The Reasoner

Volume 17, Number 5
September 2023

thereasoner.org
ISSN 1757-0522

Contents

Guest Editorial

Dear Reasoners, I am delighted to introduce Emiliano Lorini, a senior researcher at the Centre National de la Recherche Scientifique (CNRS) and co-head of the Logic, Interaction, Language and Computation (LILaC) team at the Institut de Recherche en Informatique de Toulouse (IRIT). Emiliano Lorini stands as a prominent scholar in the field of logic and artificial intelligence. His work on modeling interactions among cognitive agents exhibits strong connections to disciplines such as economics, philosophy, and cognitive science. The interview delves into contemporary challenges inherent in developing explainable AI and the distinctive role logicians play in this endeavor. Furthermore, we discussed practical aspects of the current academic landscape and its interface with real-world issues.

Ekaterina Kubyshkina
Logic, Uncertainty, Computation and Information Group,
University of Milan

Features

Interview with Emiliano Lorini

Ekaterina Kubyshkina: Can you please describe us your academic background?

Emiliano Lorini: I have a PhD in Cognitive Science from University of Siena (Italy) and a master degree in Computer Science from Toulouse University (France). But since the time of my master and the first year of my PhD I oriented myself towards logic. I am CNRS researcher since 2009, actually senior researcher. I am co-head of the LILaC team (Logic, Interaction, Language and Computer) at the Institut de Recherche en Informatique de Toulouse (IRIT), one of the main labs in computer science in France. My main expertise is in the development of formal languages and semantics, based on logic and game theory, for modeling the reasoning, decision-making and emotions of both human and artificial agents as well as several aspects of social interaction such as the concepts of norm, trust, responsibility, power, persuasion and social influence. I mainly focus on the axiomatic and complexity aspects and on the decision procedures for such languages and semantics (e.g., for satisfiability checking, model checking and planning) in order to automate the reasoning and decision-making of artificial agents that are designed to interact and communicate with other (artificial or human) agents. My work has a distinctly interdisciplinary character in strong interaction, both at the conceptual and formal level, with the models of reasoning, decision and interaction developed in philosophy, law and economics.

EK: Considering your specific interest on modelling the reasoning of human and AI agents, what are your perspectives on the Natural Language Processing (NLP) applications such as ChatGPT? Do you perceive them as valuable tools, or do you

identify any potential challenges or drawbacks in their utilization?

EL: The performance of "generative" AI systems such as Chat-GPT is impressive. They show how pure machine learning models trained on big data, without explicit background knowledge encoded in a symbolic way, are capable of conversation that is highly informative and fully understandable to humans. However, we are still far from having a statistical machine that learns how to reason generally and accurately, and in many situations counterfactually by exploiting the power of imagination, so that it can effectively solve problems not previously encountered during the learning phase. General problem solving relies, among other things, on an intelligent system's inferential, imaginative and decisional capabilities that allow it to creatively perform new tasks in a goal-directed manner. These capabilities cannot be fully acquired through statistical learning or reinforcement learning. Logic is needed.

Another fundamental problem is normative self-regulation. Is a system such as Chat-GTP able to learn the subtle conceptual distinction between benevolent persuasion and malevolent forms of persuasion (e.g., deception, manipulation) and to refrain from behaving in an obnoxious, unethical way? Should ethical, legal and social norms be designed and modeled in a top-down manner and then integrated into such systems, so that they can learn under normative constraints and fulfill the normative expectations of the users? I think in order to make systems such as Chat-GPT able to understand norms and to comply with them, AI researchers need to develop hybrid models and methods which combine (bottom-up) machine learning with (top-down) logic-based reasoning.

EK: You have been working in France for many years. How would you describe the current situation in the field of logic there? What are the popular trends? Which topics attract the most attention from students and young researchers?

EL: Logic plays an important role in both theoretical computer science and artificial intelligence, as a tool for formal verification of computer programs and for modeling and automating the reasoning, planning and decision-making of intelligence systems. Machine learning is the dominant paradigm in AI nowadays, both from the point of view of models and applications, especially deep neural learning models. The latter attract students and young researchers with a computer science background. So, the main challenge for logicians working in the computer science and AI areas, is to come up with interesting and meaningful integrations of logic-based models and machine learning. This is the main trend and scientific challenge not only in France, but also at an international level, in our field.

