The Reasoner

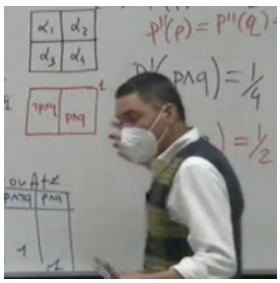
Volume 16, Number 6
September 2022

thereasoner.org
ISSN 1757-0522

Contents

Editorial

Welcome to the new issue of The Reasoner. It is my pleasure to introduce Luca Incurvati, currently Associate professor in the Department of Philosophy and the Institute for Logic, Language and Computation at the University of Amsterdam. He specialises in logic, the philosophy of mathematics and the philosophy of language. Luca is also Principal Investigator of the ERC *Starting Grant* project *EXPRESS: From the Expression of Disagreement to New Foundations for Expressivist Semantics*.



As Luca explains below, a core question he and his team in EXPRESS are working on is the very nature of disagreement, decidedly a hot topic, and not only in academia! His paper "Inferential Expressivism and the Negation Problem", published with Julian J. Schlöder was the recipient of the 2019 edition of the Sanders Prize in Metaethics. You will be able to follow the updates on EXPRESS in our Dissemination Corner, and just in case you missed it, the highlights of the project are described in The Reasoner 16(4), June 2022.

In addition to touching upon his current research and future plans, Luca was kind enough to speak a bit about his personal trajectory *and* to pin down a piece of very good advice for early-stage researchers in the field. Before leaving you to it, let me thank Luca for his time and for the very enjoyable chat.

Hykel Hosni
University of Milan

Features

Interview with Luca Incurvati

Hykel Hosni: Can you tell us a bit about your background? How did your interest for philosophy originate?

Luca Incurvati: My passion for philosophy started already in high school. I attended the so-called 'Liceo classico' in Italy, where in the last three years one is taught some philosophy (mostly from a historical point of view). I had an excellent teacher, which I guess helped. In general, I liked abstract subjects, so I was considering maths, physics (focusing on the theoretical side) and philosophy, which is what I went for in the end.

HH: That's not unheard of in this community.

LI: Indeed. Later on, I also took some maths courses, and by doing philosophy of maths and logic during my academic ca-

reer, I have been able to combine my interests; I haven't worked on philosophy of physics yet.

HH: Plenty of time ahead to do that, I guess! Where did you study?

LI: I studied philosophy at the University of Rome "La Sapienza". Most of the courses were on continental philosophy. Quite early on, though, I was captivated by analytic philosophy. Mostly, I liked the fact that I had the feeling that I could engage directly with philosophical problems.

HH: I never understood why the Italian philosophical culture requires students to go through the whole university library before granting them licence to engage with a problem. But sorry for interrupting.

LI: Indeed! It always makes me think of the passage in the Notebook of Malte Laurids Brigge about all the experiences one must have before being able to write one good verse. But now I am digressing! After taking my bachelor's degree, I decided to spend one year abroad as an Erasmus student. There were several options, including Palma de Maiorca. Instead, I opted for Glasgow, which is when I really knew that I wanted to become an academic. To go to the UK, though, I had to win one of the scholarships for English literature students, since all the scholarships in philosophy were for continental Europe. I am still very grateful to the English literature professor who, without knowing me and on the basis of my grades and my interests, decided that I should be going to Glasgow instead of one of her students.

HH: You must have been very, very keen to pack your wellies and head to Scotland then.

LI: I was! During my year in Glasgow, I got especially interested in philosophy of mathematics, partly thanks to Adam Rieger and the late Bob Hale. Gary Kemp also gave a wonderful course on philosophy of language. Gary had suggested that I stay on for a master's and I had decided to follow his advice, but I broke the ligament of my right knee shortly before starting the master's. So I decided to finish my master's in Italy instead and apply to do a research master's in the UK after that.

HH: Chance at work! Where did you go then?

