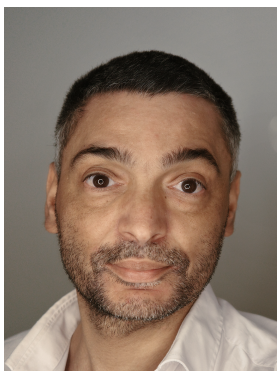
The Reasoner

Volume 18, Number 3
May 2024

thereasoner.org
ISSN 1757-0522

Contents

Editorial



Dear Reasoners, it is my pleasure to welcome you to this new issue of *The Reasoner*, which opens with an interview with Peter Grünwald. Peter is senior researcher in the Machine Learning group at CWI in Amsterdam, which he headed from 2016 to 2023. He is also full professor of statistical learning at the Mathematical Institute of Leiden University. As you can read below, his background is, as it is often the case with our guests, multidisciplinary and in fact quite unique. Peter has been recently been awarded an ERC Advanced Grant with the project *Flexible Statistical Inference* which constitutes the focus of our chat. Before leaving you to it, I would like to thank him very much for his time.

Hykel Hosni
Logic, Uncertainty, Computation and Information Lab,
University of Milan

Features

Interview with Peter Grünwald

Hykel Hosni: You have just been awarded an ERC Advanced Grant – congratulations, Peter! Can you tell us what the project is about?

Peter Grünwald: Sure! Most statistical methods require that all aspects of data collection and inference are determined *in advance*, independently of the data. These include when to stop collecting data, what decisions can be made, e.g. accept/reject a hypothesis, classify a new point, and how to measure their quality, e.g. via a loss function or significance level.

HH: Which we rarely know how to do in practice.

PG: Exactly – the demands of classical statistical methods are wildly at odds with the flexibility required in practice! It makes it impossible, for example, to achieve error control in meta-analyses and it contributes to what is called the "replication crisis" in the applied sciences. My plan is to develop a novel statistical theory in which all data-collection and decision-aspects may be unknown in advance, possibly imposed post-hoc, depending on data itself in unknowable ways.

Yet this new theory will provide small-sample frequentist error control, risk bounds and confidence sets.

HH: The core idea behind this, I take it, is the concept of *e-value*, which you introduced in a 2019 paper co-authored with Rianne de Heide and Wouter Koolen ([1906.07801]). This was presented in its final version at a discussion meeting of the Royal Statistical Society in January 2024, and is now published in their journal (JRSS B) under the name *Safe Testing*.

PG: Indeed, I base myself on far-reaching extensions of e-values and e-processes. These generalize likelihood ratios and replace p-values, capturing "evidence" in a much cleaner fashion. Both played an essential role in the development of anytime-valid inference, the one aspect of flexibility that is by now well-studied.

HH: Can you explain briefly what anytime-valid inference is?

PG: Sure. It means, for example, that one gets valid confidence intervals for one's estimates irrespective of when or why one stops collecting the data: the probability that the true parameter will *ever* fall outside of the 95

HH: This is technical work which may have a strong methodological impact.

PG: I think so. Indeed, another major innovation, that has an epistemological flavor and as such may be of special interest to readers of *The Reasoner*, is the *Sigma-Collection*, an extension of the concept of random variable that takes into account that in practical data collection, we often do not know what would have happened if the data had been different than they actually were. This is very often the case in practice, but strictly speaking, classical statistics - used by most practitioners - cannot deal with such "unknowable counterfactuals" at all. As such, we are currently living in a state of sin!

HH: I can hear "Bayesians" mumble here...

PG: This issue does go away if one strictly adheres to the likelihood principle, as, for example, subjective Bayesians do. But this is an excessive price to pay: it requires a full specification of the likelihood function. There are some very simple problems in which this is nearly impossible, yet non-likelihood-based methods exist which work quite well - I called this "the embarrassment of Bayesian statistics" in P. Grünwald 2023 ("Proposer of the vote of thanks to Waudy-Smith and Ramdas and contribution to the Discussion of 'Estimating means of bounded random variables by betting", *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86(1), 28–30)

HH: You enticed epistemologists with an approach to deal with counterfactual data. Can you say a bit more about that?

PG: By replacing random variables by sigma-collections, I can overcome the problem by formalizing, for any given problem, knowledge about counterfactual situations: *what exactly do I know about what would have happened if the data had been different from what I actually observed*? One may impose priors too, but one does not have to. Moreover, these priors have a different function from Bayesian ones. In fact, a second major innovation in the proposal is the *e-posterior*, which behaves differently from the Bayesian one: if priors are chosen badly, e-posterior based confidence intervals get wide rather than wrong.

HH: This way one acknowledges that the starting point wasn't great by being epistemically more humble, as it were.

PG: In fact, both the dominant existing Wald-Neyman-Pearson and Bayesian statistical theories will arise as special, extreme cases of the new theory, based on perfect (hence unrealistic) knowledge of the data-collection/decision problem or the underlying distribution(s), respectively.

HH: A Nature 2019 editorial called for retiring statistical significance altogether. Do you agree?

PG: I think it's impossible to retire it. People *yearn* for significance! I have heard several stories of statisticians telling journal editors to abandon it, but then the editors say "but how should we decide *then* whether the results are strong enough to be worth publishing?"

HH: On this note I heard significance being likened to democracy as the "worst form of inference except for all the other forms which have been tried from time to time". What's your take on this long-standing issue?

PG: The core of this discussion lies in the fact that modern statistical hypothesis testing is really a curious amalgam of two conceptually very different approaches, Fisher's and Neyman's. We observe some data to test a null hypothesis representing the status quo like "medication does not work", "coin is fair", "there is no effect", etc., and we summarize our findings as a p-value. Fisher saw a small p-value as evidence against the null - and that's it, no specific decision-theoretic consequences. Neyman said we should "reject the null, i.e. conclude that a medication works, for example", if $p < \text{alpha}$, where alpha is set in advance, usually to 0.05, and that's it - the p-value in and of itself then does not give more information than one bit (larger vs. smaller than alpha).

HH: In the Neyman setup we make a binary decision and forget about the degree to which the evidence speaks against the null hypothesis.

PG: With Neyman's procedure, the probability of a false positive, which is"reject the null while it's true", is bounded by alpha - we have a *Type-I error guarantee* of alpha. The idea to mention both "reject/accept" but at the same time mention the p-value (which everybody does in practice) is mired with difficulties. There is no decision-theoretic interpretation that tells you "how you can make better decisions if you observe a p that is much smaller than the alpha set in advance". It is extremely hard to explain this to practitioners - if they observe $p < 0.01$ but had set alpha to 0.05, they tend to retro-actively claim that their alpha was 0.01 but this is plain wrong, it invalidates the error guarantees.

HH: And this is one kind of practical situation in which a flexible form of inference would be extremely useful.

PG: In a way, replacing p-values by e-values in the above can be viewed as taking the sting out of this discussion. E-values are also summaries of data in an experiment, where a *large* e-value means "strong evidence against the null". Based on one's data, one may report an e-value, and just consider this as a measure of evidence against the null hypothesis that "there is no effect", without any reject/accept decision.

