's. We will therefore restrict the remainder of the discussion to case (a). Other relevant properties of an ID that we can formulate are:

- $\mathit{Acc}(x)$ (resp., $\mathit{Univ}(x)$, $\mathit{Exis}(x)$) is true of a domain element d if and only if d encodes an ID and the state in $\mathit{ID}(d)$ is an accepting state (resp., a universal state, an existential state).

- $\mathit{Step}(x, x')$ is true of elements d and d' if and only if both d and d' encode ID's and $\mathit{ID}(d')$ can follow from $\mathit{ID}(d)$ in one step of $\mathbf{M}$.

- $\mathit{Comp}(x, x')$ is true of elements d and d' if and only if both d and d' encode ID's and $\mathit{ID}(d')$ is the final ID in a maximal non-alternating path starting at $\mathit{ID}(d)$ in the computation tree of $\mathbf{M}$, and the length of this path is at most $T(n)$. A maximal non-alternating path is either a path all of whose states are existential except for the last one (which must be universal or accepting), or a path all of whose states are universal except for the last one. We can construct Comp using a divide and conquer argument, so that its length is polynomial in $t(n)$.

We remark that Acc, Step, etc. are not new predicate symbols in the language. Rather, they are complex formulas described in terms of the basic predicates R_σ. We omit details of their construction here; these can be found in [Gra83].

It remains only to describe the formula that encodes the initial configuration of $\mathbf{M}$ on input w. Since we are interested in the behavior of $\mathbf{M}$ given a particular guess string γ, we begin by encoding the computation of $\mathbf{M}$ after the initial nondeterministic phase; that is, after the string γ is already written on the guess tape and the rest of the machine is back in its original state. We now construct the formula $\mathit{Init}[R](x)$ that describes the initial configuration. This formula takes R as a parameter, and has the form $\mathit{Init}'(x) \wedge R(x)$. The formulas substituted for $R(x)$ will correspond (in a way discussed below) to possible guesses γ.

We begin by considering case (a). We assume the existence of an additional binary predicate B_0. It is easy to write a polynomial-length formula $\mathit{Init}'(x)$ which is true of a domain element d if and only if d represents an ID where: (a) the state is the distinguished state s_0 entered after the nondeterministic guessing phase, (b) the work tape contains only w, (c) the heads are at the beginning of their respective tapes, and (d) for all i, the ith location of the guess tape contains 0 iff $B_0(d, e_i)$. Here e_i is, as before, the unique element in atom A_i. Note that the last constraint can be represented polynomially using the formula

$$\forall y\, (P^*(y) \Rightarrow (B_0(x,y) \Leftrightarrow \bigvee_{\sigma \in \Sigma_W \times \{0,(0,h)\}} R_\sigma(x,y))).$$

We also want to find a formula ξ_γ that can constrain B_0 to reflect the guess γ. This formula, which serves as a possible instantiation for R, does not have to be of polynomial size. We define it as follows, where for convenience, we use B_1 as an abbreviation for $\neg B_0$:

$$\xi_\gamma(x) =_{\mathrm{def}} \bigwedge_{i=0}^{T(n)-1} \forall y\ (A_i(y) \Rightarrow B_{\gamma_i}(x,y))\ . \tag{C.2}$$

Note that this is of exponential length.

In case (b), the relation of the guess string γ to the initial configuration is essentially the same modulo the modifications necessary due to the different representation of ID's. We only sketch the construction. As in case (a), we add a predicate B_0', but in this case of arity $t(n)$. Again, the predicate B_0' represents the locations of the 0's in the guess tape following the initial nondeterministic phase. The specification of the denotation of this predicate is done using an exponential-sized formula ξ_γ', as follows (again taking B_1' to be an abbreviation for $\neg B_0'$):

$$\xi_\gamma'(x_0, x_1) =_{\mathrm{def}} B'_{\gamma_0}(x_0, \ldots, x_0, x_0) \wedge B'_{\gamma_1}(x_0, \ldots, x_0, x_1) \wedge \ldots \wedge B'_{\gamma_{T(n)-1}}(x_1, \ldots, x_1, x_1).$$

Using these formulas, we can now write a formula expressing the assertion that $\mathbf{M}$ accepts w given γ. In writing these formulas, we make use of the assumptions made above $\mathbf{M}$ (that it is initially in the state immediately following the initial guessing phase, that all computation paths make exactly $A(n)$ alternations, and so on). The formula $\varphi_w[R]$ has the following form:

$$\begin{gathered}\exists x_1\, (\mathit{Init}[R](x_1) \wedge \forall x_2\, (\mathit{Comp}(x_1,x_2) \Rightarrow \exists x_3\, (\mathit{Comp}(x_2,x_3) \wedge (\forall x_4\, \mathit{Comp}(x_3,x_4) \Rightarrow \ldots \\ \exists x_{A(n)}\, (\mathit{Comp}(x_{A(n)-1}, x_{A(n)}) \wedge \mathit{Acc}(x_{A(n)})) \ldots))).\end{gathered}$$

It is clear from the construction that $\varphi_w[R]$ does not depend on γ and that its length is polynomial in the representations of $\mathbf{M}$ and w.

Now suppose W is a world satisfying KB_w in which every possible ID is represented by at least one domain element. (As we remarked above, a random world has this property with asymptotic probability 1.) Then it is straightforward to verify that $\varphi_w[\xi_\gamma]$ is true in W iff $\mathbf{M}$ accepts w. Therefore $\Pr_\infty(\varphi_w[\xi_\gamma]|KB_w) = 1$ iff $\mathbf{M}$ accepts w given γ and 0 otherwise. Similarly, in case (b), we have shown the construction of analogous formulas $\varphi'_w[R]$, for a binary predicate R, and ξ'_γ such that $\Pr_\infty(\varphi'_w[\xi'_\gamma]|\mathit{true}) = 1$ iff $\mathbf{M}$ accepts w given γ, and is 0 otherwise. ∎

We can now use the above theorem in order to prove the #TA(EXP,LIN) lower bound.

Theorem 6.4.18: *For $\varphi \in \mathcal{L}(\Omega)$ and $KB \in \mathcal{L}(\Upsilon)$, computing $\Pr_\infty(\varphi|KB)$ is #TA(EXP,LIN)-hard. The lower bound holds even if φ, KB do not mention constant symbols and either (a) φ uses no predicate of arity > 2, or (b) KB uses no equality.*

Proof: Let $\mathbf{M}$ be a TA(EXP,LIN) Turing machine of the restricted type discussed earlier, and let w be an input of size n. We would like to construct formulas φ, KB such that from $\Pr_\infty(\varphi|KB)$ we can derive the number of accepting computations of $\mathbf{M}$ on w. The number of accepting initial existential paths of such a Turing machine is precisely the number of guess strings γ such that $\mathbf{M}$ accepts w given γ. In Theorem C.5.10, we showed how to encode the computation of such a machine $\mathbf{M}$ on input w given a nondeterministic guess γ. We now show how to force an asymptotic conditional probability to count guess strings γ appropriately.

As in Theorem C.5.10, let $T(n) = 2^{t(n)}$, and let $\mathcal{P}' = \{P'_1, \ldots, P'_{t(n)}\}$ be new unary predicates, not used in the construction of Theorem C.5.10. As before, we can view an atom A' over $\mathcal{P}'$ as representing a number in the range $0, \ldots, T(n) \Leftrightarrow 1$: if A contains P'_j, then the jth bit of the encoded number is 1, otherwise it is 0. Again, let A'_i, for $i = 0, \ldots, T(n) \Leftrightarrow 1$, denote the atom corresponding to the number i according to this scheme. We can view a simplified atomic description ψ over $\mathcal{P}'$ as representing the string $\gamma = \gamma_0 \ldots \gamma_{T(n)-1}$ such that γ_i is 1 if ψ contains the conjunct $\exists z\, A'_i(z)$, and 0 if ψ contains its negation. Under this representation, for every string γ of length $T(n)$, there is a unique simplified atomic description over $\mathcal{P}'$ that represents it; we denote this atomic description ψ_γ. Note that ψ_γ is not necessarily a consistent atomic description, since the atomic description where all atoms are empty also denotes a legal string—that string where all bits are 0.

While it is possible to reduce the problem of counting accepting guess strings γ to that of counting simplified atomic descriptions ψ_γ, this is not enough. After all, we have already seen that computing asymptotic conditional probabilities ignores all atomic descriptions that are not of maximal degree. We deal with this problem as in Theorem 6.4.10. Let Q be a new unary predicate, and let KB' be, as in Theorem 6.4.10, the sentence

$$\forall x, y \left(\left(Q(x) \wedge \bigwedge_{j=1}^{t(n)} (P'_j(x) \Leftrightarrow P'_j(y)) \right) \Rightarrow Q(y) \right) .$$

Observe that here we use KB' rather than the formula KB of Theorem 6.4.10, since we also want to count the “inconsistent” atomic description where all atoms are empty. Recall that, assuming KB', each simplified atomic description ψ_γ over $\mathcal{P}'$ corresponds precisely to a single maximal atomic description ψ'_γ over $\mathcal{P}' \cup \{Q\}$. We reduce the problem of counting accepting guess strings γ to that of counting simplified atomic description ψ'_γ.

We now consider cases (a) and (b) separately, beginning with the former. Fix a guess string γ. In Theorem C.5.10, we constructed formulas $\varphi_w[R], \xi_\gamma \in \mathcal{L}(\Omega)$ and $KB_w \in \mathcal{L}(\Upsilon)$ such that $\Pr_\infty(\varphi_w[\xi_\gamma]|KB_w) = 1$ iff $\mathbf{M}$ accepts w given γ, and is 0 otherwise. Recall that the formula $\xi_\gamma(x)$ (see Equation (C.2)) sets the ith guess bit to be γ_i by forcing the appropriate one of $B_0(x, e_i)$ and $B_1(x, e_i)$ to hold, where e_i is the unique element in the atom A_i. In Theorem C.5.10, this was done directly by reference to the bits γ_i. Now, we want to derive the correct bit values from ψ_γ, which tells us that the ith bit is 1 iff $\exists z\, A'_i(z)$. The following formula ξ has precisely the desired property:

$$\xi(x) =_{\rm def} \forall y \left(P^*(y) \Rightarrow \left(B_1(x,y) \Leftrightarrow \exists z\,(Q(z) \wedge \bigwedge_{j=1}^{t(n)} (P_j(y) \Leftrightarrow P'_j(z)))\right)\right).$$

Clearly, $\psi'_\gamma \models \xi \Leftrightarrow \xi_\gamma$.

Similarly, for case (b), the formula ξ' is:

$$\begin{array}{rcl} \xi'(x_0, x_1) & =_{\rm def} & \forall y_1, \ldots, y_{t(n)} \left((\bigwedge_{j=1}^{t(n)} (y_j = x_0 \vee y_j = x_1)) \Rightarrow \right. \\ & & \left. \left(B'_1(y_1, \ldots, y_{t(n)}) \Leftrightarrow \exists z\,(Q(z) \wedge \bigwedge_{j=1}^{t(n)} (y_j = x_1 \Leftrightarrow P'_j(z)))\right)\right). \end{array}$$

As in part (a), $\psi'_\gamma \models \xi' \Leftrightarrow \xi'_\gamma$.

Now, for case (a), we want to compute the asymptotic conditional probability $\Pr_\infty(\varphi[\xi]|KB_w \wedge KB')$. In doing this computation, we will use the observation (whose straightforward proof we leave to the reader) that if the symbols that appear in KB_2 are disjoint from those that appear in φ_1 and KB_1, then $\Pr_\infty(\varphi_1|KB_1 \wedge KB_2) = \Pr_\infty(\varphi_1|KB_1)$. Using this observation and the fact that all maximal atomic descriptions over $\mathcal{P}' \cup \{Q\}$ are equally likely given $KB_w \wedge KB'$, by straightforward probabilistic reasoning we obtain:

$$\begin{array}{rcl} \Pr_\infty(\varphi_w[\xi]|KB_w \wedge KB') & = & \sum_{\psi'_\gamma} \Pr_\infty(\varphi_w[\xi]|KB_w \wedge KB' \wedge \psi'_\gamma) \cdot \Pr_\infty(\psi'_\gamma|KB_w \wedge KB') \\ & = & \frac{1}{2^{T(n)}} \sum_{\psi'_\gamma} \Pr_\infty(\varphi_w[\xi]|KB_w \wedge KB' \wedge \psi'_\gamma). \end{array}$$

We observed before that ξ is equivalent to ξ_γ in worlds satisfying ψ'_γ, and therefore

$$\Pr_\infty(\varphi_w[\xi]|KB_w \wedge KB' \wedge \psi'_\gamma) = \Pr_\infty(\varphi_w[\xi_\gamma]|KB_w \wedge KB' \wedge \psi'_\gamma) = \Pr_\infty(\varphi_w[\xi_\gamma]|KB_w),$$

where the second equality follows from the observation that none of the vocabulary symbols in ψ'_γ or KB' appear anywhere in $\varphi_w[\xi_\gamma]$ or in KB_w. In Theorem C.5.10, we proved that

$\Pr_\infty(\varphi_w[\xi_\gamma]|KB_w)$ is equal to 1 if the ATM accepts w given γ and 0 if not. We therefore obtain that

$$\Pr_\infty(\varphi_w[\xi]|KB_w \wedge KB') = \frac{f(w)}{2^{T(n)}}.$$

Since both $\varphi_w[\xi]$ and $KB_w \wedge KB'$ are polynomial in the size of the representation of $\mathbf{M}$ and in $n = |w|$, this concludes the proof for part (a). The completion of the proof for part (b) is essentially identical. ▮

It remains only to investigate the problem of approximating $\Pr_\infty(\varphi|KB)$ for this language.

