THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Logic meets Probability: Towards Explainable AI Systems for Uncertain Worlds

Logic meets Probability:
Towards Explainable AI Systems for Uncertain Worlds*

Vaishak Belle

University of Edinburgh

United Kingdom

vaishak@ed.ac.uk

Abstract

Logical AI is concerned with formal languages to represent and reason with qualitative specifications; statistical AI is concerned with learning quantitative specifications from data. To combine the strengths of these two camps, there has been exciting recent progress on unifying logic and probability. We review the many guises for this union, while emphasizing the need for a formal language to represent a system's knowledge. Formal languages allow their internal properties to be robustly scrutinized, can be augmented by adding new knowledge, and are amenable to abstractions, all of which are vital to the design of intelligent systems that are explainable and interpretable.

1 Preface

In a seminal paper, McCarthy [1958] put forward a profound idea to realize artificial intelligence (AI) systems: he posited that what the system needs to know could be represented in a *formal language*, and a general-purpose algorithm would then conclude the necessary actions needed to solve the problem at hand. The main advantage is that the representation can be scrutinized and *understood by external observers*, and the system's behavior could be *improved* by making statements to it.

The past 60 years have yielded numerous methods for AI that attempt to learn about the world from data, many of which would arguably distance themselves from McCarthy's *declarative* approach. But as many of these data-intensive approaches are deployed in social environments, such as self-driving cars and domestic robots that care for the elderly, there is a growing concern that these systems need to be explainable, modular, re-usable, and in general, amenable to suggestions from human users [Gunning, 2016]. One cannot help but feel that this emphasis is strongly reminiscent of McCarthy's original vision.

This article reviews approaches that unify logical methods going back to McCarthy and probabilistic methods that have since dominated numerous application areas. At one extreme, we report progress on the use of symbolic representations and reasoning as *proxy* for Bayesian inference and learning. At the other extreme, we review frameworks that tackle uncertainty in a general way (that is, without committing to a single probabilistic model) where meta-reasoning plays a significant role. Our view is that *languages* that lie between these extremes are best equipped to jointly address commonsensical reasoning and learning. Overall, formal languages:

- allow their internal properties to be robustly scrutinized (internally by verification techniques and externally with conversational dialogs);

- can be augmented by adding new knowledge and naturally support operations such as composition or other programmatic constructs; and

- are amenable to abstractions, in the context of, say, describing decisions using high-level concepts;

all of which are vital to the design of intelligent systems that are explainable and interpretable.

2 Uncertain Worlds

One of the main arguments against logical AI is that in practice, there is pervasive uncertainty in almost every domain of interest: these can be in the form of measurement errors (*e.g.* readings from a thermometer), the absence of categorical assertions (*e.g.* smoking may be a factor for cancer, but cancer is not an absolute consequence for smokers), and the presence of numerous "latent" factors, including causes that the modeler may simply not have taken into account, all of which question the legitimacy of the model. The upshot is that on the one hand, logic was seen as an inappropriate tool, as it is "rigid" (sentences always evaluate to true or false), "brittle" (sentences in the knowledge base must be true in all possible worlds) and *discrete* (as opposed to the continuous error profiles for thermometers). On the other, the knowledge of the system, as posited in the declarative approach, may not only be incomplete but may be impossible to specify apriori.

Modeling uncertain worlds needed a rigorous formulation, and this came in the form of probabilistic graphical models (PGMs) [Pearl, 1988]. Given a finite set of random variables $\mathcal{X} = \{x_1, \ldots, x_n\}$ over the event space $\{0, 1\}^n$, a PGM provides the machinery to effectively factorize the joint distribution over $\mathcal{X}$. PGMs supercharged the application of statistical

---

methods in language understanding, vision, robotics and machine learning more generally [Murphy, 2012]. What makes PGMs particularly attractive is that both the probabilities of the variables in a given network, as well as the dependencies themselves can be learnt from data, thereby circumventing the requirement that the model needs to be provided by some omniscient modeler.

