Scientific Theories and Artificial Intelligence

Philippe Desjardins-Proulx[0,1,2], Timothée Poisot[2], and Dominique Gravel[1]

[0] *email: philippe.d.proulx@gmail.com*
[1] *Université de Sherbrooke, Canada.*
[2] *Université de Montréal, Canada.*

October 22, 2017

Keywords: *Artificial Intelligence, Theoretical Biology, Theoretical Ecology, Machine learning, Knowledge Representations.*

0 Summary

Artificial Intelligence presents an important paradigm shift for science. Science is traditionally founded on theories and models, most often formalized with mathematical formulas handcrafted by theoretical scientists and refined through experiments. Machine learning, an important branch of modern Artificial Intelligence, focuses on learning from data. This leads to a fundamentally different approach to model-building: we step back and focus on the design of algorithms capable of building models from data, but the models themselves are not designed by humans. This is even more true with deep learning, which requires little engineering by hand and is responsible for many of Artificial Intelligence's spectacular successes [30]. In contrast to logic systems, knowledge from a deep learning model is difficult to understand, reuse, and may involve up to a billion parameters [10]. On the other hand, probabilistic machine learning techniques such as deep learning offer an opportunity to tackle large complex problems that are out of the reach of traditional theory-making. It is possible that the more intuition-like [30] reasoning performed by deep learning systems is mostly incompatible with the logic formalism of mathematics. Yet recent studies have shown that deep learning can be useful to logic systems and vice versa. Success at unifying different paradigms of Artificial Intelligence from logic to probability theory offers unique opportunities to combine data-driven approaches with traditional theories. These advancements are susceptible to impact significantly biological sciences, where dimensionality is high and limit the investigation of traditional theories.

1 A.I. and knowledge representation

Science would greatly benefit from a unification of Artificial Intelligence with traditional mathematical theories. Modern research at the intersection of logic, probability theory, and fuzziness yielded rich representations increasingly capable of formalizing scientific knowledge. Such formal corpus could both include hand-crafted theories from Einstein's $e = mc^2$ to the Breeder's equation [38], but also harness modern A.I. algorithms for testing and learning.

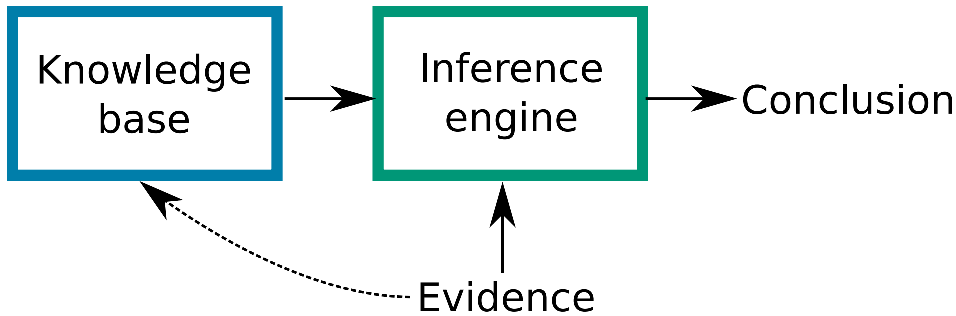

Figure 1: In his 1959's "Programs with Common Sense", McCarthy established the importance of separating knowledge and reasoning [34]. The knowledge base stores knowledge, while the inference engine exploits it, along with evidence, to reach conclusions. McCarthy believed the knowledge base should store rules in some form of predicate logic, but the idea of separation of knowledge and reasoning holds with other representations as well (Bayesian networks, detailed in figure 2, are probabilistic knowledge bases [11]). The dotted line represents automatic theory revision, where evidence is used not to a answer query but to discover new knowledge or revise existing theories.

Comprehensive synthesis is difficult in fields like biology, which have not been reduced to a small set of formulas. For example, while we have a good idea of the underlying forces driving evolution, we struggle to build effective predictive models of molecular evolution [18]. This is likely because selection changes in time and space [4], which brings population, community, and ecosystem ecology into the mix. Ecology also has a porous frontier with evolution: speciation is a common theme in community ecology theory [12].