Theorem C.5.11: *Fix rational numbers $0 \leq r_1 \leq r_2 \leq 1$ such that $[r_1, r_2] \neq [0, 1]$. For $\varphi, KB \in \mathcal{L}(\Omega)$, the problem of deciding whether $\Pr_\infty(\varphi|KB) \in [r_1, r_2]$ is TA(EXP,LIN)-hard, even given an oracle for deciding whether the limit exists.*

Proof: For the case of $r_1 = 0$ and $r_2 < 1$, the result is an easy corollary of Theorem C.5.10. We can generalize this to the case of $r_1 > 0$, using precisely the same technique as in Theorem 6.4.11. ▮

Appendix D

Proofs for Chapter 7

D.1 Unary Expressivity

Theorem 7.1.5: *Every formula in $\mathcal{L}_1^{=}$ is equivalent to a formula in canonical form. Moreover, there is an effective procedure that, given a formula $\xi \in \mathcal{L}_1^{=}$ constructs an equivalent formula $\widehat{\xi}$ in canonical form.*

Proof: We show how to effectively transform $\xi \in \mathcal{L}_1^{=}$ to an equivalent formula in canonical form. We first rename variables if necessary, so that all variables used in ξ are distinct (i.e., no two quantifiers, including proportion expressions, ever bind the same variable symbol).

We next transform ξ into an equivalent *flat* formula $\xi_f \in \mathcal{L}_1^{\approx}$, where a flat formula is one where no quantifiers (including proportion quantifiers) have within their scope any constant or variable other than the variable(s) the quantifier itself binds. (Note that in this transformation we do not require that ξ be closed. Also, observe that flatness implies that there are no nested quantifiers.)

We define the transformation by induction on the structure of ξ:

- If ξ is an unquantified formulas, then $\xi_f = \xi$.
- $(\xi' \vee \xi'')_f = \xi'_f \vee \xi''_f$
- $(\neg \xi')_f = \neg(\xi_f)$.

All that remains is to consider quantified formulas of the form $\exists x\, \xi'$, $||\xi'||_{\vec{x}}$, or $||\xi'|\xi''||_{\vec{x}}$. It turns out that the same transformation works in all three cases. We illustrate the transformation by looking at the case where ξ is of the form $||\xi'||_{\vec{x}}$. By the inductive hypothesis, we can assume that ξ' is flat. For the purposes of this proof, we define a *basic* formula to be an atomic formula (i.e., one of the form $P(z)$), a proportion formula, or a quantified formula (i.e., one of the form $\exists x\, \chi$). Let $\chi_1, \ldots, \chi_k$ be all basic subformulas of ξ' that do not mention any variable in $\vec{x}$. Let z be a variable or constant symbol not in $\vec{x}$ that is mentioned in ξ'. Clearly z must occur in some basic subformula of ξ', say χ'. By the inductive hypothesis, it is easy to see that χ'

cannot mention any variable in $\vec{x}$ and so, by construction, it is in $\{\chi_1, \ldots, \chi_\ell\}$. In other words, not only do $\{\chi_1, \ldots, \chi_\ell\}$ not mention any variable in $\vec{x}$, but they also contain all occurrences of the other variables and constants. (Notice that this argument fails if the language includes any high arity predicates, including equality. For then ξ' might include subformulas of the form $R(x, y)$ or $x = y$, which can mix variables outside $\vec{x}$ with those in $\vec{x}$.)

Now, let $B_1, \ldots, B_{2^\ell}$ be all the "atoms" over $\chi_1, \ldots, \chi_\ell$. That is, we consider all formulas $\chi'_1 \wedge \ldots \chi'_\ell$ where χ'_i is either χ_i or $\neg\chi_i$. Now consider the disjunction:

$$\bigvee_{i=1}^{2^\ell} (B_i \wedge ||\xi'||_{\vec{x}}).$$

This is surely equivalent to $||\xi'||_{\vec{x}}$, because some B_i must be true. However, if we assume that any particular B_i is true, we can simplify $||\xi'||_{\vec{x}}$ by replacing all the χ_i subformulas by *true* or *false*, according to B_i. (Note that this is allowed only because the χ_i do not mention any variable in $\vec{x}$). The result is that we can simplify each disjunct $(B_i \wedge ||\xi'||_{\vec{x}})$ considerably. In fact, because of our previous observation about $\{\chi_1, \ldots, \chi_\ell\}$, there will be no constants or variables outside $\vec{x}$ left within the proportion quantifier. This completes this step of the induction. Since the other quantifiers can be treated similarly, this proves the flatness result.

It now remains to show how a flat formula can be transformed to canonical form. Suppose $\xi \in \mathcal{L}_1^{\approx}$ is flat. Let $\xi^* \in \mathcal{L}_1^{=}$ be the formula equivalent to ξ obtained by using the translation of Section 3.1. Every proportion comparison in ξ^* is of the form $t \leq t'\varepsilon_i$ where t and t' are polynomials over flat unconditional proportions. In fact, t' is simply a product of flat unconditional proportions (where the empty product is taken to be 1). Note also that since we cleared away conditional proportions by multiplying by t', if $t' = 0$ then so is t, and so the formula $t \leq t'\varepsilon_i$ is automatically true. We can therefore replace the comparison by $(t' = 0) \vee (t \leq t'\varepsilon_i \wedge t' > 0)$. Similarly, we can replace a negated comparison by an expression of the form $\neg(t \leq t'\varepsilon_i) \wedge t' > 0$.

The next step is to rewrite all the flat unconditional proportions in terms of atomic proportions. In any such proportion $||\xi'||_{\vec{x}}$, the formula ξ' is a Boolean combination of $P(x_i)$ for predicates $P \in \mathcal{P}$ and $x_i \in \vec{x}$. Thus, the formula ξ' is equivalent to a disjunction $\bigvee_j (A_1^j(x_{i_1}) \wedge \ldots \wedge A_m^j(x_{i_m}))$, where each A_i^j is an atom over $\mathcal{P}$, and $\vec{x} = \{x_{i_1}, \ldots, x_{i_m}\}$. These disjuncts are mutually exclusive, and the semantics treats distinct variables as being independent, so

$$||\xi'||_{\vec{x}} = \sum_j \prod_{i=1}^{m} ||A_i^j(x)||_x.$$

We perform this replacement for each proportion expression. Furthermore, any term t' in an expression of the form $t \leq t'\varepsilon_i$ will be a product of such expressions, and so will be positive.

Next, we must put all pure first-order formulas in the right form. We first rewrite ξ to push all negations inwards as far as possible, so that only atomic subformulas and existential formulas are negated. Next, note that since ξ is flat, each existential subformula must have the form $\exists x\, \xi'$, where ξ' is a quantifier-free formula which mentions no constants and only the variable x. Hence, ξ' is a Boolean combination of $P(x)$ for predicates $P \in \mathcal{P}$. The formula ξ' is easily seen to be equivalent to a disjunction of atoms of the form $\bigvee_{A \in \mathcal{A}(\xi)} A(x)$, so $\exists x\, \xi'$ is equivalent to $\bigvee_{A \in \mathcal{A}(\xi)} \exists x\, A(x)$. We replace $\exists x\, \xi'$ by this expression. Finally, we must deal with formulas of the form $P(c)$ or $\neg P(c)$ for $P \in \mathcal{P}$. This is easy: we can again replace a formula ξ of the form $P(c)$ or $\neg P(c)$ by the disjunction $\bigvee_{A \in \mathcal{A}(\xi)} A(c)$.

The penultimate step is to convert ξ into disjunctive normal form. This essentially brings things into canonical form. Note that since we dealt with formulas of the form $\neg P(c)$ in the previous step, we do not have to deal with conjuncts of the form $\neg A_i(c)$.

The final step is to check that we do not have $A_i(c)$ and either $\neg\exists x\, A_i(x)$ or $A_j(c)$ for some $j \neq i$ as conjuncts of some disjunct. If we do, we remove that disjunct. ▮

D.2 The Concentration Phenomenon

Lemma 7.1.11: *There exist some function $h : I\!N \to I\!N$ and two strictly positive polynomial functions $f, g : I\!N \to I\!R$ such that, for $KB \in \mathcal{L}_1^{\approx}$ and $\vec{u} \in \Delta^K$, if $\#worlds_N^{\vec{\tau}}[\vec{u}](KB) \neq 0$, then in fact*

$$(h(N)/f(N))e^{NH(\vec{u})} \leq \#worlds_N^{\vec{\tau}}[\vec{u}](KB) \leq h(N)g(N)e^{NH(\vec{u})}.$$

Proof: To choose a world $W \in \mathcal{W}_N$ satisfying KB such that $\pi(W) = \vec{u}$, we must partition the domain among the atoms according to the proportions in $\vec{u}$, and then choose an assignment for the constants in the language, subject to the constraints imposed by KB. Suppose $\vec{u} = (u_1, \ldots, u_K)$, and let $N_i = u_i N$ for $i = 1, \ldots, K$. The number of partitions of the domain into atoms is $\binom{N}{N_1,\ldots,N_K}$; each such partition completely determines the denotation for the unary predicates. We must also specify the denotations of the constant symbols. There are at most $N^{|\mathcal{C}|}$ ways of choosing these. On the other hand, we know there is at least one model $(W, \vec{\tau})$ of KB such that $\pi(W) = \vec{u}$, so there there at least one choice. In fact, there is at least one world $W' \in \mathcal{W}_N$ such that $(W', \vec{\tau}) \models KB$ for each of the $\binom{N}{N_1,\ldots,N_K}$ ways of partitioning the elements of the domain (and each such world W' is isomorphic to W). Finally we must choose the denotation of the non-unary predicates. However, $\vec{u}$ does not constrain this choice and, by assumption, neither does KB. Therefore the number of such choices is some function $h(N)$ which is independent of $\vec{u}$.[1] We conclude that:

$$h(N)\binom{N}{N_1,\ldots,N_K} \leq \#worlds_N^{\vec{\tau}}[\vec{u}](KB) \leq h(N)N^{|\mathcal{C}|}\binom{N}{N_1,\ldots,N_K}.$$

It remains to estimate

$$\binom{N}{N_1,\ldots,N_K} = \frac{N!}{N_1!N_2!\ldots N_K!}.$$

[1] It is easy to verify that

$$h(N) = \prod_{R \in \Phi - \Psi} 2^{N^{arity(R)}},$$

where Ψ is the unary fragment of Φ and $arity(R)$ denotes the arity of the predicate symbol R.

To obtain our result, we use Stirling's approximation for the factorials, which says that

$$m! = \sqrt{2\pi m}\, m^m e^{-m}(1 + O(1/m)).$$

It follows that exist constants $L, U > 0$ such that

$$L\, m^m e^{-m} \le m! \le U m\, m^m e^{-m}$$

for all m. Using these bounds, as well as the fact that $N_i \le N$, we get:

$$\frac{L}{U^K N^K}\frac{N^N \prod_{i=1}^K e^{N_i}}{e^N \prod_{i=1}^K N_i^{N_i}} \le \frac{N!}{N_1!N_2!\ldots N_K!} \le \frac{UN}{L^K}\frac{N^N \prod_{i=1}^K e^{N_i}}{e^N \prod_{i=1}^K N_i^{N_i}}.$$

Now, consider the expression common to both bounds:

$$\begin{aligned}
\frac{N^N \prod_{i=1}^K e^{N_i}}{e^N \prod_{i=1}^K N_i^{N_i}} &= \frac{N^N}{\prod_{i=1}^K N_i^{N_i}} \\
&= \prod_{i=1}^K \left(\frac{N}{N_i}\right)^{N_i} \\
&= \prod_{i=1}^K e^{N_i \ln(N/N_i)} \\
&= e^{-N\sum_{i=1}^K u_i \ln(u_i)} = e^{NH(\vec{u})}.
\end{aligned}$$

We obtain that

$$\frac{h(N)L}{U^K N^K} e^{NH(\vec{u})} \le \#worlds_N^{\vec{\tau}}[\vec{u}](KB) \le N^{|\mathcal{C}|} h(N)\frac{UN}{L^K} e^{NH(\vec{u})},$$

which is the desired result. ▮

We next want to prove Theorem 7.1.13. To do this, it is useful to have an alternative representation of the solution space $S^{\vec{\tau}}[KB]$. Towards this end, we have the following definition.