3 Back To Logic

Despite the success of PGMs, we observe that they are essentially propositional, but are nonetheless deployed in an inherently relational world. That is, PGMs easily make sense of "flat" data, where atomic events are treated as independent random variables. But in medical records, for example, it makes little sense to treat individual entries on patient symptoms as independent, since it ignores things like relationships between co-occurring symptoms, and symptoms common to family members. This encouraged the design of concept languages for PGMs, culminating in the area of *statistical relational learning* [Getoor and Taskar, 2007]. These formalisms borrow syntactic devices from finite-domain relational logic to define complex interactions between random variables in large-scale PGMs over classes and hierarchies. For example, in Markov logic networks [Richardson and Domingos, 2006], the weighted formula:

$$1.2 \quad \forall x, y \colon D \ [Smokes(x) \land Friends(x, y) \supset Smokes(y)]$$

is indicative of a fairly involved Markov network obtained by grounding the formula wrt a finite set of constants D and assigning a potential proportional to 1.2 to the edges of the network. (The formula expresses the statistical default that friends of smokers are smokers themselves.)

While this can be seen as a purely representational advantage in using logical machinery, it turns out that propositional satisfiability (SAT) technology can be used for probabilistic inference: given a PGM $\mathcal{G}$, an encoding of its structure as a propositional formula ϕ and an encoding of its parameters as a weight function w, computing the probability of an event q in $\mathcal{G}$ amounts to computing the weights of those assignments that satisfy $\phi \land q$. Indeed, there are natural polynomial time reductions from inference to the counting instance of SAT. A variety of exact and approximate solver techniques have since been developed that are known to be very competitive, especially in the presence of hard constraints, deterministic reasoning and logical dependencies [Chavira and Darwiche, 2008; Ermon *et al.*, 2013].

Finally, needless to say, the learning of parameters and dependencies from data have natural analogues in these logical formulations [Richardson and Domingos, 2006].[1]

So, the bridge between propositional logic (possibly defined over a relational template) and PGMs has proven to be insightful and significant.[2] In the sequel, we discuss how this

union can be enriched by turning to more expressive logical languages.

4 Beyond Propositional Logic

Since most of the formal templates for PGMs reduce to propositional logic, it might seem that the use of logical methods is restricted to *discrete random variables* over finite event spaces. However, by lifting the underlying theory from propositional logic to linear arithmetic, we can also naturally reason about *continuous properties*. The idea is that these arithmetic expressions can be used to make statements about the nature of the density function over $\mathbb{R}$: for example, according a weight of .1 to the formula $0 \leq x \leq 1$ can be understood as saying that x is uniformly distributed on the interval [0,1]. Based on a generalized SAT solving for linear arithmetic, we have shown how this intuition leads to exact and approximate inference techniques for continuous and mixed discrete-continuous probability distributions [Belle *et al.*, 2015a; 2015b; 2016b].

Formal templates notwithstanding, the expressive power of PGMs, however, is limited: variables are assumed to be fixed and finite. But this conflates with usage in practice where there is uncertainty about what things are in this world. In vision and language understanding, for example, the system would discover the existence of previously unknown objects from raw data [Russell, 2015].

In principle, then, allowing for the instantiation of possibly infinitely many propositions can provide the machinery to handle applications implicitly involving infinitely many random variables. Among other strategies [Milch *et al.*, 2005], an infinite Herbrand base would enable this, as is possible in probabilistic (logic) programming [Raedt *et al.*, 2007].

To put this to practice, we have enabled *data association* using expressions such as [Nitti *et al.*, 2015b; 2015a]:

$$obspos_{t+1}(\text{ID}) \sim \text{GAUSSIAN}(pos_t(\text{ID}); 0, 1)$$

which says that the object ID's position at time t is observed at $t+1$ with Gaussian error, where ID is instantiated in an *online fashion* as the system plans for its goals in an environment. Analogously, handling objects *without known identifiers* is a common occurrence in language and robotic sensing. For example, in a GRE test, we encounter questions such as:

> *In a group of 10 people, 60 percent have brown eyes. Two people are to be selected at random from the group. What is the probability that neither person selected will have brown eyes?*

Here, the question does not specify the eye color of the remaining 40 percent of the group. The automated solving of such problems using probabilistic programming is considered in our recent work [Dries *et al.*, 2017].

Be that as it may, although piggybacking on a programming language allows some respite in going beyond the closed-world assumption, the fundamental question of when finitely many computational steps are sufficient for inference

---

[1]Learning in formal languages can take other forms [Lake *et al.*, 2015], of course, which are nonetheless related to the approaches reviewed here. Our development is mainly in the context of formal languages with a truth theory (that is, logics).