LI: I got into Cambridge, where I was lucky enough to be part of a community of graduate students all working in the area, including Tim Button, Steven Methven, Florian Steinberger and Rob Trueman. After my master's, I stayed on in Cambridge to do a PhD on the philosophy of arithmetic. At the end of three years of the PhD, I had written a thesis on the philosophy of set theory. I didn't really make a conscious decision to change topic (albeit remaining within the philosophy of maths). But when halfway through I looked at the material that I had written or published, it was clear what my thesis was going to be on. At least in this case, just following the philosophy where it led me worked.

HH: And then?

LI: During the last year of my PhD, I started applying for Junior Research Fellowships (JRFs). These are essentially postdocs, although one is not officially attached to a project or department but to an Oxbridge college. I had a successful interview for a JRF at Magdalene, Cambridge. The panel consisted of one external expert in philosophy and college fellows in a range of non-philosophy subjects including English literature and history. I had seven minutes to explain my research programme on the philosophy of set theory to them – that was quite a challenge.

HH: What did you say to convince them?

LI: If memory serves, I began from sets of apples and pears and by the seventh minute I was talking about the justification of large cardinal axioms. The history fellow was Eamon Duffy, who I think quite liked the fact that I also had a draft paper on Berkeley. I never got round to publishing that paper – maybe one day!

HH: Some people love college life, others just can't bear it. How about you?

LI: An advantage of a JRF is the freedom that comes with the position. A slight disadvantage is that one is potentially more isolated than with other postdocs, so one needs to be a bit more proactive in seeking collaborations. I have very fond memories of my time at Magdalene. My wife and I got to live in the Cory House, a very old and charming house whose ceiling I could touch with my head in one of the rooms. At Magdalene, I also got to drink from a shared silver cup containing a secret liqueur, something not to recommend then and definitely not these days.

HH: OK, you did love it. And I guess it was during one of those arcane rituals that you came up with the main ideas for your work on rejection...

LI: I wish I could say I did! In reality it grew out of a side project. I normally like to have side projects. Perhaps strangely, it allows me to maintain more focus on what I am working on at a given time. An additional bonus is that side projects often take on a life of their own and become main projects. This is what happened with a side project on rejection. In 1996, Tim Smiley published a wonderful paper in *Analysis* in which he argued against the Fregean idea that rejection should be reduced to the assertion of a negation. During my PhD and JRF (during which I was still working on the philosophy of set theory) I published a couple of papers in *Analysis* defending Smiley. What I had not seen at the time is that Smiley's idea, later developed by Ian Rumfitt, to turn Frege on his head and explain negation in terms of rejection involved a completely general pattern of semantic explanation: explaining a linguistic expression in terms of some corresponding speech act and the attitude the speech act expresses. This general idea forms the basis of my ERC project *From the Expression of Disagreement to New Foundations for Expressivist Semantics*, which aims to develop inferential expressivism, a new approach to meaning combining aspects of the inferentialist and expressivist programmes in the theory of meaning. So, in effect, I moved from a side project in the philosophy of logic to a main project also encompassing the philosophy of language and other areas, including metaethics. Being in Amsterdam, one of the best places in formal semantics in the world, was very important for the development of these ideas.

HH: That's great. Can you tell us a bit more about what you expect from this project?

LI: One major aim of the project is to show that bringing together expressivism and inferentialism is beneficial to both views. Expressivists have long struggled with making sense of the behaviour of expressions in certain contexts, such as the antecedents of conditionals. This is what is known as the Frege-Geach problem. Inferentialism, on the other hand, has been typically applied to a very restricted set of expressions—typically, the core logical constants such as 'and' and 'or'—which has led many to deem it unsuitable as a general theory of meaning. In the project, we aim to show that inferential expressivism avoids the Frege-Geach problem and can be successfully applied to a vast range of expressions—epistemic modals,

probabilistic expressions and moral vocabulary to name a few. Another important component of the project is the logical framework we are developing in support of this semantic approach.

HH : Can you give us a rough idea about it?

LI: Sure. The framework is a multilateral logic that extends and generalises the bilateral systems developed by Ian Rumfitt and Smiley. In bilateral systems, sentences can be asserted and rejected; in the multilateral systems we are developing, one can investigate the logical relations between a variety of speech acts—assertion, rejection and many others. Finally, we are looking at language evolution from the perspective of the project. I mentioned earlier that the general idea behind the project is that it is possible to explain a linguistic expression in terms of some corresponding speech act and the attitude this speech act expresses. This leads to certain hypotheses about language evolution which we are investigating.