From a theoretical perspective, work to formalize scientific theories would reveal much about the nature of our theories. Surely, scientific theories require more flexibility than mathematical corpora of knowledge, which are based on pure logic. From a practical standpoint, a formal representation both offers ways to test large corpora of knowledge and extend it with A.I. techniques. This is arguably the killer feature of a formal representation of scientific knowledge: allowing A.I. algorithms to search for revisions, extensions, and discover new rules. This is not a new ambition. Generic techniques for rule discovery were well-established in the 1990s [37]. Unfortunately, these techniques were based on pure logic, and purely probabilistic approaches to revision cannot handle mathematical theories. Recent experiences in linguistics has shown that building a knowledge base capable of handling several problems at the same time yielded better results than attacking the problem in isolation because of the problems' interconnectedness [45]. Biology, as a complex field made of more-or-less arbitrary subfields, could gain important insights from unified approach to knowledge combining A.I. techniques with traditional mathematical theories.

2 A quick tour of knowledge representations

Deep learning is arguably the dominant approach in probabilistic machine learning, a branch of A.I. focused on learning models from data [17]. The idea of deep learning is to learn multiple levels of compositions. If we want to learn to classify images for instance, the first layer of the deep learning network will read the input, the next layer will capture lines, the next layer will capture contours based on lines, and then more complex shapes based on contours, and so on [17]. In short, the layers of the network begin with simple concepts, and then compose more complicated concepts from simpler ones [5]. Deep learning has been used to solve complex computer science problems like playing Go at the expert level [41], but it is also used for more traditional scientific problems like finding good candidate molecules for drugs, predicting how changes in the genotype affect the phenotype [31], or just recently to solving the quantum many-body problem [9].

In contrast, traditional scientific theories and models are mathematical, or logic-based. Einstein's $e = mc^2$ established a logical relationship between energy e, mass m, and the speed of light c. This mathematical knowledge can be reused: in any equation with energy, we could replace e with mc^2. This ability of mathematical theories to establish precise relationships between concepts, which can then be used as foundations for other theories, is fundamental to how science grows and forms an interconnected corpus of knowledge. Furthermore, these theories are compact and follow science's tradition of preferring theories as simple as possible. There are many different foundations for logic systems. Predicate logic is a good starting point: it is based on predicates, which are functions of terms to a truth value. For example, the predicate $PreyOn$ could take two species, a location, and return true if the first species preys on the second at that location, like $PreyOn(Wolverine, Squirrel, Quebec)$. Terms are either *constants* such as 1, π, or $Wolverine$, *variables* that range over constants, such as x or *species*, or *functions* that map terms to terms, such as additions, multiplication, integration, differentiation. In $e = mc^2$, the equal sign $=$ is the predicate, e and m are variables, c and 2 are constants, and there are two functions: the multiplication of m by c^2 and the the exponentiation of c by 2. The key point is that such formalism lets us describe compact theories and understand precisely how different concepts are related. Complex logic formulas are built by combining predicates with connectives such as negation $\neg$, "and" $\wedge$, "or" $\vee$, "implication" $\Rightarrow$. We could have a rule to say that predation is asymmetrical $s_x \neq s_y \wedge PreyOn(s_x, s_y, l) \Rightarrow \neg PreyOn(s_y, s_x, l)$, or define the classical Lotka-Volterra:

$$\frac{dx}{dt} = \alpha x - \beta xy \wedge \frac{dy}{dt} = \delta xy - \gamma y, \tag{1}$$

where x and y are the population sizes of the prey and the predator, respectively, α, β, δ, γ are constants, and the time differential d/dt, multiplication and subtraction are functions. Equality ($=$) is the sole predicate in this formula. Both predicates are connected via $\wedge$ ("and"). Not all logic formulas have mathematical functions. Simple logic rules such as $Smoking(p) \Rightarrow Cancer(p)$ ("smoking causes cancer") are common in expert systems.