HH: There are however problems for which it makes sense to have a binary decision, which normally one arrives at by fixing a threshold for 'rejecting' the null.

PG: As I showed in the paper *Beyond Neyman-Pearson* (on arxiv, under submission), with e-values this is a meaningful operation, giving guarantees on expected losses, *even if the threshold itself is allowed to depend on the data*, i.e. like adjusting alpha based on e itself - which was not allowed with p-values. Hence, it makes sense to mention the e-value as evidence *and* accompany it with an accept or reject decision (that's what I meant when I said "it takes the sting out of the discus-

sion") - one can even take much more general decisions in decision problems with more than two actions.

HH: Can you tell us how the e-value is overcoming the highly problematic p-value?

PG: The e-value overcomes several issues with the p-value. Based on e-values you can make decisions with guarantees based on their expected loss even if the decision problem is changed in light of the data - changing the threshold alpha in light of the data as above is a simple instance of this. But this is a relatively new insight.

HH: What was the original idea?

PG: The original motivation was the following: suppose you observe data from an experiment, say a clinical trial for some new medication. Now, perhaps because the data looks promising but not fully conclusive, you decide to gather additional data, perhaps in a different hospital. We call this *optional continuation*. Even though the data are independent, the decision to do a second experiment may *depend* on the outcomes on the first one - so there is a dependency after all, and it's a *murky* dependency. For example, your boss may tell you to do a second trial, and you do not exactly know for what data he would have told you to do so.

HH: So how would you use e-values here?

PG: You can simply calculate e-values for both trials and multiply them, and you get a new, valid e-value, which can be used for making decisions with Type-I error guarantees. You may also just sweep the data together and re-calculate an e-value on the full data, that's also fine. With p-values, multiplication is a mortal sin for it can only make them smaller, hence increasing the evidence against the null, since they are bounded by 1. And because of the dependency, recalculating the p-value based on the full data is also wrong. There's just no way to deal with optional continuation based on p-values! In contrast, with e-values, you can keep doing it forever - whenever new data is added, you recalculate a new e-value and it remains valid. This, to me, seems a basic requirement for a useful formalization of "evidence". It is also needed to get formal error guarantees in meta-analysis, a highly important tool in the medical sciences! Current meta-analyses do not provide any precise guarantees on error probabilities at all, because p-values and standard confidence intervals cannot deal with "optional continuation".

HH: I would like to challenge you to give the multidisciplinary readers of *The Reasoner* a motivation to study e-values

PG: I'm very happy to take it on! It is to do with the issue about counterfactuals that I alluded to before. There is a famous example (Pratt's Volt-Meter, dating back to 1961) of how the classical theory of p-values and confidence intervals goes wrong if you lack knowledge about "what would have happened in situations that did not occur". But to solve this particular issue with e-values more theory is needed, and that's part of my proposal!

HH: Go ahead, please.

HH: Suppose a statistician, or a reader of *The Reasoner*, is asked to estimate the average voltage going through a batch of electron tubes. This is measured by an engineer with an accurate volt-meter. The engineer provides data of 100 measurements. Later the statistician visits the engineer's laboratory, and notices that the volt-meter reads only as far as 10: the population appears to be censored. According to classical statistical theory, this necessitates a new, complicated, analysis. However, the engineer says she also has a super-high-range-meter, equally accurate, which she would have used if any of the measurements had turned out ≥ 10. This is a relief to the statistician, because it means the original analysis is correct after all. But the next day the engineer telephones and says, *I just discovered my high-range volt-meter was not working the day I did the experiment*. The statistician then informs her that a new analysis, leading to different confidence intervals, will be required after all! The engineer is astounded. She says, "But the experiment turned out just the same as if the high-range meter had been working. *I learned exactly what I would have learned if the high-range meter had been available*. Next you'll be asking about my oscilloscope!"

HH: How about situations in which e-values make a *practical* difference compared to p-values?

PG: There are so many! For example, we already did a first "live" meta-analysis with data coming in from eight different clinical trials and we could keep monitoring it, always allowing new hospitals to enter and hospitals to leave as we were going - and we could stop and publish results, adhering to a Type-I error guarantee, as soon as we had enough evidence. This is completely beyond the realm of classical methods.

Another one: currently, in order to get a medication approved by the FDA, you need to get a significant result (conclusion: "it works") in three independent studies, at significance level alpha=0.05. What do you do if in the first two studies,, you observed very strong evidence: $p1 = 0.005$, $p2 = 0.003$. But $p3 = 0.06$, slightly too large.

HH: Hopefully you would not invoke "clear trends towards significance"!

PG: No, this counts as not being good enough, and there is simply no clear way to combine the three results and employing that the first two gave apparently very strong evidence. With e-values, you can simply multiply and get a new, valid e-value. Note though that each individual e-value will give you usually somewhat less evidence than the corresponding p-value based on the same data. That's the inevitable price to pay to get the desired flexibility!

Another really nice one - where using e-value methodology might have saved at least one live - is the SWEPIS study - let met simply refer to the paper I coauthored with Rosanne Turner and Alexander Ly (Journal of Statistical Planning and Inference 2024) for more on this.

HH: Well, that is as practical as it can get! Do you see e-values also being directly applicable to data-intensive and AI-driven science?

PG : Yes. In fact, Amazon and Netflix are using them as we speak for online A/B testing. Also the thriving literature on *bandits* in the AI world often uses e-value-like constructs, just not under this name. Having said that, we currently have fast implementations only for very simple statistical settings, like testing whether a distribution is normal or not, or basic linear regression. Part of my ERC proposal is to develop efficiently computable e-values for more sophisticated statistical problems such as generalized linear models (e.g. logistic regression). We already have algorithms for calculating e-values for such problems, but they are excruciatingly slow and impractical.

HH: Thanks for sharing so many insights about your exciting new project. Can I ask you a bit about your background? How did you get interested in data-driven inference?

PG: My undergraduate studies, what one would now call a combined Bachelor and Master, was in computer science, not mathematics! However, during the course of my study I found out that my high school math teacher, who had been

disappointed I did not study math, had been right all along: I liked the math courses best, and I really disliked the (several!) courses on large software engineering projects. So I did several extra math courses along the way. Interestingly though, these were mostly about *logic*.

HH: What a surprise! Tell us more.

PG: I was totally fascinated by Gödel's (in)completeness results back then and profited from taking a course by Johan van Benthem, a highly influential logician who teaches amazingly well. I did like probability theory, but I did not like statistics - I never completed any course on it.

HH: Which maybe was instrumental in you taking issue with the orthodox views! What about your PhD?

PG: It was officially about machine learning theory. In it I got more and more interested in statistics and reasoning (yes!) under uncertainty and basically taught myself. My supervisor, Paul Vitányi, is an expert on Kolmogorov complexity, not statistics or uncertain reasoning, but he gave me full trust and freedom to pursue my own interests.

HH: Quite remarkable.

PG: I am still very grateful for that. He was mostly a supervisor in a meta-sense, as a kind of intellectual mentor. Delightfully politically incorrect and original. One might see him as a visionary - he took me as a student on machine learning theory and Harry Buhrman as a postdoc on quantum computing in the early-to-mid 90s. Back then both topics were considered completely fringe! I think there is a lesson to be learned here for administrators and senior scientists - everyone who wants to "steer" science.