Definition D.2.1: Let $\Pi_N^{\vec{\tau}}[KB] = \{\pi(W) \in \mathcal{W}_N : (W, \vec{\tau}) \models KB\}$. Let $\Pi_\infty^{\vec{\tau}}[KB]$ be the limit of these spaces; formally,

$$\Pi_\infty^{\vec{\tau}}[KB] = \{\vec{u} \;:\; \exists N_0 \; s.t.\; \forall N \ge N_0 \; \exists \vec{u}^N \in \Pi_N^{\vec{\tau}}[KB] \; s.t. \lim_{N\to\infty} \vec{u}^N = \vec{u}\}. \quad ▮$$

The following theorem establishes a tight connection between $S^{\vec{\tau}}[KB]$ and $\Pi_\infty^{\vec{\tau}}[KB]$.

Theorem D.2.2:

(a) For all N and $\vec{\tau}$, $\Pi_N^{\vec{\tau}}[KB] \subseteq S^{\vec{\tau}}[KB]$. That is, for any $\vec{\tau}$ and W, if $(W, \vec{\tau}) \models KB$ then $\pi(W) \in S^{\vec{\tau}}[KB]$.

(b) For all sufficiently small $\vec{\tau}$, $\Pi_\infty^{\vec{\tau}}[KB] = S^{\vec{\tau}}[KB]$.

Proof: Part (a) is immediate: For any $\vec{\tau}$ and any N, consider a world W such that $(W, \vec{\tau}) \models KB$. It is almost immediate from the definitions that $\pi(W)$ must satisfy , $(KB[\vec{\tau}])$, so $\pi(W) \in$ *Sol*[, $(KB[\vec{\tau}])$]. The inclusion $\Pi_N^{\vec{\tau}}[KB] \subseteq S^{\vec{\tau}}[KB]$ now follows by definition.

One direction of part (b) follows immediately from part (a). Recall that $\Pi_N^{\vec{\tau}}[KB] \subseteq S^{\vec{\tau}}[KB]$ and that the points in $\Pi_\infty^{\vec{\tau}}[KB]$ are limits of a sequence of points in $\Pi_N^{\vec{\tau}}[KB]$. Since $S^{\vec{\tau}}[KB]$ is closed, it follows that $\Pi_\infty^{\vec{\tau}}[KB] \subseteq S^{\vec{\tau}}[KB]$.

For the opposite inclusion, the general strategy of the proof is to show the following:

(i) If $\vec{\tau}$ is sufficiently small, then for all $\vec{u} \in S^{\vec{\tau}}[KB]$, there is some sequence of points $\left\{\vec{u}^{N_0}, \vec{u}^{N_0+1}, \vec{u}^{N_0+2}, \vec{u}^{N_0+3}, \ldots\right\} \subset$ *Sol*[, $(KB[\vec{\tau}])$] such that, for all $N \geq N_0$, the coordinates of $\vec{u}^N$ are all integer multiples of $1/N$ and $\lim_{N \to \infty} \vec{u}^N = \vec{u}$.

(ii) if $\vec{w} \in$ *Sol*[, $(KB[\vec{\tau}])$] and all its coordinates are integer multiples of $1/N$, then $\vec{w} \in \Pi_N^{\vec{\tau}}[KB]$.

This clearly suffices to prove that $\vec{u} \in \Pi_\infty^{\vec{\tau}}[KB]$.

The proof of (ii) is straightforward. Suppose $\vec{w} = (r_1/N, r_2/N, \ldots, r_K/N)$ is in *Sol*[, $(KB[\vec{\tau}])$]. We construct a world $W \in \mathcal{W}_N$ such that $\pi(W) = \vec{w}$ as follows. The denotation of atom A_1 is the set of elements $\{1, \ldots, r_1\}$, the denotation of atom A_2 is the set $\{r_1 + 1, \ldots, r_1 + r_2\}$, and so on. It remains to choose the denotations of the constants (since the denotation of the predicates of arity greater than 1 is irrelevant). Without loss of generality we can assume KB is in canonical form. (If not, we consider $\widehat{KB}$.) Thus, KB is a disjunction of conjunctions, say $\bigvee_j \xi_j$. Since $\vec{w} \in$ *Sol*[, $(KB[\vec{\tau}])$], we must have $\vec{w} \in$ *Sol*[, $(\xi_j[\vec{\tau}])$] for some j. We use ξ_j to define the properties of the constants. If ξ_j contains $A_i(c)$ for some atom A_i, then we make c satisfy A_i. Note that, by Definition 7.1.6, if ξ_j has such a conjunct then $u_i > 0$. If ξ_j contains no atomic conjunct for a constant c then we make c satisfy A_i for some arbitrary atom with $u_i > 0$. Note that this construction is possible precisely because we never assign a constant to an empty atom (with $u_i = 0$). It should now be clear that $(W, \vec{\tau})$ satisfies ξ_j, and so satisfies KB. (Note that this is not the case for an arbitrary point $\vec{w}$ in $S^{\vec{\tau}}[KB]$, since this space is the closure of the actual solution space, so that the points in it do not necessarily satisfy , $(KB[\vec{\tau}])$.)

The proof of (i) is surprisingly difficult, and involves techniques from algebraic geometry. Our job would be relatively easy if *Sol*[, $(KB[\vec{\tau}])$] were an open set. Unfortunately, it is not. On the other hand, it would be an open set if we could replace the occurrences of $\leq$ in , $(KB[\vec{\tau}])$ by $<$. It turns out that we can essentially do this.

Let , $^<(KB[\vec{\tau}])$ be the same as , $(KB[\vec{\tau}])$ except that every (unnegated) conjunct of the form $(t \leq \tau_i t')$ is replaced by $(t < \tau_i t')$. (Notice that this is essentially the opposite transformation to the one used in the defining essential positivity in Definition 7.2.3.) Finally, let $S^{<\vec{\tau}}[KB]$ be $\overline{Sol[, (KB[\vec{\tau}])]}$.

Lemma D.2.3: *For all sufficiently small $\vec{\tau}$, $S^{<\vec{\tau}}[KB] = S^{\vec{\tau}}[KB]$.*

We defer the proof of this lemma until after the proof of the main theorem.

Consider some $\vec{u} \in S^{\vec{\tau}}[KB]$. It suffices to show that for all $\delta > 0$ there exists N_0 such that for all $N > N_0$, there exists a point $\vec{u}^N \in Sol[, ^{<}(KB[\vec{\tau}])]$ all of whose coordinates are integer multiples of $1/N$. (For then, we can take smaller and smaller δ's to create a sequence $\vec{u}^N$ converging to $\vec{u}$.) Hence, let $\delta > 0$. By Lemma D.2.3, we can find some $\vec{u}' \in Sol[, ^{<}(KB[\vec{\tau}])]$ such that $|\vec{u} \Leftrightarrow \vec{u}'| < \delta/2$. By definition, every conjunct in $, ^{<}(KB[\vec{\tau}])$ is of the form $q'(\vec{w}) = 0$, $q'(\vec{w}) > 0$, $q(\vec{w}) < \tau_i q'(\vec{w})$, or $q(\vec{w}) > \tau_i q'(\vec{w})$, where q' is a positive polynomial. Ignore for the moment the constraints of the form $q'(\vec{w}) = 0$, and consider the remaining constraints that $\vec{u}'$ satisfies. These constraints all involve strict inequalities, and the functions involved (q and q') are continuous. Thus, there exists some $\delta' > 0$ such that for all $\vec{w}$ for which $|\vec{u}' \Leftrightarrow \vec{w}| < \delta'$, these constraints are also satisfied by $\vec{w}$. Now, consider a conjunct of the form $q'(\vec{w}) = 0$ that is satisfied by $\vec{u}'$. Since q' is positive, this happens if and only if the following condition holds: for every coordinate w_i that that actually appears in q', $u'_i = 0$. In particular, if $\vec{w}$ and $\vec{u}'$ have the same coordinates with value 0, then $q'(\vec{w}) = 0$. It follows that for all $\vec{w}$, if $|\vec{u}' \Leftrightarrow \vec{w}| < \delta'$ and $\vec{u}'$ and $\vec{w}$ have the same coordinates with value 0, then $\vec{w}$ also satisfies $, ^{<}(KB[\vec{\tau}])$.

We now construct $\vec{u}^N$ that satisfies the requirements. Let i^* be the index of that component of $\vec{u}'$ with the largest value. We define $\vec{u}^N$ by considering each component, u_i^N, for $1 \leq i \leq K$:

$$u_i^N = \begin{cases} 0 & u'_i = 0 \\ \lceil N u'_i \rceil / N & i \neq i^* \text{ and } u'_i > 0 \\ u_i^N \Leftrightarrow \sum_{j' \neq i^*} (u_{j'}^N \Leftrightarrow u'_{j'}) & j = i^* \end{cases}$$

It is easy to verify that the components of $\vec{u}^N$ sum to 1. All the components in $\vec{u}'$, other than the i^*'th, are increased by at most $1/N$. The component $u_{i^*}^N$ is decreased by at most K/N. We will show that $\vec{u}^N$ has the right properties for all $N > N_0$, where N_0 is such that $1/N_0 < \min(u_{i^*}, \delta/2, \delta')/2K$. The fact that $K/N_0 < u_{i^*}$ guarantees that $\vec{u}^N$ is in Δ^K for all $N > N_0$. The fact that $2K/N_0 < \delta/2$ guarantees that $\vec{u}^N$ is within $\delta/2$ of $\vec{u}'$, and hence within δ of $\vec{u}$. Since $2K/N_0 < \delta'$, it follows that $|\vec{u}' \Leftrightarrow \vec{u}^N| < \delta'$. Since $\vec{u}^N$ is constructed to have exactly the same 0 coordinates as $\vec{u}'$, we conclude that $\vec{u}^N \in Sol[, ^{<}(KB[\vec{\tau}])]$, as required. The proof of (ii), and hence the theorem, now follows. ▮

It now remains to prove Lemma D.2.3. As we hinted earlier, this requires tools from algebraic geometry. We base our definitions on the presentation in [BCR87]. A subset A of $\mathbb{R}^\ell$ is said to be *semi-algebraic* if it is definable in the language of real-closed fields. That is, A is semi-algebraic if there is a first-order formula $\varphi(x_1, \ldots, x_\ell)$ whose free variables are $x_1, \ldots, x_\ell$ and whose only non-logical symbols are 0, 1, $+$, $\times$, $<$ and $=$ such that $\mathbb{R} \models \varphi(u_1, \ldots, u_\ell)$ iff

$(u_1, \ldots, u_\ell) \in A$.[2] A function $f : X \rightarrow Y$, where $X \subseteq \mathbb{R}^h$, $Y \subseteq \mathbb{R}^\ell$, is said to be semi-algebraic if its graph (i.e., $\{(\vec{u}, \vec{w}) : f(\vec{u}) = \vec{w}\}$) is semi-algebraic. The main tool we use is the following *Curve Selection Lemma* (see [BCR87, p. 34]):

Lemma D.2.4: *Suppose that A is a semi-algebraic set in $\mathbb{R}^\ell$ and $\vec{u} \in \overline{A}$. Then there exists a continuous, semi-algebraic function $f : [0,1] \rightarrow \mathbb{R}^\ell$ such that $f(0) = \vec{u}$ and $f(t) \in A$ for all $t \in (0,1]$.*

Our first use of the Curve Selection Lemma is in the following, which says that, in a certain sense, semi-algebraic functions behave "nicely" near limits. The type of phenomenon we wish to avoid is illustrated by $2x + x \sin \frac{1}{x}$ which is continuous at 0, but has infinitely many local maxima and minima near 0.

Proposition D.2.5: *Suppose that $g : [0,1] \rightarrow \mathbb{R}$ is a continuous, semi-algebraic function such that $g(u) > 0$ if $u > 0$ and $g(0) = 0$. Then there exists some $\epsilon > 0$ such that g is strictly increasing in the interval $[0, \epsilon]$.*

Proof: Suppose, by way of contradiction, that g satisfies the hypotheses of the proposition but there is no ϵ such that g is increasing in the interval $[0, \epsilon]$. We define a point u in $[0,1]$ to be *bad* if for some $u' \in [0, u)$ we have $g(u') \geq g(u)$. Let A be the set of all the bad points. Since g is semi-algebraic so is A, since $u' \in A$ iff

$$\exists u' \, ((0 \leq u' < u) \wedge g(u) \leq g(u')).$$

Since, by assumption, g is not increasing in any interval $[0, \epsilon]$, we can find bad points arbitrarily close to 0 and so $0 \in \overline{A}$. By the Curve Selection Lemma, there is a continuous semi-algebraic curve $f : [0,1] \rightarrow \mathbb{R}$ such that $f(0) = 0$ and $f(t) \in A$ for any $t \in (0,1]$. Because of the continuity of f the range of f — $f([0,1])$ — is $[0, r]$ for some $r \in [0,1]$. By the definition of f, $(0, r] \subseteq A$. Since $0 \notin A$, it follows that $f(1) \neq 0$; therefore $r > 0$ and so, by assumption, $g(r) > 0$. Since g is a continuous function, it achieves a maximum $v > 0$ over the range $[0, r]$. Consider the minimum point in the interval where this maximum is achieved. More precisely, let u be the infimum of the set $\{u' \in [0, r] \; : \; g(u') = v\}$. Clearly, $g(u) = v$; since $v > 0$ we obtain that $u > 0$ and therefore $u \in A$. Thus, u is bad. But that means that there is a point $u' < u$ for which $g(u') \geq g(u)$, which contradicts the choice of v and u. ▮

We can now prove Lemma D.2.3, which we restate for convenience:

Lemma D.2.3: *For all sufficiently small $\vec{\tau}$, $S^{<\vec{\tau}}[KB] = S^{\vec{\tau}}[KB]$.*[3]

Proof: Clearly $S^{<\vec{\tau}}[KB] \subseteq S^{\vec{\tau}}[KB]$. To prove the reverse inclusion, it is simpler to consider each disjunct of the canonical form of KB separately (recall that $\widehat{KB}$ is a disjunction of conjunctions).