EK: Could you please recommend any specific Master's or PhD programs in France or abroad that are well-suited for a young student with a keen interest in integrating logic-based approaches with the study of AI?

EL: I think Toulouse is a great place for that. The AI department at the Institut de Recherche en Informatique de Toulouse (IRIT) (https://www.irit.fr/en/departement/dep-artificial-intelligence/) has a long standing tradition in logic-based AI. A Master's program in AI is proposed by Université Toulouse III Paul Sabatier and PhD positions are regularly offered under the supervision of the researchers from the AI department.

The Artificial and Natural Intelligence Toulouse Institute (ANITI, https://aniti.univ-toulouse.fr/) was recently created, with special emphasis on so-called "hybrid AI" aimed at combining symbolic methods with machine learning methods based on statistics. It offers PhD as well as postdoc positions.

Another great place is the "Centre de Recherche en Informatique de Lens" (CRIL) at the Université d'Artois (https://www.cril.univ-artois.fr/).

EK: You are currently the principal investigator of the ANR PRCE CoPains project titled "Cognitive Planning in Persuasive Multimodal Communication." Could you please provide more information about this project and the results you have achieved so far?

EL: In the CoPains project we have studied the concept of cognitive planning at different levels: theoretical, algorithmic and applicative. Roughly speaking, cognitive planning is the problem of computing a sequence of speech (or more generally communicative) acts aimed at changing, influencing, affecting the cognitive state of the interlocutor. It relies on Theory of Mind, since for being successful, the planning agent must have a model of the interlocutor's cognitive state. We have developed a logic of mental attitudes for providing a formal specification of this problem, studied its computational complexity and developed a number of algorithms for solving it. We have also designed and implemented an intelligent conversational agent based on our theoretical model, in collaboration with researchers from linguistics and cognitive psychology, and with an industrial partner specialized in conversational and emotional artificial intelligence.

EK: This sounds great! What is the implementation of this conversational agent? Could you say more on the theoretical and practical benefits of its usage?

EL: We have applied the conversational agent to two fields: sport counseling and game playing. In the first application, the agent had to motivate the user to practice a physical activity regularly and then, after having motivated her/him, to help the user to find the sport more in line with her/his interests and preferences. The interaction with the psychologists has been fundamental. The logical specification of the agent was made in conformity with motivational interviewing, a counseling method used in clinical psychology for eliciting behavior change. In the second application, the agent was used in the context of a collaborative card game in which Theory of Mind plays a crucial role. We have shown that the performance of the human-machine (H-M) team outperforms that of the human-human (H-H) team.

Implementing our logic-based theory of cognitive planning has been a beneficial and interesting experience. We have been able to show that the theory can be "engineered", in the sense that it can be implemented in a real machine and used in the context of a practical application. To achieve this objective, we spent quite a lot of time to identify an interesting and sufficiently expressive fragment of the logic of mental attitudes and to automate it using existing SAT and QBF solvers.

EK: Logic is a significant field of research in both philosophy and computer science. From your perspective, what are the fruitful outcomes of the dialogue between philosophers and computer scientists? What are the promising advancements from this dialogue?

EL: Computer scientists are mainly interested in computabil-

ity and complexity of a given logic. Philosophers are mainly interested in its conceptual and axiomatic properties, and in the clarity of its formal semantics. The main added value of a dialogue between philosophers and computer scientists is to come up with a more exhaustive and in-depth analysis covering both aspects of the logic under development.

EK: So, do you believe that advancing current research in logic in both fields necessitates a greater presence of philosophers within computer science departments and an increased involvement of computer scientists within philosophy departments?

EL: I strongly believe in interdisciplinary research especially in the field of AI. The presence of philosophers in this field and their collaboration with computer scientists is fundamental. Logical investigation of concepts such as explanation, causality, trust and trustworthiness, responsibility, morality cannot be done accurately and comprehensively without the aid of philosophers with a training in logic.

EK: In your opinion, what are the main advantages of modelling cognitive attitudes in formal settings?