Artificial Intelligence researchers have long being interested in logic systems capable of scientific discoveries, or simply capable of storing scientific and medical knowledge in a single coherent system (Figure 1). DENDRAL, arguably the first expert system, could form hypotheses to help identify new molecules using its knowledge of chemistry [32]. In the 1980s, MYCIN was used to diagnose blood infections (and did so more accurately than professionals) [8]. Both systems were based on logic, with MYCIN adding a "confidence factor" to its rules to model uncertainty. Other expert systems were based on probabilistic graphical models [27], a field that unites graph theory with probability theory to model the conditional dependence structure of random variables [27, 2]. For example, Munin had a network of more than 1000 nodes to analyze electromyographic data [14], while PathFinder assisted medical professional for the diagnostic of lymph-node pathologies [22] (Figure 2). While these systems performed well, they are both too simple to store generic scientific

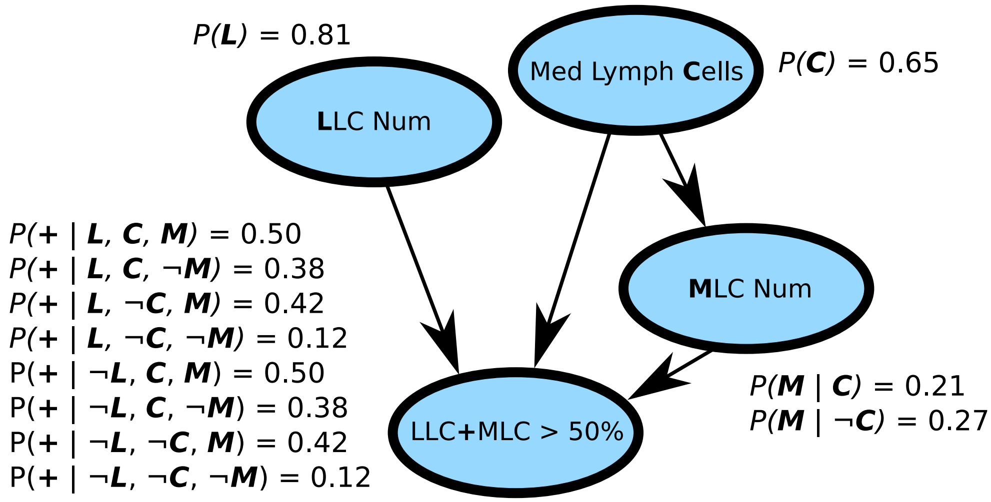

Figure 2: A Bayesian network with four binary variables and possible conditional probability tables. These four nodes were taken from PathFinder, a Bayesian network with more than 1000 nodes used to help diagnose blood infections [22]. The nodes represent four variables related to blood cells and are denoted by a single character (in bold in the figure): $C, M, L, +$. All variables are binary, and negation is denoted with $\neg$. Since $P(\neg x|\mathbf{y}) = 1 - P(x|\mathbf{y})$, we need only $2^{|Pa(x)|}$ parameters per nodes, with $|Pa(x)|$ being the number of parents of node x. The structure of Bayesian networks both highlights the conditional independence assumptions of the distribution and reduces the number of parameters for learning and inference. Example query: $P(L, \neg C, M, \neg +) = P(L)P(\neg C)P(M|\neg C)P(\neg + |L, \neg C, M) = 0.81 \times (1 - 0.65) \times 0.27 \times (1 - 0.42) = 0.044$. See [11] for a detailed treatment of Bayesian networks and [27] for a more general reference on probabilistic graphical models.

knowledge and too static to truly unify Artificial Intelligence with scientific research. The ultimate goal is to have a representation rich enough to encode both logic-mathematical and probabilistic scientific knowledge.