[2] In [BCR87], a set is taken to be semi-algebraic if it is definable by a quantifier-free formula in the language of real closed fields. However, as observed in [BCR87], since the theory of real closed fields admits elimination of quantifiers [Tar51], the two definitions are equivalent.

[3] We are very grateful to Professor Gregory Brumfiel, of the Department of Mathematics at Stanford University, for his invaluable help with the proof of this lemma.

Let ξ be a conjunction that is one of the disjuncts in $\widehat{KB}$. It clearly suffices to show that $Sol[, (\xi[\vec{\tau}])] \subseteq S^{<\vec{\tau}}[\xi] = \overline{Sol[, ^{<}(\xi[\vec{\tau}])]}$. Assume, by way of contradiction, that for arbitrarily small $\vec{\tau}$, there exists some $\vec{u} \in Sol[, (\xi[\vec{\tau}])]$ which is "separated" from the set $Sol[, ^{<}(\xi[\vec{\tau}])]$, i.e., is not in its closure. More formally, we say that $\vec{u}$ is *δ-separated* from $Sol[, ^{<}(\xi[\vec{\tau}])]$ if there is no $\vec{u}' \in Sol[, ^{<}(\xi[\vec{\tau}])]$ such that with $|\vec{u} \Leftrightarrow \vec{u}'| < \delta$.

We now consider those $\vec{\tau}$ and those points in $Sol[, (\xi[\vec{\tau}])]$ that are separated from $Sol[, ^{<}(\xi[\vec{\tau}])]$:

$$A = \{(\vec{\tau}, \vec{u}, \delta) \;:\; \vec{\tau} > \vec{0},\ \delta > 0,\ \vec{u} \in Sol[, (\xi[\vec{\tau}])] \text{ is } \delta\text{-separated from } Sol[, ^{<}(\xi[\vec{\tau}])]\}.$$

Clearly A is semi-algebraic. By assumption, there are points in A for arbitrarily small tolerance vectors $\vec{\tau}$. Since A is a bounded subset of $\mathbb{R}^{m+K+1}$ (where m is the number of tolerance values in $\vec{\tau}$), we can use the Bolzano–Weierstrass Theorem to conclude that this set of points has an accumulation point whose first component is $\vec{0}$. Thus, there is a point $(\vec{0}, \vec{w}, \delta')$ in $\overline{A}$. By the Curve Selection Lemma, there is a continuous semi-algebraic function $f : [0,1] \rightarrow \mathbb{R}^{m+K+1}$ such that $f(0) = (\vec{0}, \vec{w}, \delta')$ and $f(t) \in A$ for $t \in (0,1]$.

Since f is semi-algebraic, it is semi-algebraic in each of its coordinates. By Lemma D.2.5, there is some $v > 0$ such that f is strictly increasing in each of its first m coordinates over the domain $[0, v]$. Suppose that $f(v) = (\vec{\tau}, \vec{u}, \delta)$. Now, consider the constraints in $, (\xi[\vec{\tau}])$ that have the form $q(\vec{w}) > \tau_j q'(\vec{w})$. These constraints are all satisfied by $\vec{u}$, and they are all based on strong inequalities. By the continuity of the polynomials q and q', there exists some $\epsilon > 0$ so that, for all $\vec{u}'$ such that $|\vec{u} \Leftrightarrow \vec{u}'| < \epsilon$, $\vec{u}'$ also satisfies these constraints.

Now, by the continuity of f, there exists a point $v' \in (0, v)$ sufficiently close to v so that if $f(v') = (\vec{\tau}', \vec{u}', \delta')$, then $|\vec{u} \Leftrightarrow \vec{u}'| < \min(\delta, \epsilon)$. Since $f(v) = (\vec{\tau}, \vec{u}, \delta) \in A$, and $|\vec{u} \Leftrightarrow \vec{u}'| < \delta$ it follows that $\vec{u}' \notin Sol[, ^{<}(\xi[\vec{\tau}])]$. We conclude the proof by showing that this is impossible. That is, we show that $\vec{u}' \in Sol[, ^{<}(\xi[\vec{\tau}])]$. The constraints appearing in $, ^{<}(\xi[\vec{\tau}])$ can be of the following forms: $q'(\vec{w}) = 0$, $q'(\vec{w}) > 0$, $q(\vec{w}) < \tau_j q'(\vec{w})$, or $q(\vec{w}) > \tau_j q'(\vec{w})$, where q' is a positive polynomial. Since $f(v') \in A$, we know that $\vec{u}' \in Sol[, (\xi[\vec{\tau}'])]$. The constraints of the form $q'(\vec{w}) = 0$ and $q'(\vec{w}) > 0$ are identical in $, (\xi[\vec{\tau}'])$ and in $, ^{<}(\xi[\vec{\tau}])$, and are therefore satisfied by $\vec{u}'$. Since $|\vec{u}' \Leftrightarrow \vec{u}| < \epsilon$, our discussion in the previous paragraph implies that the constraints of the form $q(\vec{w}) > \tau_j q'(\vec{w})$ are also satisfied by $\vec{u}'$. Finally, consider a constraint of the form $q(\vec{w}) < \tau_j q'(\vec{w})$. The corresponding constraint in $, (\xi[\vec{\tau}'])$ is $q(\vec{w}) \leq \tau'_j q'(\vec{w})$. Since $\vec{u}'$ satisfies this latter constraint, we know that $q(\vec{u}') \leq \tau'_j q'(\vec{u}')$. But now, recall that we proved that f is increasing over $[0, v]$ in the first m coordinates. In particular, $\tau'_j < \tau_j$. By the definition of canonical form, $q'(\vec{u}') > 0$, so that we conclude $q(\vec{u}') \leq \tau'_j q'(\vec{u}') < \tau_j q'(\vec{u}')$. Hence the constraints of this type are also satisfied by $\vec{u}'$. This concludes the proof that $\vec{u}' \in Sol[, ^{<}(KB[\vec{\tau}])]$, thus deriving a contradiction and proving the result. ∎

We are finally ready to prove Theorem 7.1.13.

Theorem 7.1.13: *For all sufficiently small $\vec{\tau}$, the following is true. Let Q be the points with greatest entropy in $S^{\vec{\tau}}[KB]$ and let $\mathcal{O} \subseteq \mathbb{R}^K$ be any open set containing Q. Then for all $\theta \in \mathcal{L}^{\approx}$ and for* $\lim^* \in \{\limsup, \liminf\}$*:*

$$\lim_{N\rightarrow\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\theta | KB) = \lim_{N\rightarrow\infty}{}^{*} \frac{\#worlds_N^{\vec{\tau}}[\mathcal{O}](\theta \wedge KB)}{\#worlds_N^{\vec{\tau}}[\mathcal{O}](KB)}.$$

Proof: Let $\vec{\tau}$ be small enough so that Theorem D.2.2 applies, and let $\mathcal{Q}$ and $\mathcal{O}$ be as in the statement of the theorem. It clearly suffices to show that the set $\mathcal{O}$ contains almost all of the worlds that satisfy KB. More precisely, the fraction of such worlds that are in $\mathcal{O}$ tends to 1 as $N \to \infty$.

Let ρ be the entropy of the points in $\mathcal{Q}$. We begin the proof by showing the existence of $\rho_L < \rho_U (< \rho)$ such that (for sufficiently large N) (a) every point $\vec{u} \in \Pi_N^{\vec{\tau}}[KB]$ where $\vec{u} \notin \mathcal{O}$ has entropy at most ρ_L, and (b) there is at least one point $\vec{u} \in \Pi_N^{\vec{\tau}}[KB]$ with $\vec{u} \in \mathcal{O}$ and entropy at least ρ_U.

For part (a), consider the space $S^{\vec{\tau}}[KB] \Leftrightarrow \mathcal{O}$. Since this space is closed, the entropy function takes on a maximum value in this space; let this be ρ_L. Since this space does not include any point with entropy ρ (these are all in $\mathcal{Q} \subseteq \mathcal{O}$), we must have $\rho_L < \rho$. By Theorem D.2.2, $\Pi_N^{\vec{\tau}}[KB] \subseteq S^{\vec{\tau}}[KB]$. Therefore, for any N, the entropy of any point in $\Pi_N^{\vec{\tau}}[KB] \Leftrightarrow \mathcal{O}$ is at most ρ_L.

For part (b), let ρ_U be some value in the interval (ρ_L, ρ) (for example $(\rho_L + \rho)/2$), and let $\vec{v}$ be any point in $\mathcal{Q}$. By the continuity of the entropy function, there exists some $\delta > 0$ such that for all $\vec{u}$ with $|\vec{u} \Leftrightarrow \vec{v}| < \delta$, $H(\vec{u}) \geq \rho_U$. Because $\mathcal{O}$ is open we can, by considering a smaller δ if necessary, assume that $|\vec{u} \Leftrightarrow \vec{v}| < \delta$ implies $\vec{u} \in \mathcal{O}$. By the second part of Theorem D.2.2, there is a sequence of points $\vec{u}^N \in \Pi_N^{\vec{\tau}}[KB]$ such that $\lim_{N \to \infty} \vec{u}^N = \vec{v}$. In particular, for N large enough we have $|\vec{u}^N \Leftrightarrow \vec{v}| < \delta$, so that $H(\vec{u}^N) > \rho_U$, proving part (b).

To complete the proof, we use Lemma 7.1.11 to conclude that for all N,

$$\#worlds_N^{\vec{\tau}}(KB) \geq \#worlds_N^{\vec{\tau}}[\vec{u}^N](KB) \geq (h(N)/f(N))e^{NH(\vec{u}^N)} \geq (h(N)/f(N))e^{N\rho_U}.$$

On the other hand,

$$\begin{aligned}
\#worlds_N^{\vec{\tau}}[\Delta^K \Leftrightarrow \mathcal{O}](KB) &\leq \sum_{\vec{u} \in \Pi_N^{\vec{\tau}}[KB] - \mathcal{O}} \#worlds_N^{\vec{\tau}}[\vec{u}](KB) \\
&\leq |\Pi_N^{\vec{\tau}}[KB] \Leftrightarrow \mathcal{O}|\, h(N)g(N)e^{N\rho_L} \\
&\leq (N+1)^K h(N)g(N)e^{N\rho_L}.
\end{aligned}$$

Therefore the fraction of models of KB which are outside $\mathcal{O}$ is at most

$$\frac{(N+1)^K h(N)f(N)g(N)e^{N\rho_L}}{h(N)e^{N\rho_U}}.$$

Since $h(N)$ cancels out and $(N+1)^k f(N)g(N)$ is a polynomial in N, this fraction tends to 0 as N grows large. The result follows. ▮

D.3 Computing for Simple Queries

Proposition 7.2.5: *Assume that KB is essentially positive and let $\mathcal{Q}$ be the set of maximum entropy points of $S^{\vec{0}}[KB]$ (and thus also of $S^{\leq \vec{0}}[KB]$). Then for all $\epsilon > 0$ and all sufficiently* *small tolerance vectors $\vec{\tau}$ (where "sufficiently small" may depend on ϵ), every maximum entropy point of $S^{\vec{\tau}}[KB]$ is within ϵ of some maximum entropy point in $\mathcal{Q}$.*

Proof: Fix $\epsilon > 0$. By way of contradiction, assume that that there is some sequence of tolerance vectors $\vec{\tau}^m$, $m = 1, 2, \ldots$, that converges to $\vec{0}$, and for each m a maximum entropy point $\vec{u}^m$ of $S^{\vec{\tau}^m}[KB]$, such that for all m, $\vec{u}^m$ is at least ϵ away from $\mathcal{Q}$. Since the space Δ^K is compact, we can assume without loss of generality that this sequence converges to some point $\vec{u}$. Since , (KB) is a finite combination (using "and" and "or") of constraints, where every such constraint is of the form: the form: $q'(\vec{w}) = 0$, $q'(\vec{w}) > 0$, $q(\vec{w}) \leq \varepsilon_j q'(\vec{w})$, or $q(\vec{w}) > \varepsilon_j q'(\vec{w})$, where q' is a positive polynomial. Since the overall number of constraints is finite we can assume, again without loss of generality, that the $\vec{u}^m$'s satisfy precisely the same conjuncts for all m. We claim that the corresponding conjuncts in , $^{\leq}(KB[\vec{0}])$ are satisfied by $\vec{u}$. For a conjunct of the form $q'(\vec{w}) = 0$ note that, if $q'(\vec{u}^m) = 0$ for all m, then this also holds at the limit, so that $q(\vec{u}) = 0$. A conjunct of the form $q'(\vec{w}) > 0$ translates into $q(\vec{w}) \geq 0$ in , $^{\leq}(KB[\vec{0}])$; such conjuncts are trivially satisfied by any point in Δ^K. If a conjunct of the form $q(\vec{w}) \leq \varepsilon_j q'(\vec{w})$ is satisfied for all $\vec{u}^m, \vec{\tau}^m$, then at the limit $q(\vec{u}) \leq 0$, which is precisely the corresponding conjunct in , $^{\leq}(KB[\vec{0}])$. Finally, for a conjunct of the form $q(\vec{w}) > \varepsilon_j q'(\vec{w})$, if $q(\vec{u}^m) > \tau_j^m q'(\vec{u}^m)$ for all m, then at the limit we have $q(\vec{u}) \geq 0$, which again is the analogous conjunct in , $^{\leq}(KB[\vec{0}])$. (The analogous conjunct in , $(KB[\vec{0}])$ is $q(\vec{w}) > 0$, but $>$ is replaced by $\geq$ in the move to , $^{\leq}(KB[\vec{0}])$.) Thus, $\vec{u}'$ is in $S^{\leq\vec{0}}[KB]$.