HH: So what did you bring to your meetings with him?

PG: I got really interested in things like maximum entropy and axiomatic derivations of such principles, Dutch book theorems, the Savage expected utility axioms and so on. I actually read Paris' and Savage's books almost cover-to-cover! I always found such approaches impressive, but never got fully convinced. To be honest - I hope you'll forgive me the critique - I am rather stunned by interviews I've read earlier in *The Reasoner* about people who still think that if we want to be rational decision-makers, we should all be "Good Bayesians" or adhere to MaxEnt, and so on.

HH: I take your point, but here is a slightly different way of looking at it. If your goal is a beautiful result in the sense of those put forward by the names you named, then you are somewhat forced to make significant abstractions. And I guess there is no real issue with that, so long as one does not fall in the trap of thinking that the practical problems should fit the abstraction. Continuing on your PhD, how was the defense?

PG: By the time my defense came, there was so much statistics in my thesis that I asked famous statistician Philip Dawid on the committee. I had met him at a summer school in Sicily and he seemed interested in my work. But I hesitated, since my mathematical background was still quite minimal - I essentially only worked with countable sample spaces, for example. Philip has also been very influential for my career. He liked my thesis, when I apologized to him "perhaps you find some things a bit primitive, I have to admit I never took a statistics class!" He replied "that's what saved you!".

HH: I told you! How did you move forward in Academia?

PG: I went on to do a postdoc at Stanford, still in computer science, then a second postdoc at EURANDOM in Eindhoven, in statistics, and then back at CWI, where I had also done my PhD. Again a piece of luck: Joe Halpern did a sabbatical at CWI when I came back and we wrote several papers about uncertain reasoning together. I then managed to secure a major Dutch grant which got me a permanent position. I relatively soon obtained a 1-day-a-week full professorship in Leiden as well - that was probably related to me becoming active, together with Richard Gill - now I've mentioned all the people who've strongly influenced me I think! - to reopen the case against a nurse who had been falsely convicted of murdering patients. The case was partly based on flawed statistics - whenever somebody died, she was on duty, and somebody had made a flawed calculation concluding "this couldn't have been a coincidence".

HH: A sad textbook example, I'm afraid. Let's close the circle, then. Can you tell our readers, especially the early-stage researcher, what you think led the path to your successful ERC proposal?

PG: As with life in general, one needs to be creative, convinced of one's own ideas, do hard work, get good advice from others, but certainly one also needs to have considerable luck.

HH: Can you make an example?

PG: Sure! One piece of luck was that in 2010 I was asked to *review* a paper on test martingales by Shafer, Vovk and collaborators that changed my whole way of thinking about testing and the like. I do not know why the editor sent it to me, but that might have been the start of the ERC path! From then on I started thinking, not publishing!, about e-values and flexible inference a lot even back when hardly anybody else did so - I had a kind of breakthrough moment in 2016 when I derived the first version of what is now the main theorem in our *Safe Testing* paper, which basically shows that you can construct nontrivial e-values for an extremely wide variety of statistical settings.

HH: That's a very good argument for not turning down review work! And I find it particularly interesting that you started thinking about e-values without pushing for immediate publications on the topic.

PG: I also had the insight early on that e-value like concepts could help for getting other types of flexibility beyond anytime-valid inference, but I couldn't prove anything formal in this direction until 2022. Then, I managed to prove something about data-driven decision problems and saw "the big picture". The lucky part was that by then, e-values had become widely known in the theoretical statistics community - and that was because in 2019 suddenly, almost out of the blue, many different strong statistics groups started writing about them. If my group had been the only one, this might have been seen as crackpot research; now it clearly wasn't. So in 2022 I decided: this is the time!

HH: And it was.

PG: I also had/have some clear and pretty radical ideas, like changing the notion of random variable which has been used for 80 years as the basis for statistical modeling! Before 2022, the topic would have been too obscure and my ideas too narrow; if I had waited a few more years, I'm sure it would have lost the new-exciting-radical-yet-mathematically-sound flavor.

HH: Did *you* get good advice?

PG: I did not listen to the advice of many colleagues who have told me for years I should apply for an ERC Advanced - preparing proposal and interview takes about 3 months of full-time work, so you should only do that if you're really convinced you have an extremely good idea! At the same time, when I gave my first practice presentation to a "mock committee", they completely trashed me: the presentation was incomprehensible,

focused on the wrong things etc. At *that* point, I knew I had to listen to advice very carefully! In the end I obtained the ERC advanced grant the first time I applied. But note that I have applied for many other types of grants in the past and often failed!

HH: You will be hiring early stage researchers soon. What will be your first advice for them?

PG: This is difficult, because if there is one thing I learnt from 20 years of supervising students, it's that everyone is different! Advice that would work for some would perhaps not work for others. So, perhaps I can give the following meta-advice: don't assume that others are like you!

HH: Can you unpack this a bit?

PG: Some researchers work best if they almost literally hide in an ivory tower, read papers and scribble down formulas; others thrive if they talk in a group before a whiteboard. Some can listen for hours to talks, others can't take up information like that. Other than that – as I already indicated above– , it's crucial to listen to others. I have made some grave mistakes by not doing so, but also, you shouldn't listen too much either [giggles]. This research on e-values would never have happened if I had let others tell me what to do. Until about 2019 I only got blank stares when I talked about it - and believe me, I did spend a *lot* of time thinking about it anyway!

HH: Finally, do you have reading suggestions for anyone who is serious about the methodology of data-driven inference?

PG: Unfortunately, easy-to-read introductions are still lacking. For the time being, I'd recommend reading the first 10 pages of our 2024 JRSSB paper *Safe Testing*. There does exist a good overview of the field up till 2023 but it is highly technical, and requires knowledge of discrete-time random process theory, in particular martingales. This is the paper Game-Theoretic Statistics and Safe, Anytime-Valid Inference by Ramdas, myself and others.

Leibniz's Logicism and Its Aristotelian Ancestry

Bertrand Russell traces the first explicit and intentional implementation of logicism as the doctrine of reducing mathematics to logic to the works of Gottlob Frege. (1919: Introduction to Mathematical Philosophy, London: Allen and Unwin, p. 7) However, recent scholarship on the history of logicism seems to put equal, if not occasionally more, emphasis on the pioneering works of Richard Dedekind and Giuseppe Peano. (See, for example, E. Reck: 2013, 'Frege, Dedekind, and the Origins of Logicism', History and Philosophy of Logic, 34 (3): 242-65) Yet Russell himself in an earlier work suggests that the trail of the idea of logicism stretches back to Gottfried Leibniz and remarks that the general doctrine underpinning the idea 'was strongly advocated by Leibniz'. (1903/1996: The Principles of Mathematics, New York/London: W.W. Norton & Company, p. 5) For his part, Max Black seems to take umbrage with Russell's overestimation of the Leibnizian contribution in this context and offers a somewhat conservative characterization that Leibniz's 'work contained the germ of' the logistic conception. (1933/1958: The Nature of Mathematics, London: Routledge and Kegan Paul, p. 16)