[2]Although discussions in the article are restricted to probabilistic reasoning, we have argued elsewhere that there are also venues for

bridging logical methodologies and numerical optimization [Mladenov *et al.*, 2017].

over infinite theories remains poorly understood. For example, in the above smokes-friends rule, what type of queries can we efficiently compute when D is infinite? What about the case of function symbols? We investigate these questions in [Belle, 2017a; 2017b].

5 Meta Reasoning

Formulating PGMs as weighted logical formulas is an elegant approach to tightly decouple the constraints and dependencies from the probabilistic parameters. But at a conceptual level, there is little gained as we are still embedded in the framework of standard probability theory.

When it comes to long-lived AI systems, however, especially reasoners and learners that potentially run forever, we need to be able to reason about probabilistic events in a more general way. For example, we may need to compare the probabilities of hypothetical outcomes, or analyze the behavior of non-terminating probabilistic programs. In some cases, we may even be ignorant about the actual probabilities of events! Indeed, McCarthy and Hayes [1969] argued thus:

> 1. *It is not clear how to attach probabilities to statements containing quantifiers in a way that corresponds to the amount of conviction people have.*
> 2. *The information necessary to assign numerical probabilities is not ordinarily available. Therefore, a formalism that required numerical probabilities would be epistemologically inadequate.*

In sum, we would need a specification of (probabilistic) belief that can be partial or incomplete, in keeping with whatever information is available about the application domain. We would additionally need to construct a logic to reason about statements involving beliefs, not only for the current epoch, but also for all possible executable behaviors of the system.

On the subject of the expressiveness of such a logic, we take the view that as a language, it should be as general as possible: we are then in a position to investigate fragments that enjoy reasonable computational properties. In fact, at the very least, such a language should fully subsume what is already representable by purely probabilistic or purely logical accounts.

Of course, having uncertainty about the actual probabilities of events raises the following question: what does the system know about its own state? If it cannot ascribe a unique probability to a proposition q, does it know of its ignorance? To be able to talk about such meta-beliefs, we turn to *formal epistemology* [Fagin *et al.*, 1995; Reiter, 2001], where one is interested in knowledge and ignorance, and how that changes over actions.

Building on earlier efforts to combine first-order logic and probability [Bacchus, 1990], perhaps the most general formalism for dealing with degrees of belief in formulas, and in particular, with how belief should evolve in the presence of noisy sensing and acting is the account proposed by Bacchus, Halpern, and Levesque (BHL) [1999]. The BHL scheme is embedded in a rich theory of action called the *situation calculus* [Reiter, 2001], but now extended for nondeterminism and probabilistic likelihoods.[3]

The continued relevance of the situation calculus, and related knowledge representation languages, can be explained as follows. As a dialect of first-order logic, it is well-equipped to embrace domain formalizations that require classes, hierarchies and quantification. Given a language with atoms $\mathcal{P}$, actions $\mathcal{T}$ and sequences $\mathcal{Z}$ built from $\mathcal{T}$, we presuppose the existence of *possible worlds* $\mathcal{W}$ that map $\mathcal{P} \times \mathcal{Z}$ to $\{0, 1\}$. In other words, a world is a *tree*. Semantically, this allows us to not only reason about past events, but also about hypotheticals in the future. By defining epistemic operators that are interpreted over these worlds, we can talk about meta-beliefs over infinite histories [Belle and Lakemeyer, 2017].

We reiterate that the advantage of investigating results in a general language is that more constrained (and thus, often tractable) frameworks follow as special cases. Indeed, the situation calculus has proven helpful in providing formal underpinnings to automated planning formalisms [Lin and Reiter, 1997; McIlraith and Son, 2002], updates in closed- and open-world databases [Hariri *et al.*, 2013], and dynamic Bayesian networks over complex actions [Hajishirzi and Amir, 2010], among others. Interestingly, tractability can often be decoupled in terms of the expressiveness of the knowledge base and that of the action model [Belle and Lakemeyer, 2011].

Below we emphasize recent results and extensions to the BHL account, which arguably demonstrates its generality and compatibility with a number of threads of research in knowledge representation and machine learning.