By assumption, all points $\vec{u}^m$ are at least ϵ away from $\mathcal{Q}$. Hence, $\vec{u}$ cannot be in $\mathcal{Q}$. If we let ρ represent the entropy of the points in $\mathcal{Q}$, and since $\mathcal{Q}$ is the set of all maximum entropy points in $S^{\leq\vec{0}}[KB]$, it follows that $H(\vec{u}) < \rho$. Choose ρ_L, ρ_U such that $H(\vec{u}) < \rho_L < \rho_U < \rho$. Since the entropy function is continuous, we know that for sufficiently large m, $H(\vec{u}^m) \leq \rho_L$. Since $\vec{u}^m$ is a maximum entropy point of $S^{\vec{\tau}^m}[KB]$, it follows that the entropy achieved in this space for sufficiently large m is at most ρ_L. We derive a contradiction by showing that for sufficiently large m, there is some point in $Sol[, (KB[\vec{\tau}^m])]$ with entropy at least ρ_U. The argument is as follows. Let $\vec{v}$ be some point in $\mathcal{Q}$. Since $\vec{v}$ is a maximum entropy point of $S^{\vec{0}}[KB]$, there are points in $Sol[, (KB[\vec{0}])]$ arbitrarily close to $\vec{v}$. In particular, there is some point $\vec{u}' \in Sol[, (KB[\vec{0}])]$ whose entropy is at least ρ_U. As we now show, this point is also in $Sol[, (KB[\vec{\tau}])]$ for all sufficiently small $\vec{\tau}$. Again, consider all the conjuncts in , $(KB[\vec{0}])$ satisfied by $\vec{u}'$, and the corresponding conjunct in , $(KB[\vec{\tau}])$. Conjuncts of the form $q'(\vec{w}) = 0$ and and $q'(\vec{w}) > 0$ in , $(KB[\vec{0}])$ remain unchanged in , $(KB[\vec{\tau}])$. Conjuncts of the form $q(\vec{w}) \leq \tau_j q'(\vec{w})$ in , $(KB[\vec{\tau}])$ are certainly satisfied by $\vec{u}'$, since the corresponding conjunct in , $(KB[\vec{0}])$, namely $q(\vec{w}) \leq 0$, is satisfied by $\vec{u}'$, so that $q(\vec{u}') \leq 0 \leq \tau_j q'(\vec{u}')$ (recall that q' is a positive polynomial). Finally, consider a conjunct in , $(KB[\vec{\tau}])$ of the form $q(\vec{w}) > \tau_j q'(\vec{w})$. The corresponding conjunct in , $(KB[\vec{0}])$ is $q(\vec{w}) > 0$. Suppose $q(\vec{u}') = \delta > 0$. Since the value of q' is bounded over the compact space Δ^K, it follows that for all sufficiently small τ_j, $\tau_j q'(\vec{u}') < \delta$. Thus, $q(\vec{u}') > \tau_j q'(\vec{u}')$ for all sufficiently small τ_j, as required. It follows that $\vec{u}'$ is in $Sol[, (KB[\vec{\tau}])]$ for all sufficiently small $\vec{\tau}$, and in particular in $Sol[, (KB[\vec{\tau}^m])]$ for all sufficiently large m. But $H(\vec{u}') \geq \rho_U$, whereas we showed that the maximum entropy achieved in $S^{\vec{\tau}^m}[KB]$ is at most $\rho_L < \rho_U$. This contradiction proves that our assumption was false, so that the conclusion of the proposition necessarily holds. ∎

Theorem 7.2.8: *Suppose $\varphi(c)$ is a simple query for KB. For all $\vec{\tau}$ sufficiently small, if $\mathcal{Q}$ is the set of maximum entropy points in $S^{\vec{\tau}}[KB]$ and $F_{[\psi]}(\vec{v}) > 0$ for all $\vec{v} \in \mathcal{Q}$, then for* $\lim^* \in \{\limsup, \liminf\}$

$$\lim_{N\to\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\varphi(c)|KB) \in \left[\inf_{\vec{v}\in\mathcal{Q}} F_{[\varphi|\psi]}(\vec{v}), \sup_{\vec{v}\in\mathcal{Q}} F_{[\varphi|\psi]}(\vec{v})\right].$$

Proof: Let $W \in \mathcal{W}^*$, and let $\vec{u} = \pi(W)$. The proportion expression $||\psi(x)||_x$ is clearly equal to

$$\sum_{A_j\in\mathcal{A}(\psi)} ||A_j(x)||_x = \sum_{A_j\in\mathcal{A}(\psi)} u_j = F_{[\psi]}(\vec{u}).$$

If $F_{[\psi]}(\vec{u}) > 0$, then by the same reasoning we conclude that the value of $\|\varphi(x)|\psi(x)\|_x$ at W is equal to $F_{[\varphi|\psi]}(\vec{u})$.

Now, let λ_L and λ_R be $\inf_{\vec{v}\in\mathcal{Q}} F_{[\varphi|\psi]}(\vec{v})$ and $\sup_{\vec{v}\in\mathcal{Q}} F_{[\varphi|\psi]}(\vec{v})$ respectively; by our assumption, $F_{[\varphi|\psi]}(\vec{v})$ is well-defined for all $\vec{v} \in \mathcal{Q}$. Since the denominator is not 0, $F_{[\varphi|\psi]}$ is a continuous function at each maximum entropy point. Thus, since $F_{[\varphi|\psi]}(\vec{v}) \in [\lambda_L, \lambda_R]$ for all maximum entropy points, the value of $F_{[\varphi|\psi]}(\vec{u})$ for $\vec{u}$ "close" to some $\vec{v}$, will either be in the range $[\lambda_L, \lambda_U]$ or very close to it. More precisely, choose any $\epsilon > 0$, and define $\theta[\epsilon]$ to be the formula

$$\|\varphi(x)|\psi(x)\|_x \in [\lambda_L \Leftrightarrow \epsilon, \lambda_U + \epsilon].$$

Since $\epsilon > 0$, it is clear that there is some sufficiently small open set $\mathcal{O}$ around $\mathcal{Q}$ such that this proportion expression is well-defined and within these bounds at all worlds in $\mathcal{O}$. Thus, by Corollary 7.1.14, $\Pr_\infty^{\vec{\tau}}(\theta[\epsilon]|KB) = 1$. Using Theorem 7.1.16, we obtain that for $\lim^*$ as above,

$$\lim_{N\to\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\varphi|KB) = \lim_{N\to\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\varphi|KB \wedge \theta[\epsilon]).$$

But now we can use the direct inference technique outlined earlier. We are interested in the probability of $\varphi(c)$, where the only information we have about c in the knowledge base is $\psi(c)$, and where we have statistics for $\|\varphi(x)|\psi(x)\|_x$. These are precisely the conditions under which Theorem 4.2.1 applies. We conclude that

$$\lim_{N\to\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\varphi|KB) \in [\lambda_L \Leftrightarrow \epsilon, \lambda_U + \epsilon].$$

Since this holds for all $\epsilon > 0$, it is necessarily the case that

$$\lim_{N\to\infty}{}^{*} \Pr{}_N^{\vec{\tau}}(\varphi|KB) \in [\lambda_L, \lambda_U],$$

as required. ▮

Theorem 7.2.10: *Suppose $\varphi(c)$ is a simple query for KB. If the space $S^{\vec{0}}[KB]$ has a unique maximum entropy point $\vec{v}$, KB is essentially positive, and $F_{[\psi]}(\vec{v}) > 0$, then*

$$\Pr{}_\infty(\varphi(c)|KB) = F_{[\varphi|\psi]}(\vec{v}).$$

Proof: Note that the fact that $S^{\vec{0}}[KB]$ has a unique maximum entropy point does not guarantee that this is also the case for $S^{\vec{\tau}}[KB]$. However, Proposition 7.2.5 implies that the maximum entropy points of the latter space are necessarily close to $\vec{v}$. More precisely, if we choose some $\epsilon > 0$, we conclude that for all sufficiently small $\vec{\tau}$, all the maximum entropy points of $S^{\vec{\tau}}[KB]$ will be within ϵ of $\vec{v}$. Now, pick some arbitrary $\delta > 0$. Since $F_{[\psi]}(\vec{v}) > 0$, it follows that $F_{[\varphi|\psi]}$ is continuous at $\vec{v}$. Therefore, there exists some $\epsilon > 0$ such that if $\vec{u}$ is within ϵ of $\vec{v}$, $F_{[\varphi|\psi]}(\vec{u})$ is within δ of $F_{[\varphi|\psi]}(\vec{v})$. In particular, this is the case for all maximum entropy points of $S^{\vec{\tau}}[KB]$ for all sufficiently small $\vec{\tau}$. This allows us to apply Theorem 7.2.8, and conclude that for all sufficiently small $\vec{\tau}$ and for $\lim^* \in \{\limsup, \liminf\}$, $\lim^*_{N\to\infty} \Pr^{\vec{\tau}}_N(\varphi(c)|KB)$ is within δ of $F_{[\varphi|\psi]}(\vec{v})$. Since this holds for all $\delta > 0$, it follows that

$$\liminf_{N\to\infty} \Pr^{\vec{\tau}}_N(\varphi(c)|KB) = \limsup_{N\to\infty} \Pr^{\vec{\tau}}_N(\varphi(c)|KB) = F_{[\varphi|\psi]}(\vec{v}).$$

Thus, by definition, $\Pr_\infty(\varphi(c)|KB) = F_{[\varphi|\psi]}(\vec{v})$. ▮

Theorem 7.2.13: *Let Λ be a conjunction of constraints of the form $\Pr(\beta|\beta') = \lambda$ or $\Pr(\beta|\beta') \in [\lambda_1, \lambda_2]$. There is a unique probability distribution μ^* of maximum entropy satisfying Λ. Moreover, for all β and β', if $\Pr_{\mu^*}(\beta') > 0$, then*

$$\Pr_\infty(\xi_\beta(c)|\xi_{\beta'}(c) \wedge KB'[\Lambda]) = \Pr_{\mu^*}(\beta|\beta').$$

Proof: Clearly, the formulas $\varphi(x) = \xi_\beta(x)$ and $\psi(x) = \xi_{\beta'}(x)$ are essentially propositional. The knowledge base KB'_Λ is in the form of a conjunction of very simple proportion formulas, none of which are negated. Let $KB = \psi(c) \wedge KB'$. Notice that , (KB) is a conjunction of linear constraints, none of which is negated, and involving only weak inequalities ($\geq$) except for conjuncts of the form $u_j > 0$. It is therefore obvious that $S^{\vec{0}}[KB] = S^{\leq\vec{0}}[KB]$, so that KB is essentially positive. As we observed earlier, the fact that these are linear constraints also means that $S^{\leq\vec{0}}[KB]$ is convex, and hence has a unique maximum entropy point, say $\vec{v}$. Let $\mu^* = \mu_{\vec{v}}$ be the distribution over Ω corresponding to $\vec{v}$. It is clear that the constraints of , $(KB'_\Lambda[\vec{0}])$ on the points of Δ^K are precisely the same ones as those of Λ. Therefore, μ^* is the unique maximum entropy distribution satisfying the constraints of Λ. By Remark 7.2.12, it follows that $F_{[\xi_{\beta'}]}(\vec{v}) = \mu^*(\beta')$. Since we have assumed that $\mu^*(\beta') > 0$, we can now apply Theorem 7.2.10 to conclude that

$$\Pr_\infty(\varphi(c)|\psi(c) \wedge KB'_\Lambda) = F_{[\varphi|\psi]}(\mu^*) = \Pr_{\mu^*}(\beta|\beta'). \quad ▮$$