3 Beyond monolithic systems

In terms of representation, expert systems generally used a simple logic system, not powerful enough to handle uncertainty, or purely probabilistic approaches unable to handle complex mathematical formulas. In terms of flexibility, the expert systems were hand-crafted by human experts. After the experts established either the logic formulas (for logic systems like DENDRAL) or probabilistic links (in the case of systems like Munin), the expert systems act as static knowledge bases, capable of answering queries but unable of discovering new rules and relationships. While no system has completely solved these problems yet, much energy has been put in unifying logic-based systems with probabilistic approaches [16]. Also, several algorithms have been developed to learn new logic rules [37], find the probabilistic structure in a domain with several variables [46], and even transfer knowledge between tasks [36]. Together, these discoveries bring us closer to the possibility of flexible knowledge bases contributed both by human experts and Artificial Intelligence algorithms. This has been made possible in great part by efforts to unify three distinct languages: probability theory, predicate logic, and fuzzy logic (Fig 3).

The core idea behind unified logic/probabilistic languages is that formulas can be weighted, with higher values meaning we have greater certainty in the formula. In pure logic, it is impossible to violate a single formula. With weighted formulas, an assignment of concrete values to variables is only *less likely* if it violates formulas. The higher the weight of the formula violated, the less likely the assignment is. It is conjectured that all perfect numbers are even ($\forall x : Perfect(x) \Rightarrow Even(x)$), if we were to find a single odd perfect number, that formula would be refuted. It makes sense for mathematics but for many disciplines, such as biology, important principles are only expected to be true *most* of the times. To illustrate, in ecology, predators generally have a larger body weight than their preys, which can expressed in predicate logic as $PreyOn(predator, prey) \Rightarrow M(predator) > M(prey)$, with $M(x)$ being the body mass of x. This is obviously false for some assignments, for example $predator : greywolf$ and $prey : moose$. However, it is useful knowledge that underpins many ecological theories [44]. When our domain involves a great number of variables, we should expect useful rules and formulas that are not always true.

The idea of weighted formulas is not new. Markov logic, invented a decade ago, allows for logic formulas to be weighted [39, 13]. It supports algorithms to add weights to existing formulas given a data-set, learn new formulas or revise existing ones, and answer probabilistic queries. For example, Yoshikawa et al. used Markov logic to understand how events in a document were time-related [45]. Their research is a good case study of interaction between traditional theory-making and artificial intelligence. The formulas they used as a starting point were well-established logic rules to understand temporal expressions. From there, they used Markov logic to weight the rules, adding enough flexibility to their system to beat the best approach of the time. Brouard et al. [7] used Markov logic to understand gene regulatory network, noting how the resulting model provided clear insights, in contrast to more traditional machine learning techniques. Expert systems can afford to make important sacrifices to flexibility in exchange for a simple representation. Yet, a system capable of representing a large body of scientific knowledge will require a great deal of flexibility to accommodate various theories. While a step in the right direction, even Markov logic may not be powerful enough.

4 Case study: The niche model

To show some of the difficulties of representing scientific knowledge, we will build a small knowledge base for an established ecological theory: the niche model of trophic interactions [44]. The first iteration of the niche model posits that all species are described by a niche position N (their body size for instance) in the $[0, 1]$ interval, a diet D in the $[0, N]$ interval, and a range R such that a species preys on all species with a niche in the $[D - R/2, D + R/2]$ interval. We can represent these ideas with three formulas:

$$\forall x, y : \neg PreyOn(x, y), \tag{2a}$$

$$\forall x : D(x) < N(x), \tag{2b}$$

$$\forall x, y : PreyOn(x, y) \Leftrightarrow D(x) - R(x)/2 < N(y) \wedge N(y) < D(x) + R(x)/2, \tag{2c}$$

where $\forall$ reads *for all* and $\Leftrightarrow$ is logical equivalence (it is true if and only if both sides of the operator have the same truth value, so for example $False \Leftrightarrow False$ is true and

$True \Leftrightarrow False$ is false). As pure logic, this knowledge base makes little sense. Formula 2a is obviously not true all the time. It is mostly true, since most pairs of species do not interact. We could also add that cannibalism is rare $\forall x : \neg PreyOn(x, y)$ and that predator-prey are generally asymmetrical $\forall x, y : PreyOn(x, y) \Rightarrow \neg PreyOn(y, x)$. In hybrid probabilistic/logic approaches like Markov logic, these formulas would have a weight that essentially defines a marginal probability [13, 25]. Formulas that are often wrong are assigned a lower weight but can still provide useful information about the system. The second formula says that the diet is smaller than the niche value. The last formula is the niche model: species x preys on y if and only if species y's niche is within the diet interval of x.