EL: Cognitive attitudes such as beliefs, preferences, desires, ethical values and intentions have properties. For example, intentions of a rational agent (causally) depend on its preferences and beliefs. Logic, thanks to its axiomatics and semantics, can help to clarify and to have a better understanding of such properties. Moreover, logic is useful for modeling and then implementing agents that are expected to reason and make decisions in conformity with certain standards of rationality and to efficiently comply with certain legal and ethical norms. Such rationality standards are usually expressed by making reference to cognitive attitudes of both epistemic type (e.g., beliefs) and motivational type (e.g., preferences, values and intentions), and to their interrelations.

EK: In this context, would you characterize logic as assuming a normative role, a purely descriptive role, or perhaps a combination of both?

EL: We need a combination of the two, normative and descriptive. In many applications, we would like the artificial agent to adhere to normative standards of rationality, given its unbounded reasoning capabilities. But, we would also like the artificial agent to have a model of the human's imperfect rationality, limited reasoning capabilities and cognitive biases.

EK: You have done significant work in the field of explainable AI. In your opinion, what role does logic play in the current development of AI?

EL: The role of logic in the field of explainable AI (XAI) is fundamental. Researchers in this field are interested in understanding, formalizing and then applying concepts such as causal and counterfactual explanation, abductive explanation, actual causality. I firmly believe logic, in combination with probability theory, is the right tool and methodology to make this research successful. Logical languages and formal semantics for causal reasoning and counterfactuals have been developed for decades both in philosophical logic and in the area of knowledge representation and reasoning (KR). I think research in XAI can get great inspiration from such previous works. This is already taking place in the field.

EK: I noticed in your CV that you are currently working on a book titled Logics for multi-agent systems. Could you please tell us about the main objectives of this book and its intended audience?

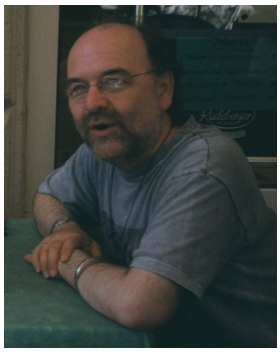
EL: The book is devoted to providing a synthetic and systematic overview of existing logics for modeling interaction in a multi-agent system with a special emphasis on STIT logic (the "logic of seeing to it that").

EK: Recently, several academic events were organized with the aim of drawing attention to the current geopolitical situation. For example, the "Logic4Peace" conference, where IRIT was one of the co-organizing institutions. Do you believe that academic scholars, especially logicians, can play a particular role in addressing real-world problems? If so, what actions can we take, and where should we focus our attention?

EL: I think it is extremely important, as "producers" and disseminators of knowledge and given the societal implications of our research, to actively promote ethical values that we consider a priority (e.g., pacifism, equity and fairness). Machine ethics is a subfield of artificial intelligence which takes this issue into serious consideration. I think logic, especially deontic logic and logics for ethical reasoning, can play an important role in machine ethics. They can help to model ethical values and norms and to implement them in real machines in order to endow them with normative and ethical reasoning capabilities.

EK: Thank you very much for this interview and your replies!

The Burali-Forti Paradox



Any article about the Burali-Forti paradox has to begin with an apology. There has been far too much written about the Burali-Forti paradox and much of what has been written is *terrible*. So why should this article be any different? The answer is that readers may find its approach to be more in harmony with the modern appreciation of the fertility of type-theory.

The Burali-Forti paradox, like lots of other paradoxes, is a proof that something is impossible. (By some definitions a paradox is a nothing more than a proof of $\bot$). The point I wish to emphasise below is that there may be lots of different ways of describing the impossibility, so that if one is to fully understand the situation, one needs to engage with all those different ways. The thought here is that the different ways are not competing resolutions of the contradiction—Mathematics is not firefighting after all—but rather are *different takes on the underlying mathematics*.

There is a famous trope about the five blind men and the elephant. My friend Kelsang Rabten, a Buddhist monk, tells me that in his tradition the meaning of this parable is taken to be that if you have only one teacher you will have only one insight.