Theorem 7.2.14: *Let c be a constant symbol. Using the translation described in Section 7.2.3, for any set $\mathcal{R}$ of defeasible rules, $B \to C$ is an ME-plausible consequence of $\mathcal{R}$ iff*

$$\Pr_\infty\left(\xi_C(c) \,\middle|\, \xi_B(c) \wedge \bigwedge_{r\in\mathcal{R}} \theta_r\right) = 1.$$

Proof: Let KB' denote $\bigwedge_{r\in\mathcal{R}} \theta_r$. For all sufficiently small $\vec{\tau}$, and for $\epsilon = \tau_1$ let μ^* denote $\mu^*_{\epsilon,\mathcal{R}}$. It clearly suffices to prove that

$$\Pr^{\vec{\tau}}_\infty(\xi_C(c)|\xi_B(c) \wedge KB') = \Pr_{\mu^*}(C|B),$$

where by equality we also mean that one side is defined iff the other is also defined. We leave it to the reader to check that a point $\vec{u}$ in Δ^K satisfies , $(KB'[\vec{\tau}])$ iff the corresponding distribution μ ϵ-satisfies $\mathcal{R}$. Therefore, the maximum entropy point $\vec{v}$ of $S^{\vec{\tau}}[KB']$ corresponds precisely to μ^*. Now, there are two cases: either $\mu^*(B) > 0$ or $\mu^*(B) = 0$. In the first case, by Remark 7.2.12, $\Pr_{\mu^*}(\xi_B(c)) = F_{[\xi_B(c)]}(\vec{v})$, so the latter is also positive. This also implies that $\vec{v}$ is consistent with the constraints entailed by $\psi = \xi_B(c)$, so that $\vec{v}$ is also the unique maximum entropy point of $S^{\vec{\tau}}[KB]$ (where $KB = \xi_B(c) \wedge KB'$). We can therefore use Corollary 7.2.9 and Remark 7.2.12 to conclude that $\Pr^{\vec{\tau}}_\infty(\xi_C(c)|KB) = F_{[\xi_C(c)|\xi_B(c)]}(\vec{v}) = \Pr_{\mu^*}(C|B)$, and that all three terms are well-defined. Assume, on the other hand, that $\mu^*(B) = 0$, so that $\Pr_{\mu^*}(C|B)$ is not well-defined. In this case, we can use a known result (see [PV89]) for the maximum entropy point over a space defined by linear constraints, and conclude that for all μ satisfying $\mathcal{R}$, necessarily $\mu(B) = 0$. Using the connection between distributions μ satisfying $\mathcal{R}$ and points $\vec{u}$ in $S^{\vec{\tau}}[KB']$, we conclude that this is also the case for all $\vec{u} \in S^{\vec{\tau}}[KB']$. By part (a) of Theorem D.2.2, this means that in any world satisfying KB', the proportion $||\xi_B(x)||_x$ is necessarily 0. Thus, KB' is inconsistent with $\xi_B(c)$, and $\Pr^{\vec{\tau}}_\infty(\xi_C(c)|\xi_B(c) \wedge KB')$ is also not well-defined. ▮

D.4 Extending the Class of Queries

Theorem 7.2.23: *If KB and $\vec{\tau} > \vec{0}$ are stable for σ^* then* $\Pr^{\vec{\tau}}_\infty(\sigma^*|KB) = 1$.

Proof: By Theorem 7.1.14, it suffices to show that there is some open neighborhood containing $\mathcal{Q}$, the maximum entropy points of $S^{\vec{\tau}}[KB]$, such that every world W of KB in this neighborhood has $\sigma(W) = \sigma^*$. So suppose this is not the case. Then there is some sequence of models $W_1, W_2, \ldots$ such that $(W_i, \vec{\tau}) \models KB \wedge \neg\sigma^*$, and $\lim_{i\to\infty} \max_{\vec{v}\in\mathcal{Q}} |\pi(W_i) \Leftrightarrow \vec{v}| = 0$. Since Δ^K is compact the sequence $\pi(W_1), \pi(W_2), \ldots$ must have at least one accumulation point, say $\vec{u}$. This point must be in the closure of the set $\mathcal{Q}$. But, in fact, $\mathcal{Q}$ is a closed set (because entropy is a continuous function) and so $\vec{u} \in \mathcal{Q}$. By part (a) of Theorem D.2.2, $\pi(W_i) \in S^{\vec{\tau}}[KB \wedge \neg\sigma^*]$ for every i, and so, since this space is closed, $\vec{u}' \in S^{\vec{\tau}}[KB \wedge \neg\sigma^*]$ as well. But this means that $\vec{u}'$ is an unsafe maximum entropy point, contrary to our assumption. ▮

In the remainder of this section we prove Theorem 7.2.27. For this purpose, fix $KB = \psi \wedge KB'$, φ, and σ^* to be as in the statement of this theorem, and let $\vec{v}$ be the unique maximum entropy point of $S^{\vec{0}}[KB]$.

Let $\mathcal{Z} = \{c_1, \ldots, c_m\}$ be the set of constant symbols appearing in ψ and in φ. Due to the separability assumption, KB' contains none of the constant symbols in $\mathcal{Z}$. Let $\chi^{\neq}$ be the formula $\bigwedge_{i\neq j} c_i \neq c_j$. We first prove that $\chi^{\neq}$ has probability 1 given KB'.

Lemma D.4.1: *For $\chi^{\neq}$ and KB' as above,* $\Pr_\infty(\chi^{\neq}|KB') = 1$.

Proof: We begin by showing that $\Pr_\infty(\neg\chi^{\neq}|KB') = 0$. Let c and c' be two constant symbols and consider $\Pr_\infty(c = c'|KB')$. We again use the direct inference technique. More precisely, fix N and consider the value of the proportion expression $||x = x'||_{x,x'}$ in any world of size N.

This value is clearly $1/N$. Since c and c' appear nowhere in KB' we can use Theorem 4.2.1 to conclude that $\Pr_N(c = c'|KB') = 1/N$. Therefore, $\Pr_\infty(c = c'|KB') = 0$. It is straightforward to verify that, since $\neg\chi^{\neq}$ is equivalent to a finite disjunction each disjunct of which implies $c = c'$ for at least one pair of constants c and c', we must have $\Pr_\infty(\neg\chi^{\neq}|KB') = 0$. ▮

As we stated in Section 7.2.4, our general technique for computing the probability of an arbitrary formula φ is to partition the worlds into a finite collection of classes, such that φ behaves uniformly over each class, and compute the relative weights of the classes. As we will show later, the classes are essentially defined using complete descriptions. Their relative weight corresponds to the probabilities of the different complete descriptions given KB.

Proposition D.4.2: *Let $KB = KB' \wedge \psi$ and $\vec{v}$ be as above. Assume $\Pr_\infty(\psi|KB') > 0$. Let D be a complete description over $\mathcal{Z}$ that is consistent with ψ. Then:*

(a) If D is inconsistent with $\chi^{\neq}$, then $\Pr_\infty(D|KB) = 0$.

(b) If D is consistent with $\chi^{\neq}$, then

$$\Pr_\infty(D|KB) = \frac{F_{[D]}(\vec{v})}{\sum_{D' \in \mathcal{A}(\psi \wedge \chi^{\neq})} F_{[D']}(\vec{v})}.$$

Proof: First, observe that if all limits exist and the denominator is nonzero then

$$\Pr_\infty(\neg\chi^{\neq}|\psi \wedge KB') = \frac{\Pr_\infty(\neg\chi^{\neq} \wedge \psi|KB')}{\Pr_\infty(\psi|KB')}.$$

By assumption, the denominator is, indeed, nonzero. Furthermore, by Lemma D.4.1, $\Pr_\infty(\neg\chi^{\neq} \wedge \psi|KB') \leq \Pr_\infty(\neg\chi^{\neq}|KB') = 0$. Hence $\Pr_\infty(\chi^{\neq}|KB)$ is also 1. We can therefore use Theorem 7.1.16 to conclude that

$$\Pr_\infty(D|KB) = \Pr_\infty(D|KB \wedge \chi^{\neq}).$$

Part (a) of the proposition follows immediately.

To prove part (b), recall that $\psi \wedge \chi^{\neq}$ is equivalent to the disjunction $\bigwedge_{D' \in \mathcal{A}(\psi \wedge \chi^{\neq})} D$. By simple probabilistic reasoning, and using the assumption that $\Pr_\infty(\psi|KB') > 0$, we conclude that:

$$\Pr_\infty(D|\psi \wedge KB') = \frac{\Pr_\infty(D \wedge \psi|KB')}{\Pr_\infty(\psi|KB')} = \frac{\Pr_\infty(D \wedge \psi|KB')}{\sum_{E \in \mathcal{A}(\psi \wedge \chi^{\neq})} \Pr_\infty(E|KB')}.$$

By assumption, D is consistent with $\chi^{\neq}$ and is in $\mathcal{A}(\psi)$. Since D is a complete description, we must have that $D \Rightarrow \psi$ is valid. Thus, the numerator on the right-hand side of this equation is simply $\Pr_\infty(D|KB')$. Hence, the problem of computing $\Pr_\infty(D|KB)$ reduces to a series of computations of the form $\Pr_\infty(E|KB')$ for some complete description E.

Fix any such description E. Recall that E can be decomposed into three parts: the unary part E^1, the non-unary part $E^{>1}$, and the equality part $E^=$. Since E is in $\mathcal{A}(\chi^{\neq})$, we conclude that $\chi^{\neq}$ is equivalent to $E^=$. Using Theorem 7.1.16 twice and some probabilistic reasoning:

$$\begin{aligned}
\Pr_\infty(E^{>1} \wedge E^1 \wedge E^= | KB') &= \Pr_\infty(E^{>1} \wedge E^1 \wedge E^= | KB' \wedge \chi^{\neq}) \\
&= \Pr_\infty(E^{>1} \wedge E^1 | KB' \wedge \chi^{\neq}) \\
&= \Pr_\infty(E^{>1} | KB' \wedge \chi^{\neq} \wedge E^1) \cdot \Pr_\infty(E^1 | KB' \wedge \chi^{\neq}) \\
&= \Pr_\infty(E^{>1} | KB' \wedge \chi^{\neq} \wedge E^1) \cdot \Pr_\infty(E^1 | KB').
\end{aligned}$$

In order to simplify the first expression, recall that none of the predicate symbols in $E^{>1}$ occur anywhere in $KB' \wedge \chi^{\neq} \wedge E^1$. Therefore, the probability of $E^{>1}$ given $KB' \wedge \chi^{\neq}$ is equal to the probability that the elements denoting the $|\mathcal{Z}|$ (different) constants satisfy some particular configuration of non-unary properties. It should be clear that, by symmetry, all such configurations are equally likely. Therefore, the probability of any one of them is a constant, equal to 1 over the total number of configurations.[4] Let ρ denote the constant which is equal to $\Pr_\infty(E^{>1} | KB' \wedge \chi^{\neq} \wedge E^1)$ for all E.

The last step is to show that, if E^1 is equivalent to $\bigwedge_{j=1}^m A_{i_j}(c_j)$, then

$$\Pr_\infty(E^1 | KB') = F_{[D]}(\vec{v}).$$

We can now apply standard probabilistic reasoning to show that

$$\begin{aligned}
&\Pr_\infty(\bigwedge_{j=1}^{m} A_{i_j}(c_j) | KB') \\
&= \Pr_\infty(A_{i_1}(c_1) | \bigwedge_{j=2}^{m} A_{i_j}(c_j) \wedge KB') \cdot \Pr_\infty(A_{i_2}(c_2) | \bigwedge_{j=3}^{m} A_{i_j}(c_j) \wedge KB') \\
&\quad \cdot \ldots \cdot \Pr_\infty(A_{i_{m-1}}(c_{m-1}) | A_{i_m}(c_m) \wedge KB') \cdot \Pr_\infty(A_{i_m}(c_m) | KB') \\
&= v_{i_1} \cdot \ \ldots \ \cdot v_{i_m} \text{ (using Theorem 7.2.10; see below)} \\
&= F_{[D]}(\vec{v}).
\end{aligned}$$

The second-last step is derived from m applications of Theorem 7.2.10. Our assumptions guarantee that $A_{i_j}(c_j)$ is a simple query for $A_{i_{j+1}}(c_{j+1}) \wedge \ldots A_{i_m}(c_m) \wedge KB'$.

We can now put everything together to conclude that

$$\Pr_\infty(D|KB) = \frac{\Pr_\infty(D|KB')}{\sum_{E \in \mathcal{A}(\psi \wedge \chi^{\neq})} \Pr_\infty(E|KB')} = \frac{F_{[D]}(\vec{v})}{\sum_{E \in \mathcal{A}(\psi \wedge \chi^{\neq})} F_{[E]}(\vec{v})},$$

proving part (b). ∎

[4] Although we do not need the value of this constant in our calculations below, it is in fact easy to verify that its value is $\prod_{R \in (\Phi - \Psi)} 2^{m^{arity(R)}}$, where $m = |\mathcal{Z}|$.