Interestingly, Frege himself in his classic logicist landmark, The Foundations of Arithmetic, approvingly quotes Leibniz that 'algebra derives its advantages from a much higher art, namely, true logic.' (1884/2007: The Foundations of Arithmetic, D. Jacquette (trans.), New York: Pearson Longman, p. 31) John Austin renders the same quotation from Leibniz in his translation of Frege thus: 'the benefits of algebra are due to its borrowings from a far superior science, that of the true logic.' (1884/1978: The Foundations of Arithmetic, J.L. Austin (trans.), Oxford: Basil Blackwell, p. 21^{e}) Indeed, in New Essays on Human Understanding, Leibniz observes that 'geometer's logic – that is, the methods of arguments which Euclid explained and established through his treatment of propositions – can be regarded as an extension or particular application of general logic.' (1985: New Essays on Human Understanding, P. Remnant & J. Bennett (trans. & ed.), Cambridge: Cambridge University Press, p. 370)

In tracing the logicist threads of the Leibnizian corpus, where historically the first explicit Aristotelian connections appear on the horizons of logicism, Russell highlights Leibniz's idea of *Characteristica Universalis* or Universal Mathematics: 'This was an idea which he [Leibniz] cherished throughout his life, and on which he already wrote at the age of 20. He seems to have thought that the symbolic method ... could produce everywhere the same fruitful results as it has produced in the sciences of number and quantity.' (1937/1958: *A Critical Exposition of the Philosophy of Leibniz*, London: George Allen & Unwin, p. 169) Similarly, in one of his unpublished papers dating back to 1880/81, Frege notes that this particular proposal of Leibniz is one of 'a profusion of seeds of ideas ... that is now to all appearances dead and buried [but] will one day enjoy a resurrection' and sees his own work in *Begriffschrift* published in 1879 as 'a fresh approach to' it in anticipation of the implementation of his own logicist agenda. (1979: *Posthumous Writings*, Oxford: Basil Blackwell, pp. 9-10)

Russell then goes on to say that for Leibniz the 'Universal Characteristic seems to have been something very like the syllogism.' (Ibid., p. 170) In fact, Leibniz himself portrays the significance of the Aristotelian syllogism in the following way: 'I hold that the invention of the syllogistic form is one of the finest, and indeed one of the most important, to have been made by the human mind. It is a kind of universal mathematics whose importance is too little known.' (Ibid., p. 478) In a letter dating to 1696 to Gabreil Wagner on the value of logic against Wagner's anti-Scholasticist attack on Aristotelian logic, Leibniz interestingly describes Aristotle in his attempt at syllogistic formalization as being the first one to write mathematically outside of mathematics: 'It is certainly no small matter that Aristotle reduced these forms [paralogisms] to unerring laws, having been the first actually to write mathematically outside of mathematics.' (L. Loemker (ed.): 1969, Leibniz: Philosophical Papers and Letters. Second Edition. Dordrecht-Holland: D.Reidel Publishing, p. 465) And the profuse portrayal is expanded to such an extent that Leibniz makes his fictional representative of John Locke in New Essays on Human Understanding, viz. Philalethes, to backtrack from his dismissal of the syllogism and to admit that: 'I am beginning to form an entirely different idea of logic from my former one. I took it to be a game for schoolboys, but I now see that, in your conception of it, it involves a sort of universal mathematics.' (Ibid., pp. 486-87)

The significance of the relationship between Aristotle's syllogistic formalization and Leibniz's *ars charateristica universalis* can be better appreciated when it is set against the backdrop of the pivotal prerequisite for the logicist program. That is, for the logicism project to get off the ground, the initial necessary step is to set up a formal deductive system of logic ade-

quate for formalizing the reasoning of one domain into another one. Particularly, in the case of Fregean logicism and its recent descendants in the form of neo-logicism, the formal deductive system must possess the ability to formalize mathematical reasoning. This indeed constitutes the principal *prerequisite* or *precondition* of any attempt in the implementation of logicism. In fact, this is where Aristotle's syllogistic formalization looms large in Leibniz's pathbreaking logicist endeavors.

Specifically, in a didactic discussion of the nature and merit of syllogism in his *New Essays on Human Understanding*, Leibniz focuses on two crucial characteristics of syllogism in relation to his own work: (i) touting syllogistic structure as a sharp and *precise* language through which arguments can be represented with clarity and unambiguously: 'It can be said to include an *art of infallibility*', and (ii) introducing the concept of *argument* or *logical form* as the fundamental feature of syllogistic structuring for the purpose of any analysis and argumentation: 'any reasoning in which the conclusion is reached by virtue of the form, with no need for anything to be added.' (*Ibid.*, p. 478) That is, reducing argumentation to 'the bare bones of 'logical form','. (*Ibid.*, p. 480) In fact, insofar as the second feature of syllogistic theory is concerned, Leibniz goes on to emphasize the importance of logical form by citing the example of reasonings in Euclidean geometry: 'Most of Euclid's demonstrations, too, are close to being formal arguments.' (*Ibid.*, p. 479) It, therefore, seems one of the most fecund logical legacies of Aristotle that Leibniz inherits is the notion of formal proof that also plays a significant role in his anticipated project of logicism. Although, readily admits Leibniz, 'scholastic syllogistic form' – the emphasis is on scholastic not Aristotle's – is prone to 'prolixity and confusion' (*Ibid.*, p. 478), being 'ridiculous' (*Ibid.*, p. 481), and are 'usually inconvenient, inadequate and poorly handled' (*Ibid.*, p. 483), in view of the later developments in the foundations of mathematics and logic, Leibniz serendipitously saw in Aristotle's logical work that there could be no rigor in the absence of formality. Anecdotally Leibniz remarks: 'I have had personal experience of controversies – even ones in writing, with people of good faith – where mutual understanding began only after we had resorted to *formal arguments* to sort out our tangle of reasonings.' (*Ibid.*, p. 481; emphasis added)

There is, however, a second aspect of the Aristotelian ancestry of Leibniz's logicism that is often neglected: namely, the status and significance of the law of non-contradiction. Aristotle in his groundbreaking role as the first metalogician (Jonathan Lear: 1980, Aristotle and Logical Theory, Cambridge: Cambridge University Press) attempts to shed light on the nature of proof and consequence as well as the status of the law of non-contradiction in his Metaphysics with the ultimate aim of demonstrating the intelligibility of the broad structure of reality in the same breath. Similarly, Gottfried Martin notes, 'Aristotelian logic is ... a complicated mixture of logic, metalogic and metaphysics, and Aristotelian metaphysics contains logical and metalogical considerations.' (1964: Leibniz: Logic and Metaphysics, Manchester: Manchester University Press, p. 85) In Aristotle's own articulation, this metaphysical and metalogical interplay and interaction takes place in the following manner: 'Obviously then it is the work of one science to examine being qua being, and the attributes which belong to it qua being, and the same science will examine not only substances but also their attributes.' (Richard McKeon, ed.: 1941, The Basic Works of Aristotle, New York: Random House, 1005^{a} 13-16, p. 735)

Consequently, the question is which discipline or branch of knowledge has the necessary wherewithal and the *logical* capability to deliver the objectives and goals of the universal or special science of being. Aristotle's answer is unhesitatingly categorical with a tantalizing twist: "Evidently then it belongs to the philosopher, i.e. to him who is studying the nature of all substance, to inquire also into *the principles of syllogism*." (*Ibid.*, 1005^{b} 6-8, p. 736; emphasis added) The weight of the twist, *viz.* the reference to the theory of syllogism, can be best appreciated against the backdrop of the Leibnizian *ars charateristica universalis* discussed earlier in terms of the availability of a formal deductive system of logic adequate for formalizing the reasoning of one domain into another one as a prerequisite to realize the logicist project.