The Burali-Forti paradox is of course an elephant, if a rather abstract one, and there are plenty of abstractions groping it. One of them is a set-theoretic foundationalist. If you are such a person (as far too many people are), so that for you—in the final analysis—everything is a set, you will have no option but to see Burali-Forti as a theorem about sets; you will see the

elephant as a proof that the collection of ordinals (however implemented) cannot be a set. This is in all the textbooks, and it's true—in some particularly thin sense of 'true'. The problem for us here and now is not that it isn't *true*; the problem is that it isn't particularly *enlightening*. It's not so much an explanation as an artefact of the equipment used to contrive the explanation. And—since that equipment is set-theory—it tells you more about set theory than it does about the Burali-Forti paradox: artefacts never tell you about the world, they only tell you about the equipment that you are using to investigate that world.

The particular blind man talking to you now is a set theorist who—because of his peculiar early conditioning history—was compelled to think about (and think *through*) the Quine systems, which are—admittedly—set theories, but are set theories that are sufficiently different from ZF-like theories to bring one up short with the thought that *There Might Be Another Way*. He was also exposed to the stimulus of Theoretical Computer Science through being a postdoc in a Computer Science department. Anyone who has been through that has a chance of groping parts of the elephant that set-theoretic foundationalism cannot reach. One is given a chance to reëxamine the Mathematics to which the Burali-Forti paradox beckons us.

Ordinals are a kind of generalisation of natural numbers; natural numbers with an additional `limit` constructor. One can usefully think of the ordinals as an end-extension of the naturals. What do natural numbers do? What are they *for*? They measure the lengths of lists. In most typed programming languages `list`s are a polymorphic data type. For two distinct types `a` and `b` the types `a-list` and `b-list` are distinct. We say that the data type `list` is *polymorphic*. However the datatype of natural numbers that measure the lengths of those lists is always taken to be *monomorphic*. *Prima facie*, I suppose, naturals ought to be polymorphic too—since they arise by abstracting away from a polymorphic datatype—but it seems pretty clear that it is actually safe to take them to be monomorphic. The natural numbers that measure the lengths of `a-list`s are the same natural numbers that measure the lengths of `b-list`s. Take-home thought: naturals *prima facie* ought to be polymorphic but it turns out to be OK to take them to be monomorphic. Keep this thought in mind when approaching ordinals.

So: what about ordinals? Are they monomorphic too? Well, finite ordinals are natural numbers and *they* are monomorphic. So far so good. Right from the dawn of ordinal arithmetic Cantor knew that the natural order relation on ordinals is wellfounded, so that, for any ordinal α, the ordinals below α form a wellordering and that wellordering will have an ordinal. Now you don't have to be a paid-up type theorist to think that perhaps—whatever the ADT of an ordinal α—the ADT of the ordinal of the wellordering of ordinals-below-α might be distinct from the ADT of the ordinal α ... an ADT somehow *derived* from the ADT of α, and with an intimate relation to that ADT, but not actually *identical* to it. Thus ordinals (like naturals) are in principle polymorphic, and for the same reason: they *start off* polymorphic, just as natural numbers did. Might it be safe to take them, too, to be monomorphic? It does seem to be perfectly safe to think of countable ordinals as monomorphic ... however the Burali-Forti paradox tells us that the answer is "no"! In particular the ordinal of the wellordering of *all* ordinals of type `a` cannot itself be of type `a`.

We shouldn't expect the BF paradox to tell us *at what point* ordinals cease to be monomorphic, merely to tell us that there is such a point. This question of quite where ordinals cease to be monomorphic is a good one to think about. BF tells us that if the monomorphic ordinals form an initial segment then it is a *proper* initial segment, so it is natural to ask what operations that proper initial segment is closed under... `successor` for one. The thought that it might be closed under some other operations is a rich source of axioms. (Does every normal function from ordinals → ordinals have a monomorphic fixed point, for example?) It seems to me that this is how we should understand large cardinal axioms: as assertions that the initial segment of monomorphic ordinals is closed under ever more operations.

So: there we have another take on the elephant, one that I hope may be helpful to other blind men engaging with it. This is not in competition or contradiction with the set theoretic analysis; it's merely a shot from another angle. I am not objecting to the ZF-iste take on Burali-Forti; all I object to is the mistake of thinking that it is the last word on the matter. The Burali-Forti paradox is a birthmark on the skin of Mathematics and we pick at it to get at the Mathematics beneath it. Set theory is merely one way of doing it among many, and we handicap ourselves if it is the only weapon we use.