So far so good! Using Markov logic networks and a data-set, we could learn a weight for each formula in the knowledge base. This step alone is useful and provide insights into which formulas hold best. With the resulting weighted knowledge base, we could make probabilistic queries and even attempt to revise the theory automatically. We could find, for example, that the second rule does not apply to parasites or some group and get a revised rule such as $\forall x : \neg Parasite(x) \Rightarrow D(x) < N(x)$. However, Markov logic networks struggle when the predicates cannot easily return a simple true-or-false truth values. For example, let's say we wanted to express the idea that when populations are small and have plenty of resources, they grow exponentially [29].

$$\forall x, l, t : SmallP(x, l, t) \text{ and } Resources(x, l, t) \Rightarrow P(x, l, t + 1) = G(x) \times P(x, l, t), \quad (3)$$

where $P(x, l, t)$ is the population size of species x in location l at time t, G is the rate of growth, $SmallP$ is whether the species has a small population and $Resources$ whether it has resources available. The problem with hybrid probabilistic/logic approach is that predicates do not capture the inherent vagueness well. We can establish an arbitrary cutoff for what a small population is, for example by saying that if it is less than 10% the average population size for the species, it is small. Similarly, resource availability is not a binary thing, there is a world of grey between starvation and satiety. Perhaps worst of all, the prediction that $P(x, l, t + 1) = G(x) \times P(x, l, t)$ is almost certainly never be exactly true. If we predict 94 rabbits and observe 93, the formula is false. Weighted formulas help us understand *how often a rule is true*, but in the end the formula has to give a binary truth value: true or false, there is no place for vagueness.

Fuzzy sets and many-valued ("fuzzy") logics were invented to handle vagueness [47, 24, 6, 3]. In practice, it simply means that predicates can return any value in the $[0, 1]$ closed interval instead of only true and false. It is used in both probabilistic soft logic [26, 1] and deep learning approaches to predicate logic [48, 23]. For our formula 3, $SmallP$ could be defined as $1 - P(x, l, t)/P_{max}(x)$, where $P_{max}(x)$ is the largest observed population size for the species. $Resources$ could take into account how many preys are available, and $P(x, l, t+1) = G(x) \times P(x, l, t)$ would return a truth value based on how close the observed population size is the predicted population size. Fuzzy logic then defines how operators such as *and* and $\Rightarrow$ behave with fuzzy values.

Both Markov logic networks and probabilistic soft logic define a probability distribution over logic formulas, but what about the large number of probabilistic models? For example, the niche model has a probabilistic counter-part defined as [43]:

$$\forall x, y : PPreyOn(x, y) = \alpha \times \exp\left[-\left(\frac{N(y) - D(x)}{R(x)/2}\right)^2\right], \quad (4)$$

where $PPreyOn(x, y)$ is the probability that x preys on y. Again, this formula is problematic in Markov logic because we cannot easily force the equality into a binary true-or-false, but fuzziness can help model the nuance of probabilistic predictions.

5 Where's our unreasonably effective paradigm?

Wigner's *Unreasonable Effectiveness of Mathematics in the Natural Sciences* led to important discussions about the relationship between physics and its main representation [42, 21]. The Mizar Mathematical Library and the Coq library [33] host tens of thousands of mathematical propositions to help build and test new proofs. In complex domains with many variables, Halevy et al. argued for the *Unreasonable Effectiveness of Data* [19], noting that simple algorithms, when fed large amount of data, would do wonder. High-dimensional problems like image imputation, where an algorithm has to fill missing parts from an image, require hundred of thousands of training images to be effective. Goodfellow et al. noted that roughly 10 000 data-points per possible labels were necessary to train deep neural networks [17]. These approaches are unsatisfactory for fields like biology where theories and principles are seldom exact. We cannot afford the pure logic-based knowledge representations favoured by mathematicians and physicists, and fitting a model to data is a different task than building a corpus of interconnected knowledge.