Now, one may pose the question: what is after all the outcome of the study of being as being by inquiring into 'the principles of syllogism'? The result is a principle, remarks Aristotle, that 'is the most certain of all': 'It is, that the same attribute cannot at the same time belong and not belong to the same subject and in the same respect': that is, the law of non-contradiction. (Ibid., 1005^{b} 17-20, p. 736; emphasis added) Also, to leave no room for doubt as to the core fundamentality and centrality of this principle vis-à-vis any other principles including mathematical ones, Aristotle sharpens his stance by the following observation: 'This, then, is the most certain of all principles ... that all who are carrying out a demonstration reduce it to this as an ultimate belief; for this is naturally the starting-point even for all the other axioms.' (Ibid., 1005^{b} 22 and 31-34, pp. 736-7; emphasis added)

Given this Aristotelian angle on the law of non-contradiction, it is worth noting Leibniz's take on the law of non-contradiction here. He writes: 'The great foundation of mathematics is the principle of contradiction ... This single principle is sufficient to demonstrate every part of arithmetic and geometry, that is, all mathematical principles. (Loemker: p. 677) Clearly this statement of Leibniz not only displays an exact echo of Aristotle's approach to the law of non-contradiction vis-à-vis arithmetic and geometry but also highlights the logicist implication of it in an important and immediate manner. The only divergence between Aristotle and Leibniz is when moving from mathematics to natural philosophy, Leibniz claims that, 'another principle is requisite ... the principle of a sufficient reason'. (Loemker: p. 677) Otherwise, in terms of the classical conception of logicism, Leibniz finds the law of non-contradiction sufficient to carry out the enterprise.

Majid Amini
Virginia State University

The Reasoner Speculates

Diagnostic reasoning in practice is frequentist

In theory, Bayesian reasoning has been prescribed as the normatively correct approach in diagnosis (Weinstein MC et al.1980: Clinical Decision Analysis. Philadelphia: WB Saunders Company), but it does not appear to be employed in diagnosis in practice. We do not find, for example, a disease to be diagnosed from its posterior probability in a Bayesian manner in any of the scores of published diagnostic exercises in real patients, such as in clinical-pathologic conferences (CPCs) and in

clinical problem-solving exercises. (Jain BP. 2016, An investigation into method of diagnosis in clinical-pathologic conferences (CPCs). Diagnosis 3: 61-64; Jain BP. 2016, Why is diagnosis not probabilistic in clinical-pathologic conferences (CPCs). Point. Diagnosis 32: 95-97.) I shall argue in this paper it is frequentist reasoning (Mayo DG. 2018, Statistical Inference as Severe Testing: How to get beyond the Statistics Wars. Cambridge: Cambridge University Press). which is employed in diagnosis in practice as it achieves accurate diagnosis of a disease with a high degree of reliability in the environment of diagnosis in practice in which practically every disease is known to occur with varying presentations and thus with varying prior probabilities.

With this reasoning, a disease that is suspected from a presentation in a patient with symptoms is formulated as a hypothesis only regardless of its prior probability. This hypothesis is tested by performing a test, and if a highly informative test result with likelihood ratio (LR) greater than 10 is observed (Guyatt G et al. 2008. Users' guide to the medical literature: A manual for evidence-based clinical practice, New York: The McGraw-Hill Companies, p 428), it is interpreted as strong evidence, based on its performance in diagnosing the disease accurately with a high frequency in other patients with varying prior probabilities. From this strong evidence, the hypothesis is inferred to be correct, and the disease diagnosed with a high degree of confidence in the patient. Frequentist reasoning differs from Bayesian reasoning in not interpreting a prior probability as a prior degree of belief and in not diagnosing a disease from its posterior probability that is generated by combining its prior probability and LR of a test result.

I shall illustrate frequentist reasoning in diagnosis with its use for diagnosis of the disease, acute myocardial infarction (MI) in practice. I first look at a real patient discussed in a problem-solving exercise (Pauker SG et al. 1992.How sure is sure enough? N Engl J Med 326: 688-91) in whom frequentist reasoning is employed to diagnose acute MI. This patient is a healthy, 40 year old woman with no cardiac risk factor who presents with highly uncharacteristic chest pain, in whom acute MI is suspected and formulated as a hypothesis. This hypothesis is tested by performing the test, an EKG, which reveals the highly informative test result, acute ST elevation EKG changes, with LR of 13 (Rude RE et al. 1983.Electrocardiographic and clinical criteria for recognition of acute myocardial infarction based on analysis of 3,697 patients. Am J Card 52: 936-42). This test result is interpreted as strong evidence, based, I suggest, on its performance in diagnosing acute MI accurately with the high frequency of about 86 percent or in nearly 9 out of 10 other patients with varying prior probabilities (Larson DM et al. 2007. "False positive" cardiac catheterization laboratory activation among patients with suspected ST-segment elevation myocardial infarction. JAMA 298: 2754-60). Based on this strong evidence acute MI is diagnosed accurately with a high degree of confidence in this patient.

We note, acute MI is diagnosed in practice from the test result, acute ST elevation EKG changes in a patient in whom it is suspected with a high degree of confidence, regardless of its prior probability, all over the world including in USA and Europe (Myocardial Infarction redefined-a consensus document of the Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction. 2000. Eur Heart J 21: 1502-13.), India (Guha S et al.2017, Cardiological Society of India: Position statement for the management of ST elevation myocardial infarction in India. Indian Heart J April 69 (Suppl 1) S63-S97) and in Africa (Shavadia J et al. 2012. A prospective review of acute coronary syndromes in an urban hospital in sub-Saharan Africa. Cardiovasc J of Africa 6: 318-21). This uniformity in diagnosis of acute MI is achieved primarily, I suggest, due to a series of patients with varying prior probabilities in whom acute MI is suspected, over a period of time at some place, being a random sample as I discuss below.

We do not know in advance about prior probability of acute MI in the next patient in whom we suspect it, and this prior probability is independent of its prior probability in any other patient in this series. Therefore, the prior probability of acute MI in a patient in this series, can be looked upon, I suggest, as being a random variable (Blitzstein JK et al. 2019. Introduction to Probability, London: Chapman and Hall, p 103). and this series as being a random sample. Different series of patients with varying prior probabilities in whom acute MI is suspected in different parts of the world, can all be looked upon, I suggest, as being random samples which are drawn from a population of patients with varying prior probabilities in whom acute MI is suspected.