HH: You also mentioned the relevance of your project to metaethics. The pandemic reminded us that we are quite far from settling societal priorities. Do you see a long-term role for your field in this very challenging problem?

LI: Philosophy has always had and will no doubt continue to have a role to play in decision-making within society. This can range from understanding the very processes of decision-making and disagreement within society to evaluating the options that we are confronted with. Moreover, the societal impact of philosophy can be aided by a general recent trend in philosophy towards de-idealization, a trend which attempts to engage more directly with the analysis and understanding of real-world phenomena. In fact, one of my next projects is to develop an overall theory of real-world conversations, which can be seen as part of this general de-idealization trend. Having said all of this, in my view it is mistaken to think that philosophical understanding can replace actual political action. To paraphrase Kripke, there's no philosophical substitute for politics.

HH: Indeed! Speaking of worldly things, do you have any advice for early stage researchers in your field?

LI: Plenty of things to say here, of course, so I'll pick one: realise that what works for you need not work for others and vice versa. There's no golden rule for how to write a paper, how to give a talk or how to structure your working hours (or perhaps you don't even need to structure them). Also, you might even find that what worked for you at one stage of your career or of your life, no longer works at a later stage. So try to find your own style and habits by experimenting with as many as possible, but also continue to question them and possibly change them along the way.

HH: I'm sure this can be treasured by not-so-early stage researchers as well. Finally, would you like to share the title of a book/paper which you think is very important but vastly underappreciated in your field?

LI: Well, rather than a book or a paper, let me mention Cambridge Pragmatism, a strand in the pragmatist philosophical tradition that is not as well known as it should be, though of course several of its mains figures, such as Ramsey and Wittgenstein, are individually very well known. Fortunately, Cambridge Pragmatism is starting to be properly recognised, in large part thanks to Cheryl Misak's work, including her book *Cambridge Pragmatism: From Peirce and James to Ramsey and Wittgenstein* (so here's a book recommendation after all). I consider myself a pragmatist, and pragmatism is probably one of the main unifying themes of my work. In part, this is probably due my philosophical temperament, but I'm sure it is also in part due to the influence that Cambridge Pragmatism had on my philosophical development during my graduate studies.

Logic and Apophatic Theology

Apophatic or negative theology is to a good measure motivated by logical problems and paradoxes arising from divine attributes. In particular, three characteristics stand out in this (in)famous array of absurdities: omnipotence, omniscience, and omnibenevolence. Among the most notorious members of this scandalous set, one may refer to paradoxes of omnipotence, paradox of omniscience, and paradox of omnibenevolence. For omnipotence, paradoxes such as "Can an omnipotent being create an immovable stone?" or "Can an omnipotent being undo the past?" highlight the logical contentiousness of the concept. For omniscience, self-determined actions of free agents draw attention to the logical untenability of foreknowledge of open futures. And, for omnibenevolence, the putative problem of evil that "Can the existence of an omnibenevolent and omnipotent entity be reconciled with the existence of evil?" underscores the logical unsustainability of such a deity with the overwhelming presence of pain and suffering in the universe.

Apophatic theologians thereby propose the suspension of ascribing any attributes to the divine being and maintain that such an entity defies description in terms of any *positive* properties or characteristics. Thus, the deity in discussion is ineluctably indescribable and ineffable. However, cataphatic or positive theologians have been quick to charge apophaticism with a *self-referential* paradox: namely, the admission of apophatic theologians that the divine cannot be described breaches their own constraint by *describing it as an indescribable*. That is, the claim that the divine defies all description, nonetheless, seems to describe it since it attributes to the divinity the *property* of defying all definition or description.