Fortunately, we do not need to choose between mathematical theories, probabilistic models, and learning. New inventions such as Markov logic networks and probabilistic soft logic are moving Artificial Intelligence toward rich representations capable of formalizing and even extending scientific theories. This is a great opportunity for synthesis. There are still problems: inference is often difficult in those rich representations. Recently, Garnelo et al. [15] designed a prototype to extract logic formulas from a deep learning system, while Hu et al. [23] created a framework to learn predicate logic rules using deep learning. Both studies used flexible fuzzy predicates and weighted formulas while exploiting deep learning' ability to model complex distributions via composition. The end result is a set of clear and concise weighted formulas supported by deep learning for scalable inference. The potential for science is important. Not only these new researches allow for deep learning to interact with traditional theories, but it opens many exciting possibilities, like the creation of large databases of scientific knowledge. The only thing stopping us from building a unified corpus of, say, ecological knowledge, is that normal pure-logic systems are too inflexible. They do not allow imperfect, partially-true theories, which are fundamental to many sciences. Recent developments in Artificial Intelligence make these corpora of scientific knowledge possible for complex domains, allowing us to combine a traditional approach to theory with the power of Artificial Intelligence.

It is tempting to present deep learning as a threat to traditional theories. Yet, there is a real possibility that the union of Artificial Intelligence techniques with mathematical theories is not only possible, but would help the integration of knowledge across various disciplines. Otherwise, short of discovering a small set of elegant theories, what is our plan to combine ideas from ecosystem ecology, community ecology, population ecology, and evolution?

6 Acknowledgements

PDP has been funded by an Alexander Graham Bell Graduate Scholarship from the National Sciences and Engineering Research Council of Canada, an Azure for Research award from