In one such random sample (Larson DM et al. 2007, 2754-60), in which an EKG is performed to test suspected acute MI in patients, the frequency of acute MI in presence of the test result, acute ST elevation EKG changes, is observed to be 86+/-2 percent with confidence level 95 percent. This means this frequency will be observed to be between 84 and 88 percent in 95 percent random samples drawn anywhere from the parent population.

The observed frequency of about 86 percent can be looked upon, as Cox DR (2006. Principles of Statistical Inference. Cambridge: Cambridge University Press) has proposed, as calibrating accuracy of the test result, acute ST elevation EKG changes, in diagnosing acute MI by repeated testing, just as accuracy of a measuring instrument is calibrated by taking repeated measurements. It is due to this calibrated high accuracy, I suggest, that this test result is interpreted as strong evidence from which acute MI is diagnosed with a high degree of confidence in any patient in whom it is suspected, regardless of its prior probability everywhere.

We find any other disease which has a test capable of generating a highly informative result with LR greater than 10 (Guyatt G et al. 2008 p 428) is diagnosed in a similar manner by the frequentist method in practice. For example, pulmonary embolism is diagnosed from positive chest CT angiogram, LR 20 (Stein PD et al. 2006, Multi-detector computed tomography for pulmonary embolism. N Engl J Med 353: 2317-27); deep vein thrombosis from positive venous ultrasound study, LR 16 (Zierler BK 2004. Ultrasonography and diagnosis of venous thromboembolism. Circulation 109: 1-9-1-4.) and covid-19 disease from positive covid-19 antigen test, LR 14 (Watson J et al. 2020. Interpreting a covid-19 test result. BMJ 369 doi.org/1136/bmj.m1808, published 12 May 2020) with a high degree of confidence in any patient in whom it is suspected, regardless of its prior probability all over the world.

We note, prior probability of a disease does not play any direct role in diagnosis by frequentist reasoning. In this reasoning, it is interpreted, I suggest, as chance of a disease in a patient and its only role in diagnosis is in prioritizing testing of various suspected diseases in a differential diagnosis in a non-urgent diagnostic situation. In this situation, the disease with

the highest prior probability is tested first, as it has the greatest chance of being present in a patient.

Frequentist reasoning for diagnosis in practice is highly accurate as overall diagnostic accuracy in practice has been reported to be high at 85 to 90 percent (Berner ES et al. 2008. Overconfidence as a cause of diagnostic errors in medicine. Am J Med 121: S2-S23).

Bimal P Jain MD
Mass General Brigham/Salem Hospital

Dissemination Corner

BRIO

The Third BRIO Meeting

For the third time, the annual research meeting of the national project BRIO took place, this time in the headquarter of Alkemy – the industrial partner of the project – in Milan, on March 8, 2024. Once again, the event has represented an occasion for each research unit to share their advancements, and to envisage future collaborations. Before going through the new ideas that emerged from the dense programme of this single-day event, let us spend a few words on the project itself.

BRIO (Bias, Risk and Opacity) is a research project funded by the Italian Ministry of University and Research https://sites.unimi.it/brio/ and has four main objectives: 1) To formulate an epistemological and normative analysis of Trustworthy AI as undermined by bias and risk, not only with respect to their reliability, but also to their social acceptance; 2) To define a comprehensive formal ontology, including a taxonomy of biases and risks and their mutual relations for autonomous decision systems; 3) To design (sub)-symbolic formal models to reason about safe TAI, and produce associated verification tool; 4) To develop a novel computational framework for TAI systems explanation capabilities, aimed at mitigating the opacity of Machine Learning (ML) models.

The opening talk was given by Viola Schiaffonati (PoliMi), who presented a preliminary work on the risks related to AI. With the AI Act, the notion of risk plays a pivotal role in the current European approach to AI regulation. In the talk, it was argued that the standard distinction between hazard, exposure, and vulnerability is still relevant when it comes to reasoning about the possible harms of AI, both at a philosophical and at a normative level. Nonetheless, it was pointed out that this multi-component analysis is challenged by the difficulty of performing ex-ante and ex-post risk assessments of AI systems in practice.

After that, the workshop continued with a joint talk by Giacomo Zanotti (PoliMi) and Salvatore Giuliano (UniNa) on their current research on the epistemological aspect of "ablation studies" in machine learning. By analogy with the practice of ablative brain surgery, machine learning ablation refers to the practice of removing a component of the AI system, in order to observe the effects in its behaviour. The removal can either regard hyperparameters of the model, such as neurons or entire layers of a neural network (model ablation), or the input features (hence, feature ablation, with its proximity to feature importance techniques), where a feature is replaced with random or constant values. The talk highlighted both technical and epistemological challenges related to such practices, stressing their close connection to the notion of intervention in philosophy of science.

The presentation by Roberto Prevete (UniNa) shed light on the critical issue of data leakage in machine learning and transfer learning contexts. Data leakage occurs when unintended information contaminates the data used to train a model, in a way that makes the evaluation of its performance unreliable. This happens, for instance, when the train and the test sets end up overlapping. The incorrectness of performance estimates is, of course, a significant concern when the system is deployed in real, high-stakes situations, performing worse than expected. The talk presented not only an analysis, but also an exhaustive classification of types of data leakage along the machine learning pipeline.

Two theoretical works on the notion of trust and AI trustworthiness were presented. The talk by Daniele Porello (UniGe) aimed at formalising the concept of trust in Unified Foundational Ontology (UFO) in a way that models the relationship between a trustor and a trustee (either human or artificial) in different contexts of interest. Francesco Genco (UniMi), departing from some considerations in the philosophy of language, walked us through an analysis of trust as an hyperintensional operator. The talk by Emanuele Bottazzi (CNR Trento) focused on another aspect of the AI-human interaction, namely the complex system of linguistic expectations that humans inevitably naturally have when interacting with conversational AI systems, such as ChatGPT. Finally, Greta Coraglia (UniMi), Giuseppe Primiero (UniMi) and Davide Posillipo (Alkemy) presented two new features of the bias detection tool developed in collaboration with Alkemy within the BRIO project. On the one hand, they presented the design of a novel fairness measure called "Correction Distance" based on a previous work by Manganini and Primiero (2023: "Reasoning with Bias", *Proceedings of the 1st Workshop on Fairness and Bias in AI co-located with 26th European Conference on Artificial Intelligence*, pp. 1-16). This measure captures possible disparities in the uncertainty associated to the predictions of an AI system. On the other hand, a new module was realized to encapsulate all the fairness violations detected by the tool into a unique, indicative measure of risk, that will help the final user to have a synthetic view of the possible harms of an AI system.

An informal round table closed the event. Many points emerged as potentially interesting to explore and problematise further: among them, whether and how the operationalisation of ethical notions pursued by the current research on trustworthy AI is adequate and desirable, given that AI systems are not merely mathematical objects, but socio-technical artefacts situated in complex equilibria between stakeholders.