Indeed, Sam Lebens takes this logical quandary of apophatic theology as the reason for the prevalent antipathy of analytic philosophers, as opposed to the continental tradition, towards negative theology. ('Negative Theology as Illuminating and/or Therapeutic Falsehood', in *Negative Theology as Jewish Modernism*, ed. M. Fagenblat, Indiana University Press, 2017, pp. 85-108) However, the contention here is that within the logical repertoire of analytic philosophy, there may be devices available to rescue apophaticism from self-referential stultification and thereby, at least in this respect, rehabilitate it in the domain of analytic philosophy. However, if there are still difficulties in assuaging analytic philosophers' apathy *vis-à-vis* negative theology, it may be other issues that transcend the analytic/continental dualism.

To see whether apophatic theology can be exonerated from a self-referential paradox, it should be first noted that at this level of discourse between apophatic and cataphatic theologians, there is a common cognitivist semantic theory of religious language to which both parties seem to subscribe. It appears that both camps converge on the idea that theological statements are truth apt and possess propositional content. In more detail, they seem to agree at least on three theses: (i) indicative theological statements are intended as representation of the world being in a certain way, (ii) they are typically used to make assertions about matters of fact, and (iii) in uttering such statements, they convey speakers' beliefs. Obviously

denying any of these theses opens the vista of theological non-cognitivism that simply renders the positive/negative divide in theology irrelevant and otiose.

So, is apophatic theology guilty of a paradox of self-reference? Apparently, there may be a way of extricating apophaticism from this liability by drawing on Alfred Tarski's distinction between *object-language* and *meta-language*, or more precisely and generally *the hierarchy of language use*, in his classic article, 'The Semantic Conception of Truth'. (*Philosophy and Phenomenological Research* 4: 341-76, 1944) In discussing the problem of defining truth against the backdrop of the paradoxical consequence of statements such as the liar antinomy, Tarski suggests that 'we have to use two different languages': the first is the language which is "talked about" and the second is the language in which we "talk about" the first language. (*Ibid.*, p. 349) In this dichotomy, statements involving the concept of truth are strictly speaking not uttered at the same level of language use and should be lassoed into the two different levels of object and meta utterances.

Applying this Tarskian schema to the pronouncements of negative theology, the claim is basically that the two core propositions of apophaticism, *viz.*, (a) impossibility of positive descriptions of the divinity and (b) the negative description of indescribability of the divine, are not at the same level of language use and thus their utterance should be bifurcated into two different levels of linguistic expression. Should this separation of linguistic levels of utterance work, it would show that apophaticism may not be self-stultifying after all. The application of the Tarskian schema to the negative theologian's endeavour may run thus: having observed the logical problems, paradoxes, and absurdities arising from the articulation of positive properties of the deity, the apophatic theologian reaches her position that there is not a single logically unproblematic positive property of the divine. On this model, ordinary religious statements concerning the nature of the divine are at the *object* level as they are on a par with all the other religious statements made by individuals in other societies and cultures across time and space. In other words, common-or-garden theological statements about the divine are uttered at the same level of language that is "talked about". Yet, the apophatic theologian's own pronouncement is at a *meta* level of language use since she is "talking about" the language in which religious believers "talked about" their theological beliefs concerning the nature of the deity.

Thus far there is no self-referential paradox, but neither is there any talk of divine indescribability. So, where does indescribability fit into this model without degenerating itself into self-contradiction? Is not paradoxical describability at the *same* meta-level of language use as the doctrine of apophatic theology itself, thereby threatening the cogency of the position? That is, the statement that (a) 'the deity cannot be described by any positive property' is at the same level of language use as the statement that (b) 'the divine is indescribable'.

But, is (b) at the same level as (a)? At this juncture, the negative theologian seems to have two options to disentangle herself from the predicament of paradox. (I) She may concede that claims (a) and (b) are at the same meta-level but deny that they belong to the same category of statements. That is, not all meta-level statements are about the same subject: although both claims are normative by nature, the first one has its normativity rooted in logic whereas the second one is anchored in epistemology. In other words, indescribability is an *epistemic* norm given that from an epistemological perspective there is no way to privilege one property over another. The norms involved in the two statements belong to different species of normativity, and thus the *impossibility* of one is not in conflict with the *possibility* of the other.