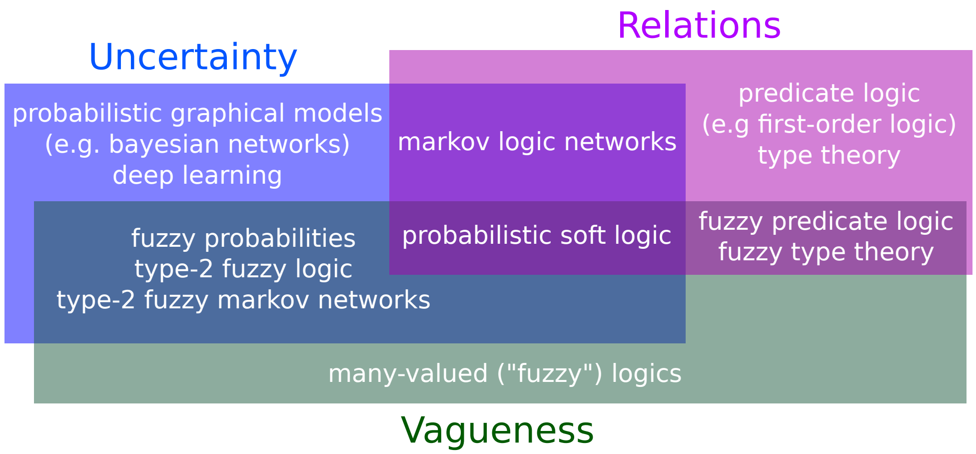

Figure 3: Various reasoning languages and their ability to model uncertainty, vagueness, and relations. The size of the rectangles has no meaning. **In the blue rectangle**: languages capable of handling uncertainty. Probabilistic graphical models combine probability theory with graph theory to represent complex distributions [27]. Deep learning is, strictly speaking, more general than its usual probabilistic interpretation, but it is arguably the most popular probabilistic Artificial Intelligence approach at the moment [17]. Alternatives to probability theory for reasoning about uncertainty include possibility theory and Dempster-Shafer belief functions, see [20] for an extended discussion. **In the green rectangle**: Fuzzy logic extends standard logic by allowing truth values to be anywhere in the $[0, 1]$ interval. Fuzziness models vagueness and is particularly popular in linguistics, engineering, and bioinformatics, where complex concepts and measures tend to be vague by nature. See [28] for a detailed comparison of probability and fuzziness. **In the purple rectangle**: relations, as in: mathematical relations between objects. Even simple mathematical ideas, such as the notion that all natural numbers have a successor ($\forall x \exists y : y = x + 1$), requires relations. *Predicate* and *Relation* are synonymous in this context. Alone, these reasoning languages are not powerful enough to express scientific ideas. We must thus focus on what lies at their intersection. Type-2 Fuzzy Logic is a fast-expanding [40] extension to fuzzy logic, which, in a nutshell, models uncertainty by considering the truth value itself to be fuzzy [35, 49]. Markov logic networks [39, 13] extends predicate logic with weights to unify probability theory with logic. Probabilistic soft logic [26, 1] also has formulas with weights, but allow the predicates to be fuzzy, i.e. have truth values in the $[0, 1]$ interval. Some recent deep learning studies also combine all three aspects [15, 23].

Microsoft, and benefited from the Hardware Donation Program from NVIDIA. DG is funded by the Canada Research Chair program and NSERC Discovery grant. TP is funded by an NSERC Discovery grant and an FQRNT Nouveau Chercheur grant.

References

[1] SH Bach, M Broecheler, B Huang, and L Getoor. Hinge-loss markov random fields and probabilistic soft logic. arXiv: 1505.04406, 2015.

[2] D Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[3] L Behounek, P Cintula, and P Hájek. Introduction to mathematical fuzzy logic. In P Cintula, P Hájek, and C Noguera, editors, *Handbook of Mathematical Fuzzy Logic volume 1*, chapter 1, pages 1–101. College Publications, London, 2011.

[4] G Bell. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Phil. Trans. R. Soc. B*, 365:87–97, 2010.

[5] Y Bengio, A Courville, and P Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[6] M Bergmann. *An introduction to many-valued and fuzzy logic*. Cambridge University Press, 2008.

[7] C Brouard, C Vrain, J Dubois, D Castel, MA D, and F d'Alche Buc. Learning a markov logic network for supervised gene regulatory network inference. *BMC Bioinformatics*, 14(1):273, 2013.

[8] B Buchanan and EH Shortliffe. *Rule-based Expert Systems: The Mycin experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, 1984.

[9] G Carleo and M Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[10] A Coates, B Huval, T Wang, DJ Wu AY Ng, and B Catanzaro. Deep learning with COTS HPC systems. *Journal of Machine Learning Research Workshop and Conference Proceedings*, 28(3):1337–1345, 2013.

[11] A Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.

[12] P Desjardins-Proulx and D Gravel. A complex speciation-richness relationship in a simple neutral model. *Ecology and Evolution*, 2(8):1781–1790, 2012.

[13] P Domingos and D Lowd. *Markov Logic: An Interface Layer for Artificial Intelligence*. Morgan & Claypool Publishers, 2009.

[14] S Andreassen et al. Evaluation of the diagnostic performance of the expert EMG assistant MUNIN. *Electroencephalogr Clin Neurophysiol*, 101(2):129–144, 1996.

[15] M Garnelo, K Arulkumaran, and M Shanahan. Towards deep symbolic reinforcement learning. arXiv:1609.05518v2, 2016.

[16] L Getoor, N Friedman, D Koller, A Pfeffer, and B Taskar. Probabilistic relational models. In L Getoor and B Taskar, editors, *Introduction to Statistical Relational Learning*. MIT Press, 2007.

[17] I Goodfellow, Y Bengio, and A Courville. *Deep Learning*. MIT Press, 2016.

[18] MW Hahn. Toward a selection theory of molecular evolution. *Evolution*, 76(3):255–265, 2008.

[19] A Halevy, P Norvig, and F Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24:8–12, 2009.

[20] JY Halpern. *Reasoning about Uncertainty*. The MIT Press, 2003.