Thomas Forster
Victoria University of Wellington

The Reasoner Speculates

Addendum to the philosophical puzzle of Theseus' ship

Imagine that the castle of the city of EG, the "EG Castle", is being rebuilt by replacing the original stones with new ones. The old stones are transported elsewhere, say to the town of PB, and the castle is rebuilt using the original stones in its original form. In EG, the castle will also be preserved, not by using the original stones but with new ones. Is there still an EG castle at that time, and if so, which of the two?

1. If the main identity criterion of a castle is its location and continued existence, therefore, there is still the Castle of EG, only renewed.

2. If the main identity criterion of the castle is the original stones within it, therefore, there is still an EG castle; it has just moved to a new location.

3. If both the place and the stones are essential to its identity, therefore, the Castle of EG has ceased to exist and no longer exists.

We need to know what we are talking about when referring to the "Castle of EG". Because no such strange occurrences have happened, everyday language provides no answer to the question; therefore, we cannot give a satisfactory answer to the puzzle.

Next, let us look at Theseus' ship and the philosophical puzzle that goes with it. The criterion for the identification of the ship was that Theseus had boarded the ship and was sailing across the sea towards his destination. During the long voyage, many parts of the ship were replaced, but there was no

question whether it was Theseus' ship because the criterion for reidentifying the ship was not done by looking at the parts of the ship but by the fact that Theseus had travelled on it. The continued existence of the ship and the hero's continued travel on it reidentifies the ship in time. I could not speak of a "ship of Theseus" if the hero had changed ships during the voyage.

The criteria for identification and existence are different between a ship on a voyage and a retired ship. They can replace all the parts of the ship during the voyage; this is not an identification problem. The identity criterion for Theseus' ship is that the hero is travelling on it. If he were to change ships on the way, there would be no point talking about his ship. A ship on a road is like a living creature; all the parts built into it become part of it, but the parts thrown out of the ship do not. However, the situation is different with a ship on exhibition, wherein the aim is to remain unchanged. Theseus' ship on the way home is a means of transport. Once it has been dismantled, it becomes a relic of the past, and it ceases to be a means of transport. The ship that was exhibited was the Theseus' ship, but now it is no longer his ship; the hero, having returned home, has a new ship. We have to decide how far and to what extent of deterioration we should consider the exhibited ship to be a descendant of the ship that made the famous voyage. Afterwards, we can say that the remaining ruin resembles the original ship, but we deny that they are the same. To formulate the similarity and uniformity, we assume that we can measure the difference between two ship examples from the previous state and the original state.

Following Amie L. Thomasson's investigations, the alternatives arise as follows:

1. application condition – it is a vehicle or an object memory;
2. identification criterion – the ship was named The Ship of Theseus when he boarded it;
3. re-application criterion – a ship (or an object memory) when it is the same ship (or object memory) as before.

Imagine while walking along a beach, a friend points at a ship: "See, that is Theseus' ship, he sailed it to defeat the Minotaur." The ship's name was marked on a small sign with the inscription, "This is Theseus' ship."
This is the first time I saw the ship as a memory; the identifying criterion for the ship.

Many years later, I passed by again and found the sign and the ship behind the sign. The boat looked like it had been repaired a lot over the years, and to tell you the truth, I did not remember what it looked like when I first saw it. However, seeing the sign made me believe that the ship in front of me was the ship of Theseus. Meanwhile, I learned that it no longer possesses any of its original parts; they had all been replaced. However, someone had rebuilt the ship in its original form from the old parts in another place with only slightly rotted planks. He thinks he has the Theseus' ship. Now who is right: which ship is Theseus' ship? We need to know what we are talking about and what the logical proper name "Theseus' ship" means, otherwise, the question is meaningless.

1. If the ship is a vehicle, its identity criterion is the place where it was originally placed, and its continued existence next to the sign, therefore, Theseus' ship exists and is there next to the sign, only renewed.
2. If the ship is a memory, and its main identity criterion is the old planks in it, therefore, Theseus' ship still exists, not at the sign but in the new location.
3. If both the place at the board, the continued existence, and the parts of the ship are relevant to its identity, therefore, Theseus' ship has ceased to exist.