However, if the distinction between different types of normativity fails to have purchase on hardcore cataphatic theologians, the apophatic theologian may resort to a second option (II): unlike option (I), it is denied that propositions (a) and (b) are at the same level of language use. The negative theologian reaches proposition (a) at the *meta*-level while observing how paradoxes emerge from religious beliefs about the nature of the divinity at the *object*-level. However, we are still none-the-wiser about the apophatic theologian's epistemology: that is, what is her epistemic assessment of the paradoxicality of positive properties of the divine? Obviously, given her meta-logical position, she cannot subscribe to any positive description of the deity. Her meta-logical stance bars her from endorsing any positive divine property unless she descends to the object-level where she can approve or disapprove of any positive property considering whether it leads to any paradox or not. But, once she arrives at her meta-logical position of apophaticism, she reaches a stalemate in terms of epistemic assessment of any positive description of the deity. It is at this point that the negative theologian is forced to ascend to the next level of discourse, *viz.* the *meta-meta*-level, where she can propose the epistemic statement of divine indescribability.

Should the application of such a Tarskian hierarchy to the negative theologian's endeavor work, then apophaticism can consistently maintain its rejection of the possibility of any positive description of the divinity with the simultaneous recommendation of divine indescribability since the two theses are enunciated at two different levels of discourse.

Majid Amini
Philosophy, Virginia State University

The Paradox of Epistemic Obligation Avoided

We often claim that experts, in particular, ought to know certain things. But Lennart Åqvist pointed out that statements of the form $O(K_S\, p)$, i.e. that it is obligatory that S knows that p, appear to be paradoxically problematic (1967: "Good Samaritans, Contrary-to-duty Imperatives, and Epistemic Obligations," *Nous* 1,361-379). Here it will be shown that this paradoxicality can be avoided by careful analysis of such claims framed in terms of the standard logic of knowledge (KDT) and the standard system of deontic logic (SDL).

The axioms of KDT are as follows (see Hendricks, V. and J. Symons (2015: "Epistemic Logic," *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.)):

(P) All well-formed tautologies.
(K) $K_S(p \to q) \to (K_S\, p \to K_S\, q)$
(D) $K_S\, p \to \neg K_S \neg p$
(T) $K_S\, p \to p$
(4) $K_S\, p \to K_S\, K_S\, p$
(5) $\neg K_S\, p \to K_S \neg K_s p$
MP If p and $p \to q$, then q.
NEC If p, then $K_S\, p$

Here $K_S\,p$ is understood as "S knows that p" and the K operator is tied to the JTB+ analysis of knowledge:

(JATB+) S knows that p, if and only if,

(i) S believes that p,

(ii) S's belief that p is justified,

(iii) p is true and

(iv) S's justified belief that p meets a required "additional anti-luck condition(s)".

Conditions (i)-(iii) are fairly orthodox and condition (iv) is a more controversial placeholder intended to rule out Gettier cases. One of the most popular variants of (iv) is the modal safety condition (see, for example, Pritchard, D. (2005: *Epistemic Luck*, Oxford, Oxford University Press) and Pritchard, D. (2007: "Anti-Luck Epistemology," *Synthese* 158, 277-298)):

$$\text{(Safety)}\,(w_i \vDash K_S\,p) \rightarrow \neg\left[\langle w_i\rangle \vDash (B_{S'}\,p\ \&\ \neg p)\right].$$

Here w_i is world i, $\langle w_i\rangle$ is the set of worlds sufficiently close to w_i, and $B_S\,p$ represents that S believes that p. The safety condition is the claim that if S knows that p, then there are no worlds sufficiently similar to w_i (including w_i) where S' (S's counterpart in those worlds) believes that p and p is false.

SDL is the standard system of deontic logic and its axioms are as follows (see Hilpinen, R. (2001: "Deontic Logic," in L. Goble *The Blackwell Guide to Philosophical Logic*, Malden, Blackwell Publishers, 159-182) and McNamara, P. (2019: "Deontic Logic," *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/fall2018/entries/logicdeontic/):

(P)	All well-formed tautologies.
(K)	$O(p \rightarrow q) \rightarrow (Op \rightarrow Qq)$
(D)	$Op \rightarrow \neg O \neg p$
MP	If p and $p \rightarrow q$, then q.
NEC	If p, then Op

Here O interpreted as "obligation". Given SDL and KDT, so understood, we have the logical and conceptual tools in hand to examine statements of the form $O\,(K_S\,p)$ in a precise manner.