Compatibility and Generality The BHL account motivates the necessary extensions to the situation calculus for handling probabilistic nondeterminism, but it was left open how closely the account matches important real-world applications like *localizing* a robot. In [Belle and Levesque, 2014b; 2015b], we developed a logical theory where a number of standard and non-standard variations pertaining to localization could be tested.

In [Belle and Lakemeyer, 2017], we have also studied the compatibility between the BHL scheme and purely logical accounts of knowledge and action [Reiter, 2001].

Continuity One of the major points of divergence between logical accounts and learning frameworks is the absence of continuous distributions in the former, and their abundance in the latter. The BHL scheme is certainly restricted to discrete random variables. In [Belle and Levesque, 2013a], we showed how this limitation could be lifted.

Reduction Theorems The situation calculus provides a very rich setting for describing dynamics, such as allowing for *context-specificity*: for example, the motion model of a

---

[3]Probabilities are but one approach to quantify uncertainty, perhaps the most common and an obvious choice for many modeling situations. Of course, there are other approaches. But as discussed by BHL, much of their framework can be exported to alternate formalisms, such as Dempster-Shafer belief functions [Shafer, 1990]. This is possible by defining an alternate measure of belief, and replacing the likelihoods with non-probabilistic functions that support an alternate rule of belief update.

robot could be conditioned on whether the floor is slippery. But this richness comes at the cost of making inference more challenging. Nonetheless, given queries about outcomes after actions, as needed for automated planning, we have shown that actions can be either "compiled" away, or "injected" to the current state [Belle and Levesque, 2013b; 2014a]. (See [Kaelbling and Lozano-Pérez, 2013] for similar results.)

Programming Languages As an illustration of how the above reduction theorems can be applied to PGMs, the design of a query system was studied in [Belle and Levesque, 2014c]. Going further, [Belle and Levesque, 2015a] developed a programming language that can be used for goal-directed planning over meta-beliefs. Such programs can be seen as partial iterative policies [Reiter, 2001].

Characterizing Beliefs A crucial question ignored by many prior accounts on unifying logic and probability is this: what *exactly* do the *beliefs of the system* look like? Believing an event *p*, for example, does not the preclude the system from also believing $p \wedge q$, and so, as modelers, we would need a way to succinctly characterize knowledge and ignorance. We investigated this issue in [Belle *et al.*, 2016a].

Multiple Agents AI systems these days can rarely be expected to only reason about their own internal model, as they are often deployed in social environments, that is, in the presence of other autonomous systems, or even humans. There is a large body of work on the model theory and implementation of meta-beliefs with multiple agents [Fagin *et al.*, 1995; Muise *et al.*, 2015], but in the context of the situation calculus, reduction theorems and axiomatizations have been studied in [Kelly and Pearce, 2008; Liu and Wen, 2011; Belle and Lakemeyer, 2015; 2014a; 2014b]. In [Belle and Levesque, 2015b], a preliminary investigation on extending the BHL scheme to multiple agents is considered.

In sum, with these developments, we obtain *languages* and *compositional programming abstractions* to reason about subjective probabilities and meta-beliefs in dynamical domains. The open and exciting avenues here include a better understanding of tractable fragments, and how algorithmic proposals from statistical relational learning can be lifted for such general languages.

6 Discussion and Conclusions

Proponents of the declarative paradigm argue that expecting systems to infer the complex and interactive nature of the real-world in a fully unsupervised manner is a fantasy, whereas those building systems that learn from data argue that the provision of a sufficiently nuanced AI system by human experts is unrealistic. The unification of logic and probability may provide the best of both worlds, especially when there is partial knowledge about the domain itself. This article reviewed frameworks that show how such a unification is achievable, and is suggestive of a more integrative view of the two AI camps. These frameworks essentially sketch a spectrum of possibilities, ranging from bottom-up approaches that are symbolic analogues to Bayesian learning, to full-blown modal logics that allow us to talk about meta-beliefs and histories. Languages that lie between these extremes are perhaps

best equipped to jointly address commonsensical reasoning and learning.