Chiara Manganini
University of Milan and University of Edinburgh

SMARTEST

ONTOLOGICAL ANALYSIS AND MODELLING OF DTS

The project SMARTEST has been introduced in Volume 18, Issue 2 of *The Reasoner* and is aimed at studying, analysing and simulating probabilistic DTs (DTs). DTs are digital representations of physical systems executed in real time and they are mainly used to make predictions about the behaviour of complex systems that cannot be directly tested. DTs may simulate very different kinds of entities, ranging from industrial machineries to the human body, to very articulated systems, like smart cities. Interestingly, even the European Commission is applying the DT paradigm as a technique that may support the Green Deal policy in addressing the increasingly pressing climate change issues.

As for any model or replica, DTs keep some of the properties of the represented entity and miss some other properties. One of the main points of DTs simulation is exactly that of preserving the representation of its essential properties. For this reason, epistemological and ontological analyses become of primary importance and are preliminary to the formal representation and check of such essential properties.

One of the main challenges put forward by DTs is that of continuously monitoring the simulated entity's behaviour, in order to prevent malfunctioning and increase the entity's performances to a maximum or optimal level. To this aim, sensors and networking devices enable a bi-directional stream of data between the entity under analysis and its DT.

Formal ontologies are nowadays used in engineering application contexts to provide a transparent representation of the data stream between a DT and the corresponding physical artefact or entity in general. The approach has been applied in the past to the study and representation of artefacts and relations among them on the one side, and of copy, replica and counterpart on the other, especially in the industrial domain. Leveraging these previous researches, the ontological approach may be used to examine the extent of preservation of essential properties like safety or liveness at the two edges of the copy relation between real entity and DT; following Primiero and Angius' proposal, we may build taxonomies of copies, like exact, inexact, approximate copy, or counterfeit, based on missing or extra properties that the DTs display with respect the the real entity they represent. The construction of an ontology of simulative models and their empirical adequacy with respect to the simulated systems is the *first objective* of the project.

However, the current scenario is evolving towards an ecosystem of connected DTs and this engenders further problems. For instance, so far only very specific domain applications have been fully characterised by formal ontologies and this of course constitutes a bottleneck for the integration of multiple models and data. The Laboratory for Applied Ontology (LOA), a branch of the Institute of Cognitive Sciences and Technologies of the CNR situated in Trento, has extensive expertise in formal ontology, mathematical logic, and epistemology, with more than 15 years experience in formal and computational modelling in the engineering domain. LOA is member of the CLAIRE and TAILOR European consortia for research in AI and participates in the consortium of the project OntoCommons, (ontology-driven data documentation for industry commons) and to the Industry and Standards Technical Committee of the IAOA (International Association of Ontology and its Applications).

One of the main contributions of the lab to the Formal Ontology Community has been the development of the top-level ontology DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering). Top-level ontologies represent the ontological commitment (the entities that exist in the domain and the meaning to be attributed to terms) in an explicit way and can be used for meaning negotiation, as they account for the formal structure of the domain to be represented. They are general theories applicable across domains, composed of fundamental categories and relations, such as object, event, part, space, region, time instant, quality, parthood, constitution, location, etc.). Given their generality, top-level ontologies may be used as tools of analysis that limit *ad hoc* models, and of integration, by establishing mappings between different models. Therefore, DOLCE and other top-level ontologies may come out as very useful as building blocks of an ecosystem of DTs; since their introduction, back in the 90's, they were in fact conceived of especially for interoperability, i.e. the ability of diverse systems and organizations to work together, taking into account not only technical, but also social, political, and organizational factors that impact system to system performance.

Some of the questions the LOA is going to address within the SMARTEST project are: What kind of entities are DTs? Which relations do they entertain with the twin physical systems? In particular, can existing notions such as replica, copy and counterpart, as discussed in the philosophy of technology literature, be adapted to make sense of the engineering view on DTs, or is a new conceptual framework required?

To answer such questions, from a methodological standpoint, the LOA relies on an interdisciplinary approach at the intersection between philosophy, engineering, and knowledge representation. The adoption of existing formal ontologies like DOLCE and analytic approaches like OntoClean, which are both product of the research team, is planned. Moreover, the latter have been already used for the ontological treatment of engineering notions, among which that of technical artefact, functionality, and capability, and they will help in framing the analysis within a larger ontological picture of engineering systems. More specifically, for the characterisation of DTs, the notion of information entity, often used for the modelling of engineering specifications, will be deeply analysed.

The expected outcome of the research is an ontology of DTs contextualised within a broader view on technical artefacts and satisfying both engineering requirements and theoretical views on models and copies found in philosophy and knowledge representation. The ontology will also include the identification of different types of DTs (e.g., prototype DT, development DT, etc.) to properly classify the differences found across engineering models. In addition, the ontology will represent engineering functions and malfunctions to cover the prototypical use of DTs, i.e., predictive maintenance, which requires identifying the types of fault that the different DTs are meant to predict in simulations. As a consequence, the resulting framework will be suitable for an ontological study of reliability which in turn will ground the epistemological and formal study of property preservation that will be developed for the accomplishment of the *second* and *third objective* of SMARTEST, which will

be presented in the next issues of The Reasoner. In fact, the most original contributions of the project are expected from the cross-fertilisation of the involved disciplinary domains, namely formal ontology, epistemology and model checking.

Roberta Ferrario
Institute of Cognitive Sciences and Technologies – CNR

News

"Amplifying underrepresented voices in formal philosophy", 26 April, London

The workshop "Amplifying underrepresented voices in formal philosophy" took place on April 26, 2024, at King's College London. It was a satellite event of the LogIn Project, a podcast aiming to foster inclusivity in formal philosophy by interviewing philosophers who are either members of traditionally underrepresented groups or who work outside of what is perceived as "traditional" topics in formal philosophy, discussing both their research and themes related to diversity in academia. The workshop was organised by Beatrice Buonaguidi (KCL), Giulia Schirripa (St Andrews and Stirling), Elena Wüllhorst (KCL), and Matteo Zicchetti (University of Warsaw).

In line with the LogIn Project's aim, the workshop had the goal of creating a space to discuss the intersection of formal philosophy and feminist philosophy and the intersection of formal philosophy and feminism.

The workshop was opened by a talk by Gillian Russell (soon to be at ANU), titled "Social spheres and generics", which suggested the use of variable binary quantifiers as a tool to give the truth conditions for generics. In particular, by observing that generic statements are often used for talking about social hierarchies, she suggested that variable binary quantifiers could be used as a tool to create awareness of the truth conditions for generic statements which presuppose an underlying social hierarchy or social stereotypes, such as "Women lie", or "Immigrants are treated well in Australia".

Next, a talk by Viviane Fairbank (St Andrews), "Toward a feminist pragmatist theory of logic", aimed to draw an important yet still overlooked distinction between feminist philosophy of logic and feminist logic. Feminist philosophy of logic was defined as philosophy of logic having a distinctive relationship to feminist philosophy, whereas feminist logic was understood as a theory of logical consequence/validity that is grounded in feminist philosophy of logic. Viviane provided a historical overview of several views of feminist philosophy of logic and then suggested a conception of feminist philosophy of logic based on feminist pragmatism.