So, as Åqvist pointed out (Åqvist 1967), there appear to be paradoxically odd consequences of such statements framed in terms of KDT/SDL. Let us begin by introducing a case in order to help show this:

> Sarah is a competent and professional doctor. John, a patient, comes to see her. John exhibits a fever, fatigue, swollen lymph nodes, a headache, and Koplik's spots. Sarah sends a sample of his blood to the lab for testing and the test report comes back indicating that John's blood contains significant infection by *Measles morbillivirus*.

Here it is natural to think that Sarah *should know* that John has measles. But consider what the SDL/KDT logic implies about statements of the form $O(K_S\,p)$:

$O(K_S\,p)$	[Assumption]
$K_S\,p \rightarrow p$	[Principle T of KDT]
$\therefore$ Op	

In this case, this amounts to the idea that if Sarah should know that John has measles, then it is obligatory that John has measles. This is, of course, absurd. But, in virtue of what does this paradoxicality arise and can it be avoided?

The suggestion made here is that statements of the form $O(K_S\,p)$ are conceptually problematic due to the opacity of the K operator and that there are really no such obligations to know. Thus, the solution to the paradox of epistemic obligation offered here is a skeptical one. This is the case despite the fact that expressions of the form "You ought to know that p" are commonplace in natural languages. So, why can't we have such obligations? We cannot have such obligations because we cannot have obligations with respect to conditions (ii)-(iv) of the JTB+ analysis of knowledge. This is specifically due to the Kantian "ought implies can" principle, the idea that if p is obligatory, then it must be possible for S to directly and voluntarily bring about p:

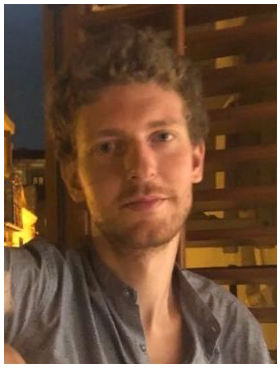
$$(OC)\ Op \rightarrow \diamond p$$

OC is a theorem of SDL. But, due to how OC applies to the K operator, it turns out that what we ordinarily take to be obligations to know can't be anything more than obligations to believe-which we (and Sarah) can have-that are conjoined with (1) non-obligatory and objective relations between what is believed and external factual and modal states and (2) non-obligatory and objective relations between evidence and what is believed. So, the problem arises in virtue of some confusions about obligations and the distinction between knowledge and belief. This confusion of obligations to know and to believe is, however, obscured by the fact that in looking at epistemic obligations in terms of the form $O(K_S\,p)$ the K operator hides the externalist aspects of knowledge about which there can be no obligations. This includes that one's possessed evidence is adequate for one's belief that p, that p is true, and that p meets (iv). So, once we see this the paradox of epistemic obligation vanishes and need to revise SDL/KTD to reflect this.