There is much we do not yet understand on relating the representation language of the AI system and the (possibly human) decision maker [Gunning, 2016]. Moreover, the question of how one can effectively learn representations for expressive frameworks remains mostly open. But that question, albeit challenging, may provide the necessary bridge between incomplete and quantitative specifications, on the one hand, and meta-beliefs and noisy observations, on the other. Logical AI has a long and distinguished history in dealing with thorny topics in commonsense reasoning [Davis, 2014]; integrating these with probabilistic models may prove fruitful. We reiterate that formal languages enable scrutiny, augmentations and abstractions, and when additionally armed with meta-reasoning, systems can not only describe which of their beliefs influenced their decisions, but also how this internal model would change over interactions with the environment and other agents. These features are vital to the design of intelligent and commonsensical systems that are explainable and interpretable, but also social.

Finally, we hope to have clarified that the characteristics of rigidity, discreteness and brittleness attributed to logic is a misunderstanding. Logic provides a language for talking about the world and understanding what information is conveyed by expressions in that language; the language is indeed rigid and discrete, but its worlds certainly need not be.

References

[Bacchus *et al.*, 1999] F. Bacchus, J. Y. Halpern, and H. J. Levesque. Reasoning about noisy sensors and effectors in the situation calculus. *Artificial Intelligence*, 111(1–2):171 – 208, 1999.

[Bacchus, 1990] F. Bacchus. *Representing and reasoning with probabilistic knowledge*. MIT Press, 1990.

[Belle and Lakemeyer, 2011] V. Belle and G. Lakemeyer. On progression and query evaluation in first-order knowledge bases with function symbols. In *Proc. IJCAI*, 2011.

[Belle and Lakemeyer, 2014a] V. Belle and G. Lakemeyer. Multi-agent only knowing in dynamic systems. *Journal of Artificial Intelligence Research*, 49, 2014.

[Belle and Lakemeyer, 2014b] V. Belle and G. Lakemeyer. On the progression of knowledge in multiagent systems. In *KR*, 2014.

[Belle and Lakemeyer, 2015] V. Belle and G. Lakemeyer. Semantical considerations on multiagent only knowing. *Artif. Intell.*, 223:1–26, 2015.

[Belle and Lakemeyer, 2017] V. Belle and G. Lakemeyer. Reasoning about probabilities in unbounded first-order dynamical domains. In *IJCAI*, 2017.

[Belle and Levesque, 2013a] V. Belle and H. J. Levesque. Reasoning about continuous uncertainty in the situation calculus. In *Proc. IJCAI*, 2013.

[Belle and Levesque, 2013b] V. Belle and H. J. Levesque. Reasoning about probabilities in dynamic systems using goal regression. In *Proc. UAI*, 2013.

[Belle and Levesque, 2014a] V. Belle and H. J. Levesque. How to progress beliefs in continuous domains. In *Proc. KR*, 2014.

[Belle and Levesque, 2014b] V. Belle and H. J. Levesque. A logical theory of robot localization. In *Proc. AAMAS*, 2014.

[Belle and Levesque, 2014c] V. Belle and H. J. Levesque. PREGO: An action language for belief-based cognitive robotics in continuous domains. In *Proc. AAAI*, 2014.

[Belle and Levesque, 2015a] V. Belle and H. J. Levesque. ALLEGRO: Belief-based programming in stochastic dynamical domains. In *IJCAI*, 2015.

[Belle and Levesque, 2015b] V. Belle and H. J. Levesque. A logical theory of localization. In *Studia Logica*. 2015.

[Belle et al., 2015a] V. Belle, G. Van den Broeck, and A. Passerini. Hashing-based approximate probabilistic inference in hybrid domains. In *UAI*, 2015.

[Belle et al., 2015b] V. Belle, A. Passerini, and G. Van den Broeck. Probabilistic inference in hybrid domains by weighted model integration. In *IJCAI*, 2015.

[Belle et al., 2016a] V. Belle, G. Lakemeyer, and H. J. Levesque. A first-order logic of probability and only knowing in unbounded domains. In *Proc. AAAI*, 2016.

[Belle et al., 2016b] V. Belle, A. Passerini, and G. Van den Broeck. Component caching in hybrid domains with piecewise polynomial densities. In *AAAI*, 2016.

[Belle, 2017a] V. Belle. Open-universe weighted model counting. In *AAAI*, 2017.

[Belle, 2017b] V. Belle. Weighted model counting with function symbols. In *UAI*, 2017.

[Chavira and Darwiche, 2008] M. Chavira and A. Darwiche. On probabilistic inference by weighted model counting. *Artif. Intell.*, 172(6-7):772–799, 2008.

