THE REASONER

VOLUME 18, NUMBER 2
MARCH 2024

thereasoner.org
ISSN 1757-0522

CONTENTS

Editorial 9

Features 9
Interview with Vaishak Belle 9
Conditional beliefs aren't conditional probabilities . . 12

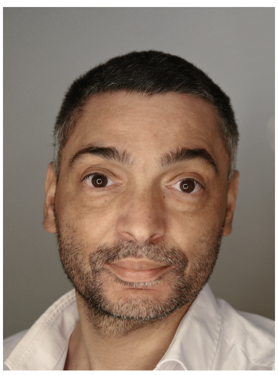
The Reasoner Speculates 13
Benefits of cybernetic models in philosophy 13

Dissemination Corner 14
SMARTEST 14
BRIO 15

EDITORIAL

Dear Reasoners,

I am very happy to welcome you to this new issue of The Reasoner. It opens with my interview with Vaishak Belle and then features Jon Williamson arguing that conditional beliefs aren't conditional probabilities, Ferenc András suggesting the benefits of cybernetic models in philosophy, and ends with the dissemination of two exciting research projects.



As a short introduction to my guest, Vaishak is Reader at the University of Edinburgh, an Alan Turing Fellow, and a Royal Society University Research Fellow. He is one of the key contributors to the field known as neurosymbolic AI which, as you will see, tries to make the most of the two traditions in the field.

Since this topic is of great interest to many readers of The Reasoner, we are soliciting a *Focussed Issue* on it (see for instance here and here for two examples). Please send short proposals with the list of contributors to `hykel.hosni@unimi.it`.

Before leaving you to the interview, I'd like to thank warmly Vaishak Belle for his time and for the generosity with which he shared his views with us.

HYKEL HOSNI
Logic, Uncertainty, Computation and Information Lab,
University of Milan

FEATURES

Interview with Vaishak Belle

HYKEL HOSNI: You are an expert in neurosymbolic AI, which is very much in the news these days.

VAISHAK BELLE: I'm not sure I would consider myself an expert, but I do find myself very interested in the area. One of the reasons I suppose it is difficult for anybody to declare themselves an expert in this field is because it is rapidly changing.

HH: Can you tell us what it is all about?

VB: In the early days, the term "neurosymbolic AI" was usually referred to formalisms that allowed neural architectures in logical languages: representations that combine some aspects of neural networks in a logic, especially fuzzy logic, which allows for real-valued truth. However, recently, neurosymbolic AI is better understood as formalisms that combine aspects of logical reasoning with deep learning.

HH: And since deep learning encompasses a variety of methods. . .

VB: . . . there is no single agreed-upon definition, indeed! This obviously opens up the space for a wide range of solutions. For example, perhaps the most common kind of solution typically seen in robotics applications is when you have deep learning systems for vision and audio and language that are interpreted using some kind of control framework – e.g., a symbolic automated planning framework might interact with one of these deep learning outputs to help the robot operate purposefully in its domain. This is often regarded as a loose coupling between logic and deep learning because it only allows a limited sense in which the semantics of the logical language captures what is happening inside the deep learning system.

HH: Can you give an example in which the connection between logic and deep learning is tighter?

VB: Sure. A deeper integration involves exploring ways to enable joint training or reasoning between logical systems and deep learning architectures. For instance, a popular area of inquiry which has recently been attracting interest is based on the idea of modifying the loss function of deep learning systems with logical formulas. This modification allows the distributions learned by the neural networks to capture the semantics of those formulas and constraints. Consequently, predictions can be ensured to adhere to physical and geometric properties of the domain. Another type of coupling involves extracting facts and relations from the web, say, using a deep learning system. These facts can then populate a database or an ontology over which a logical query-driven engine is built. Yet another area of inquiry is investigating the possibility of extracting symbolic structures, such as computer programs, from neural architectures. The idea is that these programs could be interpreted by humans and therefore become, in some sense, explainable. HH: What varieties of logic and deep learning are relevant in those applications?

VB: All the examples I just mentioned involve an interesting and often complicated mixing of model theory on the logic side and statistical learning and geometry from the deep learning side. Therefore, there is a very real possibility that neurosymbolic AI lays the foundations of a new type of AI that involves the best of both worlds.

HH: That may come as a surprise to (classical) logicians in the first place!