Michael J. Shaffer

Dissemination Corner

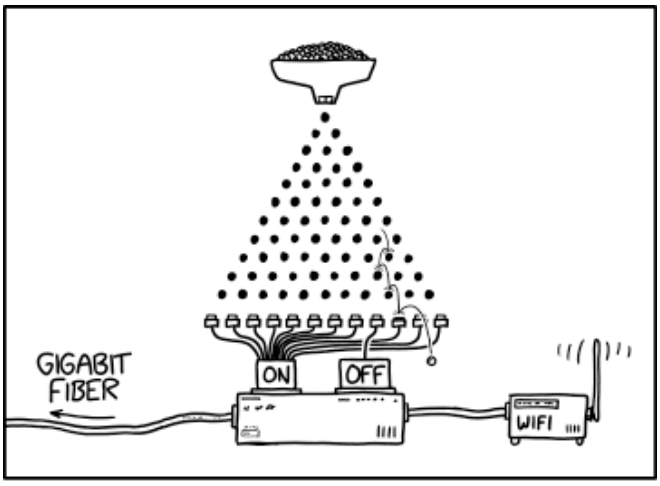
BRIO



Throughout the last decade, Machine Learning (ML) systems and the availability of huge amounts of data have incredibly advanced the state of the art of Artificial Intelligence (AI). These technologies have managed to impress non-experts with their ability to mimic human intelligence, often making the headlines of important newspapers: think, for example, of the blog post written by GPT-3 for The Guardian and the more recent widespread news about LaMBDA "passing the Turing Test" and claimed to be conscious by a Google engineer who wouldn't shut it down. Although this news usually exaggerates what these systems can *actually* do, it is not only the non-experts who find them to overachieve in domains previously thought to be prerogative of humans.

As a consequence, our society has gradually given in to the temptation of using them to support (and, sometimes, delegate) decision-making. However, it is important not to forget that the data they are trained on is primarily human-produced and may therefore replicate typical prejudices of human societies, echoing an imaginary red thread that connects François Galton's idea of eugenics with, e.g., well-known cases of discrimination such as COMPAS. For this reason, both non-experts and experts have (rightfully) started being concerned about how we use these systems: the former are looking to be reassured when AI takes sensitive decisions, especially those that can have serious (e.g., legal, medical) impact on them. On the other hand, experts are looking for (and, at least for now, struggling to find) ways to realize AI systems people can trust. This coincides with the ultimate goal of the BRIO project, which brings together researchers from different areas who share the aim of advancing our understanding of Bias, Risk and Opacity in AI with the idea of achieving Trustworthy AI (TAI). For the reader who is looking to catch up with BRIO I suggest looking at the previous issues of *The Reasoner* for a general overview of the project (see The Reasoner, Volume 16, Number 1, January 2022) as well as a description of Objective 1 (see The Reasoner, Volume 16, Number 3, May 2022) and Objective 2 (see The Reasoner Volume 16, Number 5, August 2022). BRIO is carrying out a multi-faceted analysis of biases and fairness, including a philosophical investigation of trust and ontological modeling of bias types. Together with the unit led by the LUCI Group at the University of Milan (https://luci.unimi.it), I am working as an external collaborator on formal approaches to TAI, positioning them in the broader philosophical picture of fairness drawn by the other research units. To interact with opaque ML classifiers, we have developed a natural deduction system dubbed TPTND (short for "Trustworthy Probabilistic Typed Natural Deduction", see D'Asaro and Primiero (2022: Checking Trustworthiness of Probabilistic Computations in a Typed Natural Deduction System, https://arxiv.org/abs/2206.12934) for a preprint) to check that the distribution of a classifier's outputs does not diverge much from a distribution that is assumed to be fair. TPTND is a tool that aims to ensure *equality of outcomes*, i.e., it checks that the outcome of a process fairly distributes some desirable good across the reference population and does not yield unwanted side-effects. TPTND allows for some degree of freedom when it comes to modeling how close the output distribution and the reference distribution are, and therefore several criteria can be employed, e.g. the binomial test and the Kullback–Leibler divergence. Such flexibility allows for some easy changes to the formalism, e.g., to match it with the results emerging from the other units' analyses. In this direction, other ongoing work of the LUCI unit includes Termine et al. (2021: Modelling Accuracy and Trustworthiness of Explaining Agents, LORI 2021: Logic, Rationality, and Interaction, pp. 232–245), which proposes a formal definition of trustworthiness and accuracy of an opaque model with respect to a transparent one in terms of Probabilistic Tree Computational Logic operators. In Explainable AI terminology (see Guidotti et al. (2019, A survey of methods for explaining black box models, ACM Comput. Surv. 51(5), pp. 1–42) for a classic survey on the topic) these methods are called *post-hoc* in that they aim to explain a model after its training phase has finished, and one only has access to the model's predictions. An example of a post-hoc method that produces logical theories from any opaque system was presented in D'Asaro et al. (2020: Towards an Inductive Logic Programming Approach for Explaining Black-Box Preference Learning Systems, Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, pp. 855–859) that started exploring the use of the ILASP inductive logic programming framework (available at www.ilasp.com) to build logical theories from opaque systems for preference learning. This and similar methods attempt to extract logical explanations from existing opaque models, thus making them more trustworthy. Another possible line of research is that of building *transparent-by-design* algorithms that can be interrogated and explain their decisions, as in the case of language EPEC, introduced in D'Asaro et al. (2020: Probabilistic reasoning about epistemic action narratives, Artificial Intelligence, Volume 287, 103352) and recently presented at IJCAI in D'Asaro et al. (2021, https://ijcai-21.org/videos-slides/?video=J26). Implementations of EPEC are being used for several tasks where explaining decisions is key, e.g., in medical decision-support and legal reasoning. The presented formal efforts tackle the problem of trustworthiness and explainability from different perspectives, and other teams and researchers are currently investigating several other options. We believe that a solution to TAI must be interdisciplinary. For this reason, BRIO, in collaboration with a broader community, is organizing the BEWARE Workshop (https://sites.google.com/view/beware2022), soon to be held in Udine (Italy) and co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022) that will take place from the 28th of November to the 2nd of December 2022. BEWARE will focus on some of the core objectives of BRIO, such as ethical and epistemological aspects of AI, bias, risk, explainability, the role of logic and logic programming. Come join us if you think you can contribute to the cause!