[Davis, 2014] E. Davis. *Representations of commonsense knowledge*. Morgan Kaufmann, 2014.

[Dries et al., 2017] A. Dries, A. Kimmig, J. Davis, V. Belle, and L. De Raedt. Solving probability problems in natural language. In *IJCAI*, 2017.

[Ermon et al., 2013] S. Ermon, C. P. Gomes, A. Sabharwal, and B. Selman. Embed and project: Discrete sampling with universal hashing. In *NIPS*, pages 2085–2093, 2013.

[Fagin et al., 1995] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about knowledge*. MIT Press, 1995.

[Getoor and Taskar, 2007] L. Getoor and B. Taskar, editors. *An introduction to statistical relational learning*. MIT Press, 2007.

[Gunning, 2016] D. Gunning. Explainable artificial intelligence (xai). Technical report, DARPA/I20, 2016.

[Hajishirzi and Amir, 2010] H. Hajishirzi and E. Amir. Reasoning about deterministic actions with probabilistic prior and application to stochastic filtering. In *Proc. KR*, 2010.

[Hariri et al., 2013] B. B. Hariri, D. Calvanese, M. Montali, G. De Giacomo, R. De Masellis, and P. Felli. Description logic knowledge and action bases. *JAIR*, 46:651–686, 2013.

[Kaelbling and Lozano-Pérez, 2013] L. P. Kaelbling and T. Lozano-Pérez. Integrated task and motion planning in belief space. *I. J. Robotic Res.*, 32(9-10):1194–1227, 2013.

[Kelly and Pearce, 2008] R. F. Kelly and A. R. Pearce. Complex epistemic modalities in the situation calculus. In *KR*, 2008.

[Lake et al., 2015] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.

[Lin and Reiter, 1997] F. Lin and R. Reiter. How to progress a database. *Artificial Intelligence*, 92(1-2):131–167, 1997.

[Liu and Wen, 2011] Y. Liu and X. Wen. On the progression of knowledge in the situation calculus. In *IJCAI*, 2011.

[McCarthy and Hayes, 1969] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, pages 463–502, 1969.

[McCarthy, 1958] J. McCarthy. Programs with common sense. In *Proceedings of the Symposium on the Mechanization of Thought Processes*, National Physiology Lab, Teddington, England, 1958.

[McIlraith and Son, 2002] S. A. McIlraith and T. C. Son. Adapting Golog for composition of semantic web services. In *KR*, pages 482–496, 2002.

[Milch et al., 2005] B. Milch, B. Marthi, S. J. Russell, D. Sontag, D. L. Ong, and A. Kolobov. BLOG: Probabilistic models with unknown objects. In *Proc. IJCAI*, pages 1352–1359, 2005.

[Mladenov et al., 2017] M. Mladenov, V. Belle, and K. Kersting. The symbolic interior point method. In *AAAI*, 2017.

[Muise et al., 2015] C. Muise, V. Belle, P. Felli, S. McIlraith, T. Miller, A. Pearce, and L. Sonenberg. Planning over multi-agent epistemic states: A classical planning approach. In *Proc. AAAI*, 2015.

[Murphy, 2012] K. Murphy. *Machine learning: A probabilistic perspective*. The MIT Press, 2012.

[Nitti et al., 2015a] D. Nitti, V. Belle, T. De Laet, and L. De Raedt. Sample-based abstraction for hybrid relational MDPs. In *European Workshop on Reinforcement Learning*, 2015.

[Nitti et al., 2015b] D. Nitti, V. Belle, and L. De Raedt. Planning in discrete and continuous Markov Decision Processes by probabilistic programming. In *ECML*, 2015.

[Pearl, 1988] J. Pearl. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann, 1988.

[Raedt et al., 2007] L. De Raedt, A. Kimmig, and H. Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *Proc. IJCAI*, pages 2462–2467, 2007.

[Reiter, 2001] R. Reiter. *Knowledge in action: Logical foundations for specifying and implementing dynamical systems*. MIT Press, 2001.

[Richardson and Domingos, 2006] M. Richardson and P. Domingos. Markov logic networks. *Machine learning*, 62(1):107–136, 2006.

[Russell, 2015] S. J. Russell. Unifying logic and probability. *Commun. ACM*, 58(7):88–97, 2015.

[Shafer, 1990] G. Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4(5–6):323 – 362, 1990.