VB: It should be noted that the learning of logical formulas and the use of logic in machine learning are long-standing areas of research in their own right. Perhaps the most popular representation of this is statistical relational learning, which combines machine learning and probabilistic logical languages, such as relational Bayesian networks and probabilistic logic programs. This is yet another facet of how logic and deep learning can combine: by using a probabilistic logical formalism, distributions learned by deep learning models could be directly embedded in a logical language.

HH: Regular readers of The Reasoner will recognise this, as Felix Weitkämper has been running a column on statistical relational learning for quite some time!

VB: That's great! Although there are plenty of academic communities working on neuro-symbolic AI, the industry has been taking a very serious interest as well, especially considering that deep learning on its own seems to be data-hungry and often struggles in safety-critical applications, owing to issues such as distribution drift, and generally the lack of guarantees that comes with that. Thus, verifying the robustness of neural networks is an important topic, as is explainability owing to its black-box nature. Finally, because the use of machine learning in the real world doesn't often immediately reduce simply to prediction, there is an inherent need to combine structures and symbolic grammars with neural networks.

HH: Many are tempted by the analogy with dual systems of cognition where deep learning embodies the fast and highly fallible "system 1", whereas logic is asked to play the role of the slow and arguably reliable "system 2", in the terminology made popular by Daniel Kahneman.

VB: With the advent of large language models and their capacity for confabulations, the idea that perhaps one could use symbolic reasoners as a post hoc solution for correctness and consistency has been circulating. For example, Wolfram Alpha recently started to feature an integration with ChatGPT so that mathematically correct answers can be provided for questions of a mathematical or computational nature. The general idea is that whatever is uttered in natural language is processed by ChatGPT and converted to a form that can be interpreted by Wolfram Alpha, after which the symbolic solver returns the solution.

HH: One characteristic feature of the current AI spring is that it is driven by private companies who nonetheless appear to make significant scientific contributions. Of course I am thinking of DeepMind...

VB: Indeed! The recent AlphaGeometry approach by Google DeepMind, which made sensational news in the New York Times (17 January), attempts to solve geometry problems from the International Mathematical Olympiad. The key idea here too is to use the language model to create formal constructs and have a symbolic engine interpret these constructs to not only solve them but potentially give signals back to the language model for more effective problem-solving.

HH: Do you agree with those who think that this is yet another game changer from DeepMind?

VB: It should be noted that for AlphaGeometry to work they needed to generate a hundred million synthetic data examples. Such an effort might not be possible for everyone. However, as more and more of such synthetic samples are generated for numerous domains on which the language models are trained, it might eventually be possible to use one of these models in different settings, provided you have an appropriate symbolic reasoner to ensure that the responses are correct. Thus, neurosymbolic AI has a promising future, it seems.

HH: I can see expectations being really high! Can you tell us about your background?

VB: I completed my undergraduate degree in India in a field that could be considered closer to software engineering than computer science. I then pursued my master's degree as part of an Erasmus Mundus program between Germany and Italy. This was perhaps my first exposure to formal approaches.

HH: Was it the classic love at first sight?

VB: Not sure! Initially, I wasn't entirely convinced of their applicability in the real world. In India, the emphasis was often more on software engineering, as graduates were being trained for services-oriented software companies. It took me quite some time to rewire my way of thinking to develop an appreciation for theory.

HH: But I guess that happened quite quickly. Were you set to pursue the academic path after graduation?

VB: At that point I wasn't necessarily keen on an academic career per se. To be honest, I didn't quite know what it entailed, but I did entertain the notion that a job involving writing and thinking all day was a fun career, if such a thing was possible at all.

HH: That sounds very familiar!

VB: I was also into science fiction, so in some sense, I was interested in artificial intelligence fairly early on. It was with my master's degree and the start of a Ph.D. that I slowly transitioned to becoming familiar with logic. Somewhat bizarrely, because I lacked a formal background, I ended up teaching myself about modal logic first, and never covered propositional or first-order logic in any course. Interestingly, in contrast, my master's thesis was on face recognition. Logic appealed to me, but when I began working on it, I still recognized the value of the machine-learning way of thinking, especially in the sense of extracting patterns from data through the training process.

HH: It is interesting to see how the hybrid approach to AI you are pursuing in your research is rooted in your very personal trajectory. So, after your masters, you started a PhD in Germany. What was its topic?