We can decide that Theseus' ship, as an exhibited memory, exists only as long as, for example, we have most of the original parts, or if we are stricter, we have 70% of the original parts. The change must be measured against the original state because only in this case we obtain an equivalence relation. If we measure the change to the previous state, we get a similarity relation, which is not transitive, only reflexive and symmetric.

Ferenc András
Pomáz, Hungary

Courses and Programmes

Programmes

https://www.unimi.it/it/corsi/laurea-magistrale/human-centered-artificial-intelligence: MA in Human Centered Artificial Intelligence, University of Milan.Italy
MA in Reasoning, Analysis and Modelling: University of Milan, Italy.
APhil: MA/PhD in Analytic Philosophy, University of Barcelona.
Master Programme: MA in Pure and Applied Logic, University of Barcelona.
Doctoral Programme in Philosophy: Language, Mind and Practice, Department of Philosophy, University of Zurich, Switzerland.
Doctoral Programme in Philosophy: Department of Philosophy, University of Milan, Italy.
LogiCS: Joint doctoral program on Logical Methods in Computer Science, TU Wien, TU Graz, and JKU Linz, Austria.
HPSM: MA in the History and Philosophy of Science and Medicine, Durham University.
LoPhiSC: Master in Logic, Philosophy of Science and Epistemology, Pantheon-Sorbonne University (Paris 1) and Paris-Sorbonne University (Paris 4).
Master Programme: in Artificial Intelligence, Radboud University Nijmegen, the Netherlands.
Master Programme: Philosophy and Economics, Institute of Philosophy, University of Bayreuth.
MA in Cognitive Science: School of Politics, International Studies and Philosophy, Queen's University Belfast.
MA in Logic and the Philosophy of Mathematics: Department of Philosophy, University of Bristol.
MA Programmes: in Philosophy of Science, University of Leeds.
MA in Logic and Philosophy of Science: Faculty of Philosophy, Philosophy of Science and Study of Religion, LMU Munich.
MA in Logic and Theory of Science: Department of Logic of the Eotvos Lorand University, Budapest, Hungary.
MA in Metaphysics, Language, and Mind: Department of Philosophy, University of Liverpool.

MA in Mind, Brain and Learning: Westminster Institute of Education, Oxford Brookes University.
MA in Philosophy of Biological and Cognitive Sciences: Department of Philosophy, University of Bristol.
MA programmes: in Philosophy of Language and Linguistics, and Philosophy of Mind and Psychology, University of Birmingham.
MRes in Methods and Practices of Philosophical Research: Northern Institute of Philosophy, University of Aberdeen.
MSc in Applied Statistics: Department of Economics, Mathematics and Statistics, Birkbeck, University of London.
MSc in Applied Statistics and Datamining: School of Mathematics and Statistics, University of St Andrews.
MSc in Artificial Intelligence: Faculty of Engineering, University of Leeds.
MSc in Cognitive & Decision Sciences: Psychology, University College London.
MSc in Cognitive Systems: Language, Learning, and Reasoning, University of Potsdam.
MSc in Cognitive Science: University of Osnabrück, Germany.
MSc in Cognitive Psychology/Neuropsychology: School of Psychology, University of Kent.
MSc in Logic: Institute for Logic, Language and Computation, University of Amsterdam.
MSc in Mind, Language & Embodied Cognition: School of Philosophy, Psychology and Language Sciences, University of Edinburgh.
MSc in Philosophy of Science, Technology and Society: University of Twente, The Netherlands.
MRes in Cognitive Science and Humanities: Language, Communication and Organization: Institute for Logic, Cognition, Language, and Information, University of the Basque Country (Donostia San Sebastián).
Open Mind: International School of Advanced Studies in Cognitive Sciences, University of Bucharest.
Research Master in Philosophy and Economics: Erasmus University Rotterdam, The Netherlands.
Doctoral Programme in Philosophy: Language, Mind and Practice, Department of Philosophy, University of Zurich, Switzerland.
MA in Philosophy: Dept. of Philosophy, California State University Long Beach.

MY HOBBY: GEOHYDROTYPOGRAPHY