[21] RW Hamming. The unreasonable effectiveness of mathematics. *The American Mathematical Monthly*, 87(2):81–90, 1980.

[22] D Heckerman and B Nathwani. An evaluation of the diagnostic accuracy of Pathfinder. *Computers and Biomedical Research*, 25(1):56–74, 1992.

[23] Z Hu, X Ma, Z Liu, E Hovy, and EP Xing. Harnessing deep neural networks with logic rules. arXiv:1603.06318, 2016.

[24] P Hájek. *Metamathematics of Fuzzy Logic*. Springer Netherlands, 1998.

[25] D Jain. Knowledge Engineering with Markov Logic Networks: A Review. In *DKB 2011: Proceedings of the Third Workshop on Dynamics of Knowledge and Belief*, 2011.

[26] A Kimmig, SH Bach, M Broecheler, B Huang, and L Getoor. A short introduction to probabilistic soft logic. In *Proceedings of the NIPS Workshop on Probabilistic Programming*, 2012.

[27] D Koller and N Friedman. *Probabilistic Graphical Models*. The MIT Press, 2009.

[28] B Kosko. Fuzziness vs probability. *Int J General Systems*, 17:211–240, 1990.

[29] M Kot. *Elements of Mathematical Ecology*. Cambridge University Press, 2001.

[30] Y LeCun, Y Bengio, and G Hinton. Deep learning. *Nature*, (521):436–444, 2015.

[31] MKK Leung, A Delong, B Alipanahi, and BJ Frey. Machine learning in genomic medicine: A review of computational problems and data sets. *Proceedings of the IEEE*, 104(1):176–197, 2016.

[32] RK Lindsay, BG Buchanan, EA Feigenbaum, and J Lederberg. Dendral: A case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence*, 61(2):209–261, 1993.

[33] The Coq development team. *The Coq proof assistant reference manual*. LogiCal Project, 2004. Version 8.0.

[34] J McCarthy. Programs with common sense. 1959.

[35] JM Mendel and RI Bob John. Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, 10(2):117–127, 2002.

[36] L Mihalkova, T Huynh, and RJ Mooney. Mapping and revising markov logic networks for transfer learning. In *Proceedings of the 22nd Conference on Artificial Intelligence*, pages 608–614, 2007.

[37] S Muggleton and L de Raedt. Inductive logic programming: Theory and methods. *The Journal of Logic Programming*, 19-20:629–679, 1994.

[38] SH Rice. *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer, 2004.

[39] M Richardson and P Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107–136, 2006.

[40] A Sadeghian, JM Mendel, and H Tahayori. *Advances in Type-2 Fuzzy Sets and Systems*. Springer, 2014.

[41] D Silver, A Huang, CJ Maddison, A Guez, L Sifre, G van den Driessche, J Schrittwieser, I Antonoglou, V Panneershelvam, M Lanctot, S Dieleman, D Grewe, J Nham, N Kalchbrenner, I Sutskever, T Lillicrap, M Leach, K Kavukcuoglu, T Graepel, and D Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

[42] E Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13(1):1–14, 1960.

[43] RJ Williams, A Anandanadesan, and ND Martinez. The probabilistic niche model reveals the niche structure and role of body size in a complex food web. *PLOS One*, 5(8):e12092, 2010.

[44] RJ Williams and ND Martinez. Simple rules yield complex food webs. *Nature*, 404:180–183, 2000.

[45] K Yoshikawa, S Riedel, M Asahara, and Y Matsumoto. Jointly identifying temporal relations with markov logic. 2009.

[46] C Yuan and B Malone. Learning optimal bayesian networks: A shortest path perspective. *Journal of Artificial Intelligence Research*, 48:23–65, 2013.

[47] LA Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[48] T Zahavy, N Ben-Zrihem, and S Mannor. Graying the black box: Understanding DQNS. arXiv: 1602.02658, 2016.

[49] J Zeng and Z-Q Liu. Type-2 fuzzy markov random fields and their application to handwritten chinese character recognition. *IEEE Transactions on Fuzzy Systems*, 16(3):747–760, 2008.