We now address the issue of computing $\Pr_\infty(\varphi|KB)$ for an arbitrary formula φ. In order to do that, we must first investigate the behavior of $\Pr_\infty^{\vec{\tau}}(\varphi|KB)$ for small $\vec{\tau}$. Fix some sufficiently small $\vec{\tau} > 0$, and let $\mathcal{Q}$ be the set of maximum entropy points of $S^{\vec{\tau}}[KB]$. Assume KB and $\vec{\tau}$ are stable for σ^*. By definition, this means that for every $\vec{v} \in \mathcal{Q}$, $\sigma(\vec{v}) = \sigma^*$. Let I be the set of i's for which σ^* contains the conjunct $\exists x A_i(x)$. Since for all $\vec{v}$, $\sigma(\vec{v}) = \sigma^*$, we must have that for all $i \in I$, $v_i > 0$. Since $\mathcal{Q}$ is a closed set, this implies that there exists some $\epsilon > 0$ such that for all $\vec{v} \in \mathcal{Q}$ and for all $i \in I$, $v_i > \epsilon$. Let $\theta[\epsilon]$ be the formula:

$$\bigwedge_{i \in I} ||A_i(x)||_x > \epsilon.$$

The following proposition is now easy to prove:

Proposition D.4.3: *Suppose that KB and $\vec{\tau}$ are stable for σ^* and, that $\mathcal{Q}$, i, $\theta[\epsilon]$, and $\chi^{\neq}$ are as above. Then*

$$\Pr_\infty^{\vec{\tau}}(\varphi|KB) = \sum_{D \in \mathcal{A}(\psi)} \Pr_\infty^{\vec{\tau}}(\varphi|KB' \wedge \theta[\epsilon] \wedge \sigma^* \wedge D) \cdot \Pr_\infty^{\vec{\tau}}(D|KB).$$

Proof: Clearly, $\theta[\epsilon]$ satisfies the conditions of Corollary 7.1.14, allowing us to conclude that $\Pr_\infty^{\vec{\tau}}(\theta[\epsilon]|KB) = 1$. Similarly, by Theorem 7.2.23 and the assumptions of Theorem 7.2.27, we can conclude that $\Pr_\infty^{\vec{\tau}}(\sigma^*|KB) = 1$. Since the conjunction of two assertions that have probability 1 also has probability 1, we conclude using Theorem 7.1.16 that $\Pr_\infty^{\vec{\tau}}(\varphi|KB) = \Pr_\infty^{\vec{\tau}}(\varphi|KB \wedge \theta[\epsilon] \wedge \sigma^*)$.

Now, recall that ψ is equivalent to the disjunction $\bigvee_{D \in \mathcal{A}(\psi)} D$. By straightforward probabilistic reasoning, we can therefore conclude that:

$$\Pr_\infty^{\vec{\tau}}(\varphi|KB \wedge \theta[\epsilon] \wedge \sigma^*) = \sum_{D \in \mathcal{A}(\psi)} \Pr_\infty^{\vec{\tau}}(\varphi|KB \wedge \theta[\epsilon] \wedge \sigma^* \wedge D) \cdot \Pr_\infty^{\vec{\tau}}(D|KB \wedge \theta[\epsilon] \wedge \sigma^*).$$

In order to bring the second expression to the desired form, we appeal to Theorem 7.1.16 again. That is, $\Pr_\infty^{\vec{\tau}}(D|KB \wedge \theta[\epsilon] \wedge \sigma^*) = \Pr_\infty^{\vec{\tau}}(D|KB)$. The desired expression now follows. ∎

We now simplify the expression $\Pr_\infty^{\vec{\tau}}(\varphi|KB \wedge \theta[\epsilon] \wedge \sigma^* \wedge D)$.

Proposition D.4.4: *For φ, KB, σ^*, D, $\theta[\epsilon]$ as above, if* $\Pr_\infty^{\vec{\tau}}(D|KB) > 0$, *then*

$$\Pr_\infty^{\vec{\tau}}(\varphi|KB' \wedge \theta[\epsilon] \wedge \sigma^* \wedge D) = \Pr_\infty(\varphi|\sigma^* \wedge D),$$

and its value is either 0 or 1. Note that since the latter probability only refers to first-order formulas, it is independent of the tolerance values.

Proof: If $\Pr_\infty^{\vec{\tau}}(D|KB) > 0$, then the first limit above is well-defined. That the second limit is either 0 or 1 is proved in Theorem 6.3.4, where it is shown that the asymptotic probability of any pure first-order sentence when conditioned on knowledge of the form $\sigma^* \wedge D$ (which is, essentially, what was called a *model description* in Chapter 6) is either either 0 or 1. Very similar techniques can be used to show that the first limit is also either 0 or 1, and that the conjuncts $KB' \wedge \theta[\epsilon]$ do not affect this limit (so that the left-hand side and the right-hand side are in fact equal). We briefly sketch the relevant details here, referring the reader to Section 6.3 for full details.

The idea (which actually goes back to Fagin [Fag76]) is to associate with a model description such as $\sigma^* \wedge D$ a theory T which essentially consists of *extension axioms*. Intuitively, an extension axiom says that any finite substructure of the model, defined by a complete description D' can be extended in all possible ways definable by another description D''. We say that a description D'' *extends* a description D' if all conjuncts of D' are also conjuncts in D''. An extension axiom has that form $\forall x_1, \ldots, x_j\, (D' \Rightarrow \exists x_{j+1}\, D'')$, where D' is a complete description over $\mathcal{X} = \{x_1, \ldots, x_j\}$ and D'' is a complete description over $\mathcal{X} \cup \{x_{j+1}\}$, such that D'' extends D', both D' and D'' extend D, and both are consistent with σ^*. It is then shown that (a) T is complete (so that for each formula ξ, either $T \models \xi$ or $T \models \neg\xi$) and (b) if $\xi \in T$ then $\Pr_\infty(\xi | \sigma^* \wedge D) = 1$. From (b) it easily follows that if $T \models \xi$, then $\Pr_\infty(\xi | \sigma^* \wedge D)$ is also 1. Using (a), the desired 0-1 law follows. The only difference from the proof in Section 6.3 is that we need to show that (b) holds even when we condition on $KB' \wedge \theta[\epsilon] \wedge \sigma^* \wedge D$, instead of just on $\sigma^* \wedge D$ (that is, we need to reprove Proposition C.2.2).

So, suppose ξ is the extension axiom $\forall x_1, \ldots, x_j\, (D' \Rightarrow \exists x_{j+1}\, D'')$. We must show that $\Pr_\infty(\xi | KB' \wedge \theta[\epsilon] \wedge \sigma^* \wedge D) = 1$. Fix a domain size N and consider the set of worlds satisfying $KB' \wedge \theta[\epsilon] \wedge \sigma^* \wedge D$. Now consider some particular j domain elements, say $d_1, \ldots, d_j$, that satisfy D'. Observe that, since D' extends D, the denotations of the constants are all among $d_1, \ldots, d_j$. For a given $d \notin \{d_1, \ldots, d_j\}$, let $B(d)$ denote the event that $d_1, \ldots, d_j, d$ satisfy D'', given that $d_1, \ldots, d_j$ satisfy D'. What is the probability of $B(d)$ given $\sigma^* \wedge D \wedge KB \wedge \theta[\epsilon]$? First, note that since d does not denote any constant, it cannot be mentioned in any way in the knowledge base. Thus, this probability is the same for all d. The description D'' determines two types of properties for x_{j+1}. The unary properties of x_{j+1} itself—i.e., the atom A_i to which x_{j+1} must belong—and the relations between x_{j+1} and the remaining variables $x_1, \ldots, x_j$ using the non-unary predicate symbols. Since D'' is consistent with σ^*, the description σ^* must contain a conjunct $\exists x\, A_i(x)$ if D'' implies $A_i(x_{j+1})$. By definition, $\theta[\epsilon]$ must therefore contain the conjunct $||A_i(x)||_x > \epsilon$. Hence, the probability of picking d in A_i is at least ϵ. For any sufficiently large N, the probability of picking d in A_i which is different from $d_1, \ldots, d_j$ (as required by the definition of the extension axiom) is at least $\epsilon/2 > 0$. The probability that $d_1, \ldots, d_j, d$ also satisfy the remaining conjuncts of D'', given that d is in atom A_i and $d_1, \ldots, d_j$ satisfy D', is very small but bounded away from 0. (For this to hold, we need the assumption that the non-unary predicates are not mentioned in the KB.) This is the case because the total number of possible ways to choose the properties of d (as they relate to $d_1, \ldots, d_j$) is independent of N. We can therefore conclude that the probability of $B(d)$ (for sufficiently large N), given that $d_1, \ldots, d_j$ satisfy D, is bounded away from 0 by some λ independent of N. Since the properties of an element d and its relation to $d_1, \ldots, d_j$ can be chosen independently of the properties of a different element d', the different events $B(d), B(d'), \ldots$ are all independent. Therefore, the probability that there is no domain element at all that, together with $d_1, \ldots, d_j$, satisfies

D'' is at most $(1 \Leftrightarrow \lambda)^{N-j}$. This bounds the probability of the extension axiom being false, relative to fixed $d_1, \ldots, d_j$. There are $\binom{N}{j}$ ways of these choosing j elements, so the probability of the axiom being false anywhere in a model is at most $\binom{N}{j}(1 \Leftrightarrow \lambda)^{N-j}$. This tends to 0 as N goes to infinity. Therefore, the extension axiom $\forall x_1, \ldots, x_j\, (D' \Rightarrow \exists x_{j+1}\, D'')$ has asymptotic probability 1 given $KB' \wedge \theta[\epsilon] \wedge \sigma^* \wedge D$, as desired. ∎

Finally, we are in a position to prove Theorem 7.2.27.

Theorem 7.2.27: *Let φ be a formula in $\mathcal{L}^{\approx}$, and let $KB = KB' \wedge \psi$ be an essentially positive knowledge base in $\mathcal{L}^{\approx}_1$ which is separable with respect to φ. Let $\mathcal{Z}$ be the set of constants appearing in φ or in ψ (so that KB' contains none of the constants in $\mathcal{Z}$), and let $\chi^{\neq}$ be the formula $\bigwedge_{c,c' \in \mathcal{Z}} c \neq c'$. Assume that there exists a size description σ^* such that, for all $\vec{\tau} > 0$, KB and $\vec{\tau}$ are stable for σ^*, and that the space $S^{\vec{0}}[KB]$ has a unique maximum entropy point $\vec{v}$, then:*

$$\Pr_\infty(\varphi|KB) = \frac{\sum_{D \in \mathcal{A}(\psi \wedge \chi^{\neq})} \Pr_\infty(\varphi|\sigma^* \wedge D) F_{[D]}(\vec{v})}{\sum_{D \in \mathcal{A}(\psi \wedge \chi^{\neq})} F_{[D]}(\vec{v})},$$

if the denominator is positive.

Proof: Assume without loss of generality that ψ mentions all the constant symbols in φ, so that $\mathcal{A}(\psi \wedge \chi^{\neq}) \subseteq \mathcal{A}(\psi)$. By Proposition D.4.3,

$$\Pr^{\vec{\tau}}_\infty(\varphi|KB) = \sum_{D \in \mathcal{A}(\psi)} \Pr^{\vec{\tau}}_\infty(\varphi|KB \wedge \theta[\epsilon] \wedge \sigma^* \wedge D) \cdot \Pr^{\vec{\tau}}_\infty(D|KB).$$

Note that we cannot easily take limits of $\Pr^{\vec{\tau}}_\infty(\varphi|KB \wedge \theta[\epsilon] \wedge \sigma^* \wedge D)$ as $\vec{\tau}$ goes to $\vec{0}$, because of the dependence of this expression on $\theta[\epsilon]$ (the value of ϵ used depends on the choice of $\vec{\tau}$). However, applying Proposition D.4.4, we get

$$\Pr^{\vec{\tau}}_\infty(\varphi|KB) = \sum_{D \in \mathcal{A}(\psi)} \Pr_\infty(\varphi|\sigma^* \wedge D) \cdot \Pr^{\vec{\tau}}_\infty(D|KB).$$

We can now take the limit as $\vec{\tau}$ goes to $\vec{0}$. To do this, we use Proposition D.4.2. The hypotheses of the theorem easily imply that $\Pr_\infty(\psi|KB') > 0$. Part (a) tells us we can ignore those complete descriptions that are inconsistent with $\chi^{\neq}$. We can now apply part (b) to get the desired result. ∎

Bibliography

[Ada75] E. Adams. *The Logic of Conditionals.* D. Reidel, Dordrecht, Netherlands, 1975.

[Ash93] N. Asher. Extensions for commonsense entailment. In *Proceedings of the IJCAI Workshop on Conditionals in Knowledge Representation*, pages 26–41, 1993.

[Bac90] F. Bacchus. *Representing and Reasoning with Probabilistic Knowledge.* MIT Press, Cambridge, MA, 1990.