VB: At the beginning of my Ph.D.I was quite interested in interactive epistemology as it was making its way into game theory. Then I began to consider whether those kinds of formalisms could be useful in AI. Ultimately, this led me to work on epistemic and dynamic logic for my Ph.D. And few years into my Ph.D., I began to wonder if it would be useful to examine languages that combined the capabilities of logic and probabilistic reasoning.

HH: Which you took forward in your postdoctoral years.

VB: Exactly. After the PhD my work focussed on integrating probability and logic and, ultimately, on learning and logic.

HH: Which lead you to venturing in neurosymbolic AI. Many PhDs with this background would be attracted to industry careers. Did you consider that?

VB: I did, briefly. However, I was fortunate enough to obtain a postdoc, which seemed like a more natural choice. We –my partner and me– had to move to Canada for this, but ultimately it was the start of a wonderful adventure.

HH: You spent two years in Canada. What happened next?

VB: I held a postdoc fellowship in Belgium. After a little more than a year of that, I began applying for academic positions and was fortunate to obtain one here in Edinburgh, where I have remained since.

HH: What is the most exciting problem you are working on at the moment?

VB: At the moment, I am very interested in mechanisms for extracting logical knowledge using neural architectures, as well as the ways in which logical knowledge can be embedded as constraints in neural architectures. In some sense, both of these are begging the question: what kind of semantics and formal machinery best allows the representation of neural computations with logical knowledge? How does this affect using logical solvers as part of this architecture? And where should we draw the line, from a scalability point of view, to either rely completely on neural computations or completely on logical computations? There clearly needs to be a boundary that allows us to go back and forth to have the most effective way to reason about neurosymbolic computations. And that's a broad open challenge that I find very interesting. Ultimately, I suppose, it really is a way to get at the dichotomy between deduction, abduction and other kinds of deliberative computation versus reactive complications such as predictions from a neural network.

HH: Fascinating. We have covered a lot, but I am sure there is more in the pipeline! Can you tell us about your plans for the future?

VB: I have a couple of projects related to large language models and logic that I am looking into. But I suppose what is really keeping me occupied right now is organizing some of the ideas I mentioned in a kind of unified framework and seeing how this evolves in the next few months.

HH: Sure. Is there any advice you would like to give to PhD students who just started or are about to start?

VB: Two things stand over the others: do good science, and trust the process. To get started on doing good science, the nature of which can vary wildly from area to area, we need to have an understanding of the background literature and the foundations (e.g., keep a few textbooks in hand, and not just the latest works to study the lineage), and keeping the motivation and need for this result in mind, are the best ways to have a clear-cut goal, from which you can define a path.

HH: I can imagine them now being impatient to hear how they turn this into exciting research

VB: Of course, this is only the beginning! The results will come gradually, as long as we put in the work in a disciplined manner, and are consistent, and take a scholarly approach to the related work. It is important to be honest about the kind of results we desire and to acquire the necessary skills along the way. The nature of research is that it is often unfamiliar, and mistakes will be made. However, by learning from these mistakes, acquiring new knowledge, and constantly ensuring that we are not repeating past mistakes or reinventing the work of others, we can ensure satisfaction with the end result, whether positive or negative. The related work can provide guidance and feedback too, if studied properly, similar to that of a supervisor. So, it is important to see how others in the community approach the problems, their intuition, expertise, and knowledge, and work along those lines with attention to detail.

HH: You mentioned trusting the process.

VB: The process should be as enjoyable as the outcome. After all, science is supposed to be fun, so make sure that curiosity is not hindered and that the enjoyment of the process is as rewarding as the end result. Even if the desired outcome is not achieved, a lot of valuable lessons will have been learned along the way, making it easier to tackle the next problem.

HH: Indeed! Finally, can you share any reading suggestions for anyone serious about neurosymbolic AI?

VB: There are a couple of edited volumes on neurosymbolic AI, which, although not immediately accessible for a reader who isn't working on AI, still give away the most important ideas emerging in the space right now. But to me, to really get to the heart of Neurosymbolic AI, it might be helpful to look at some major books discussing common sense and the need for combining logic and learning more generally. For instance, a book by Gary Marcus and Ernie Davis titled *Rebooting AI*, and *Machines like us* by Hector Levesque and Ron Brachman. I think they capture the essence of what is required for a commonsensical AI agent to perform in a way that is reasonable with our view of the world. And even though they don't directly speak about current developments in Neurosymbolic AI, I believe they are relevant. From a technical perspective, they strongly advocate for why people should be considering the integration of logic and learning.