Frederique Janssen-Lauret (Manchester) gave a talk titled "Ruth Barcan Marcus' Contributions to Modal Logic". In her talk, she presented Ruth Barcan's contributions in establishing the direct reference theory of names, the necessity of identity, and quantified modal logic, all of which are usually attributed mostly to Kripke or to Carnap. She situated Barcan's contributions in the context of her empiricist nominalism, and highlighted how Barcan's presentation of quantified modal logic and her account of the necessity of identity are in some aspects superior to the versions popularised by Kripke.

Ivan Restovic's (Zagreb) talk suggested a model of gender identity based on fuzzy logic, which can overcome some of the difficulties presented by spectrum models of gender identity, especially regarding the treatment of agender identities. Further, he argued that fuzzy contrariety is a promising way to model the difference between genders. Sara Uckelman (Durham) gave a talk titled "Logic, Neurodiversity and Gender". She partly focused on her own experience as a neurodivergent woman and her love of logic to draw attention to how to make academia a more inclusive environment, and to the intersections between gender, late diagnosis of neurodivergence, and feeling welcome in logic. Finally, the workshop was closed by Helen Meskhidze (Harvard), who presented some joint work with Francisco Calderón (Michigan) and Thomas Colclough (UC Irvine), titled "Feminist and trauma-informed approaches to teaching logic". In their work, Helen and her co-authors conducted an experimental study on the perception of logic by philosophy undergraduates, and investigated how to change this perception through pedagogy, for example, showing different approaches to natural deduction proofs to highlight pluralism in logic, and to convey that there may be different ways to solve a problem. They implemented feminist and trauma-informed pedagogy for two basic logic modules at UC Irvine and Michigan, and noticed that this pedagogy contributed in significantly improving the perception of logic by the students. In particular, students from underrepresented groups benefited from the study by perceiving logic as more friendly to them, and by perceiving their own abilities, the objectivity of the discipline and its applicability more positively. The workshop was closed by a panel discussion on how to make logic and formal philosophy a more inclusive environment. Several issues were highlighted over the course of the discussion, among which the usefulness of feminist pedagogy as presented by Meskhidze et al., the importance of granting access to better funding for underrepresented groups, inclusive study groups and summer schools for undergraduates and A-level students. Also, the importance of initiatives such as the LogIn workshop was emphasised, and it was suggested that the event be made a regular one.

Maria Beatrice Buonaguidi
King's College London
Giulia Schirripa
St Andrews and Stirling
Elena Wüllhorst
King's College London
Matteo Zicchetti
University of Warsaw

What's Hot in ...

Statistical relational AI

StarAI and Explainability

In March, I attended a Dagstuhl seminar on trustworthiness in artificial intelligence, where interpretability of AI systems and explainability of their decisions were a major topic of discussion. Attending this event made me think more about the role of statistical relational approaches for this burgeoning field of *explainable AI*.

At its core, statistical relational AI combines the power of first-order logic with the flexibility of statistical artificial intelligence. So to understand explainability and interpretability in statistical relatioanal AI, let us briefly consider classical symbolic ("logical") approaches and classical statistical approaches in this light. Rule-based symbolic AI, sometimes

known as "Good Old-Fashioned AI", is often considered to be the paragon of interpretability. The model itself is given by a set of rules, which are open to human inspection, and any decision reached by the model gives rise to a computation trace or proof tree which takes account of the rules that were involved in reaching a decision. Statistical approaches to AI are somewhat more varied, so let us consider two possible concretisations here. The first is a probabilistic tree, in which the inner nodes encode decision rules and the leaves are labelled with a probability of a given Boolean target predicate random variable to be true. In this model, at least superficially, explainability is secured: The decision tree itself can be read by humans, and any marginal probability can still be explained by a path along the tree, just as a binary decision could be explained by a proof tree in the deterministic case. In the case of multiple targets, a Bayesian network could be employed, with much the same explainability properties as this simple tree model. Alternatively, consider a maximum entropy model, such as a Markov random field. In such a model, correlations between the random variables can be inspected, as can the weighted factors determining the probability distribution at large. However, the factors do not have an intuitive meaning in terms of concrete probabilities, and a marginal probability calculated from such a model can not be explained in terms of a subset of rules contained in it.

These considerations also apply to the formalisms of statistical relational AI. If we consider a probabilistic logic program, which can be written as a finite set of probabilistic facts and rules, we see that the model is highly interpretable: The rule sets are easy to read for a trained human, and the probability annotations have an interpretation as independent probabilities of causes of the atom in the head of the clause. If we have a relational Bayesian network, we can again deduce a lot from the graphical structure of the model, and we also have direct access to the conditional probabilities encoded in the Bayesian network. On the other hand, Markov logic networks are plagued by much of the same issues as ordinary Markov random fields, as the weights do not have any intuitive meaning associated with them.

So far, these should all be fairly straightforward observations. However, the inherent interpretability of propositional rule systems and ordinary statistical models has a very practical caveat: The larger the underlying knowledge bases, decision trees or graphical models become, the harder they are for humans to read. Consider for instance a network of 1000 published researchers, with co-authorships and citation relations between them. An ordinary probabilistic graphical model that captures this situation would have a node for any possible co-authorship or citation, leading to up to 2,000,000 nodes. A probabilistic graphical model with that many nodes is now almost impossible to visualise, let alone does it allow for visual inspection of its basic dependencies. In this situation, a statistical relational AI system can be employed, which encodes the symmetries inherent in the relationships by abstracting away from the individual authors. Suddenly, instead of considering 2,000,000 nodes, there are only two predicates to deal with, co-authorship and citation, and their mutual dependencies are encoded abstractly on a relational level rather than separately for every pair of authors. The resulting model will be extremely compact and easy to inspect, both visually and with regards to its underlying logic. In this way, the relational abstraction turns the theoretical possibility of an interpretable model into practical reality.

However, there are also hurdles on the path to truly interpretable statistical relational AI. In my opinion one of the biggest comes in the guise of structure learning algorithms. In my last column in the January issue, I mentioned BoostSRL, a statistical relational structure learner based on functional gradient boosting. In my experience, functional gradient boosting is the closest StarAI has to a high-performing general learning algorithm for simple prediction tasks. It achieves decent accuracy with excellent runtimes even on large datasets and across a variety of domains. However, just like its non-relational counterpart, BoostSRL ultimately learns a very large family of relational decision trees, which are usually highly redundant and very hard to understand for any user that would like to get an idea about how decisions came about. So while the relational setting drastically reduces the size of the individual tree models, this is counteracted by the large number of trees that are actually constructed during learning. Hence, I was excited to see recent work published in the Machine Learning Journal (Explainable models via compression of tree ensembles, https://doi.org/10.1007/s10994-023-06463-1) that approaches the issue of tree compression, reducing the large ensembles of learned trees to a smaller set that hardly loses prediction accuracy but vastly improves interpretability. So while statistical relational learning and reasoning can go a long way to providing interpretable artificial intelligence in a variety of domains, the interplay of algorithms and interpretability is still a fertile field for researchers interested in logic, probability, computation and the human touch.

Felix Weitkämper
Computer Science, LMU Munich

Courses and Programmes

Courses

LAIS: Logic for the AI Spring 2, 2–6 September, Como, Italy.

Programmes

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: 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 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 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.
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).
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.