[BCD$^+$93] S. Benferhat, C. Cayrol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax-based entailment. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pages 640–645, 1993.

[BCR87] J. Bochnak, M. Coste, and M-F. Roy. *Géométrie Algébrique Réelle*, volume 12 of *A Series of Modern Surveys in Mathematics.* Springer-Verlag, Berlin Heidelberg, 1987.

[BGHK92] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. From statistics to belief. In *Proc. National Conference on Artificial Intelligence (AAAI '92)*, pages 602–608, 1992.

[BGHK93a] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. Forming beliefs about a changing world. In preparation, 1993.

[BGHK93b] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. A response to: "Believing on the basis of evidence". *Computational Intelligence, to appear*, 1993.

[BGK85] A. Blass, Y. Gurevich, and D. Kozen. A zero–one law for logic with a fixed point operator. *Information and Control*, 67:70–90, 1985.

[Bou91] C. Boutilier. *Conditional Logics for Default Reasoning and Belief Revision.* PhD thesis, Department of Computer Science, University of Toronto, 1991.

[Car50] R. Carnap. *Logical Foundations of Probability.* University of Chicago Press, Chicago, 1950.

[Car52] R. Carnap. *The Continuum of Inductive Methods.* University of Chicago Press, Chicago, 1952.

[Che83] P. C. Cheeseman. A method of computing generalized Bayesian probability values for expert systems. In *Proc. Eighth International Joint Conference on Artificial Intelligence (IJCAI '83)*, pages 198–202, 1983.

[Chu91] R. Chuaqui. *Truth, possibility, and probability: new logical foundations of probability and statistical inference.* North-Holland, 1991.

[CKS81] A. K. Chandra, D. Kozen, and L. J. Stockmeyer. Alternation. *Journal of the ACM*, 28:114–133, 1981.

[CM77] A. K. Chandra and P. M. Merlin. Optimal implementation of conjunctive queries in relational databases. In *Proc. 9th ACM Symp. on Theory of Computing*, pages 77–90, 1977.

[Com88] K. Compton. 0-1 laws in logic and combinatorics. In I. Rival, editor, *Proc. 1987 NATO Adv. Study Inst. on algorithms and order*, pages 353–383. Reidel, Dordrecht, Netherlands, 1988.

[DD85] K. G. Denbigh and J. S. Denbigh. *Entropy in Relation to Incomplete Knowledge.* Cambridge University Press, Cambridge, UK, 1985.

[Del88] J. P. Delgrande. An approach to default reasoning based on a first-order conditional logic: Revised report. *Artificial Intelligence*, 36:63–90, 1988.

[DG79] B. Dreben and W. D. Goldfarb. *The Decision Problem: Solvable Classes of Quantificational Formulas.* Addison-Wesley, Reading, MA, 1979.

[End72] H. B. Enderton. *A Mathematical Introduction to Logic.* Academic Press, New York, 1972.

[Eth88] D. W. Etherington. *Reasoning with incomplete information.* Morgan Kaufmann Publishers, Inc., 1988.

[Fag76] R. Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41(1):50–58, 1976.

[FHM90] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87(1/2):78–128, 1990.

[Gab84] D. Gabbay. Theoretical foundations for nonmonotonic reasoning in expert systems. In K. R. Apt, editor, *Proceedings of the NATO Advanced Study Institute on logics and models of concurrent systems.* Springer-Verlag, 1984.

[Gai64] H. Gaifman. Concerning measures in first order calculi. *Israel Journal of Mathematics*, 2:1–18, 1964.

[Gef92] H. Geffner. *Default reasoning: causal and conditional theories.* MIT Press, 1992.

[GHK92] A. J. Grove, J. Y. Halpern, and D. Koller. Random worlds and maximum entropy. In *Proc. 7th IEEE Symp. on Logic in Computer Science*, pages 22–33, 1992.

[GJ79] M. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. Freeman and Co., San Francisco, CA, 1979.

[GKLT69] Y. V. Glebskiĭ, D. I. Kogan, M. I. Liogon'kiĭ, and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Kibernetika*, 2:17–28, 1969.

[GKP89] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics—A Foundation for Computer Science*. Addison-Wesley, Reading, MA, 1989.

[GMP90] M. Goldszmidt, P. Morris, and J. Pearl. A maximum entropy approach to nonmonotonic reasoning. In *Proc. National Conference on Artificial Intelligence (AAAI '90)*, pages 646–652, 1990.

[Gol87] S. A. Goldman. Efficient methods for calculating maximum entropy distributions. Master's thesis, MIT EECS Department, 1987.

[Goo92] S. D. Goodwin. Second order direct inference: A reference class selection policy. *International Journal of Expert Systems: Research and Applications*, 1992. To appear.

[GP90] H. Geffner and J. Pearl. A framework for reasoning with defaults. In H. E. Kyburg, Jr., R. Loui, and G. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 245–26. Kluwer Academic Press, Dordrecht, Netherlands, 1990.

[GP92] H. Geffner and J. Pearl. Conditional entailment: bridging two approaches to default reasoning. *Artificial Intelligence*, 53(2–3):209–244, 1992.

[Gra83] E. Grandjean. Complexity of the first–order theory of almost all structures. *Information and Control*, 52:180–204, 1983.

[Hac75] I. Hacking. *The Emergence of Probability*. Cambridge University Press, Cambridge, UK, 1975.

[Hal90] J. Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, 46:311–350, 1990.

[HM87] S. Hanks and S. McDermott. Nonmonotonic logic and temporal projection. *Artificial Intelligence*, 33(3):379–412, 1987.

[Hun89] D. Hunter. Causality and maximum entropy updating. *International Journal of Approximate Reasoning*, 3(1):379–406, 1989.

[Jay57] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, 1957.

[Jay78] E. T. Jaynes. Where do we stand on maximum entropy? In R. D. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*, pages 15–118. MIT Press, Cambridge, MA, 1978.

[Jay82] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proc. IEEE*, 70(9):939–952, 1982.

[Jef68] R. C. Jeffrey. Probable knowledge. In I. Lakatos, editor, *International Colloquium in the Philosophy of Science: The Problem of Inductive Logic*, pages 157–185. North Holland Publishing Co., 1968.

[JL81] Larsen R. J. and Mark M. L. *An introduction to mathematical statistics and its applications.* Prentice-Hall, Englewood Cliffs, NJ, 1981.

[Joh32] W. E. Johnson. Probability: The deductive and inductive problems. *Mind*, 41(164):409–423, 1932.

[KH92] D. Koller and J. Y. Halpern. A logic for approximate reasoning. In B. Nebel, C. Rich, and W. Swartout, editors, *Proc. Third International Conference on Principles of Knowledge Representation and Reasoning (KR '92)*, pages 153–164. Morgan Kaufmann, San Mateo, CA, 1992.

[KLM90] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[KV90] Ph. G. Kolaitis and M. Y. Vardi. 0-1 laws and decision problems for fragments of second-order logic. *Information and Computation*, 87:302–338, 1990.

[Kyb61] H. E Kyburg, Jr. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, Connecticut, 1961.

[Kyb74] H. E. Kyburg, Jr. *The Logical Foundations of Statistical Inference.* Reidel, Dordrecht, Netherlands, 1974.

[Kyb83] H. E. Kyburg, Jr. The reference class. *Philosophy of Science*, 50(3):374–397, 1983.

[Kyb88] Henry E. Kyburg, Jr. Full beliefs. *Theory and Decision*, 25:137–162, 1988.

[Lew79] H. R. Lewis. *Unsolvable Classes of Quantificational Formulas.* Addison-Wesley, New York, 1979.

[Lew80] H. R. Lewis. Complexity results for classes of quantificational formulas. *Journal of Computer and System Sciences*, 21:317–353, 1980.

[Lif89] V. Lifschitz. Benchmark problems for formal non-monotonic reasoning, version 2.00. In M. Reinfrank, J. de Kleer, M. Ginsberg, and E. Sandewall, editors, *Non-Monotonic Reasoning: 2nd International Workshop (Lecture Notes in Artificial Intelligence 346)*, pages 202–219. Springer-Verlag, 1989.

[Lio69] M. I. Liogon'kiĭ. On the conditional satisfiability ratio of logical formulas. *Mathematical Notes of the Academy of the USSR*, 6:856–861, 1969.

[LM90] D. Lehmann and M. Magidor. Preferential logics: the predicate calculus case. In R. Parikh, editor, *Theoretical Aspects of Reasoning about Knowledge: Proc. Third Conference*, pages 57–72, 1990.

[LM92] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.

[Lyn80] J. Lynch. Almost sure theories. *Annals of Mathematical Logic*, 18:91–135, 1980.

[Mak] D. Makinson. General patterns of nonmonotonic reasoning. In D. Gabbay, editor, *Handbook of logic and artificial intelligence and logic programming*, volume 2. Oxford University Press. to appear.

[Mak89] D. Makinson. General theory of cumulative inference. In M. Reinfrank, J. de Kleer, M. Ginsberg, and E. Sandewall, editors, *Non-Monotonic Reasoning: 2nd International Workshop (Lecture Notes in Artificial Intelligence 346)*, pages 1–18. Springer-Verlag, 1989.

[McC80] J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13:1,2, 1980.

[McC86] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28:86–116, 1986.

[MH69] J. M. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In D. Michie, editor, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, UK, 1969.

[Moo85] R. C. Moore. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25:75–94, 1985.

[Mor93] M. Morreau. The conditional logic of generalizations. In *Proceedings of the IJCAI Workshop on Conditionals in Knowledge Representation*, pages 108–118, 1993.

[MP92] Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems*, volume 1. Springer-Verlag, Berlin/New York, 1992.

[Nil86] N. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[PB83] S. J. Provan and M. O. Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal on Computing*, 12:777–788, 1983.

[Pea88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Mateo, CA, 1988.

[Pea89] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In R. J. Brachman, H. J. Levesque, and R. Reiter, editors, *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pages 505–516, 1989. Reprinted in *Readings in Uncertain Reasoning*, G. Shafer and J. Pearl (eds.), Morgan Kaufmann, San Mateo, CA, 1990, pp. 699–710.

[Pea90] J. Pearl. System Z: A natural ordering of defaults with tractable applications to nonmonotonic reasoning. In M. Vardi, editor, *Theoretical Aspects of Reasoning about Knowledge: Proc. Third Conference*, pages 121–135. Morgan Kaufmann, 1990.

[Pol90] J. L. Pollock. *Nomic Probabilities and the Foundations of Induction.* Oxford University Press, Oxford, U.K., 1990.

[Poo89] D. Poole. What the lottery paradox tells us about default reasoning. In R. J. Brachman, H. J. Levesque, and R. Reiter, editors, *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pages 333–340. Morgan Kaufmann, San Mateo, CA, 1989.

[Poo91] D. Poole. The effect of knowledge on belief: conditioning, specificity and the lottery paradox in default reasoning. *Artificial Intelligence*, 49(1–3):282–307, 1991.

[PS72] R. E. Powell and S. M. Shah. *Summability Theory and Applications.* Van Nostrand Reinhold, 1972.

[PV89] J. B. Paris and A. Vencovska. On the applicability of maximum entropy to inexact reasoning. *International Journal of Approximate Reasoning*, 3:1–34, 1989.

[RC81] R. Reiter and G. Criscuolo. On interacting defaults. In *Seventh International Joint Conference on Artificial Intelligence (IJCAI-81)*, pages 270–276, 1981.

[Rei49] H. Reichenbach. *Theory of Probability.* University of California Press, Berkeley, 1949.

[Rei80] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Rog67] H. Rogers, Jr. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, New York, 1967.

[Rot93] D. Roth. On the hardness of approximate reasoning. In *Proc. Thirteenth International Joint Conference on Artificial Intelligence (IJCAI '93)*, pages 613–618, 1993.

[Sav54] L. J. Savage. *Foundations of Statistics.* John Wiley & Sons, New York, 1954.

[Sha] G. Shafer. Personal communication, 1993.

[Sha76] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ, 1976.

[Sha89] L. Shastri. Default reasoning in semantic networks: a formalization of recognition and inheritance. *Artificial Intelligence*, 39(3):285–355, 1989.

[Sto77] L. J. Stockmeyer. The polynomial-time hierarchy. *Theoretical Computer Science*, 3:1–22, 1977.

[Tar51] A. Tarski. *A Decision Method for Elementary Algebra and Geometry.* Univ. of California Press, 2nd edition, 1951.

[THT87] D. S. Touretzky, J. F. Horty, and R. H. Thomason. A clash of intuitions: the current state of nonmonotonic multiple inheritance systems. In *Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, pages 476–482, 1987.

[Tra50] B. A. Trakhtenbrot. Impossibility of an algorithm for the decision problem in finite classes. *Doklady Akademii Nauk SSSR*, 70:569–572, 1950.

[Val79a] L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[Val79b] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8:410–421, 1979.

[Vau54] R. L. Vaught. Applications of the Lowenheim-Skolem-Tarski theorem to problems of completeness and decidability. *Indagationes Mathematicae*, 16:467–472, 1954.