Fabio Aurelio D'Asaro
University of Verona



What's Hot in ...

Statistical Relational AI

These lines are written just a day after my return from the Federated Logic Conferences (FLoC) in Haifa, Israel. This has been my first in-person conference since the pandemic, and I am still under the spell of the wonderful atmosphere such a large gathering entails: Mediterranean evenings and streetside cafés lined with logicians bending over notepads and discussing the issues of the day.

FLoC is a consortium of several of the most prominent conferences on logical methods in computer science held at a single venue across two weeks. The first week included Logic in Computer Science (LICS), Knowledge Representation and Reasoning (KR), Satisfiability (SAT), Formal Structures for Computation and Deduction (FSCD), Constraint Programming (CP) and the International Conference on Logic Programming (ICLP), which were preceded by a large array of workshops and tutorials associated with each of the main conferences. The diversity of attendees from different communities gives FLoC its distinct flavour. One can take in the most relevant presentations at each of the parallel conferences by jumping between the various sessions. This gives a broad overview of research done across the related but often somewhat isolated fields, which is further enhanced by the large number of presentations on recently published research from journals and related conferences. In this way FLoC is truly a unique experience for the computer-science-aligned logician.

Such an impressive offering makes it more difficult to pick out a single *theme*, even when restricted to contributions in statistical relational artificial intelligence. However, something I took from Haifa was the great interest in and sophisticated work around *Probabilistic Answer Set Programming*, a variety of the probabilistic logic programming paradigm based on the *stable model semantics* of Answer Set Programming (ASP).

The roots of ASP go back to 1988, when the stable model semantics was proposed to deal with the perennial problem of how to handle negation in logic programming. Two examples may serve to illustrate the issue. Fist, consider the following definition of a predicate, remembering that ":-" stands for "is implied/caused by" and expressions such as "p." stand for facts.

```
p.
q.
r :- p, not s.
s :- q.
```

The interpretation of this program can be derived by following the rules from the facts, in the following manner:
"p and q are true as facts, hence s is true, and therefore r is not true, because r can only be caused by $\neg s$."

This seems straightforward enough; however, the following program is much harder to interpret in this way:

```
p :- not q.
q :- not p.
r :- p.
r :- q.
```

If we would try, we would end up in a twist:
"As there are no facts in the program, p and q are false. Hence, p and q are true by the first two rules. But then the rules shouldn't have been invoked, since p and q are not false."

The stable model semantics aims to resolve the issue of this cyclical negation by accepting two models as "stable" for this program: One in which p is true and q is not, and one in which q is true and p is not. These are referred to as the answer sets of the programming, giving ASP its name. Note that r would be true in either case, and is therefore contained in all answer sets of the program.

When the distribution semantics was first introduced to extend ordinary logic programs to probabilistic logic programs, its key idea was to combine a probability distribution arising from independent probabilistic facts with a logic program which defines derived predicates in terms of those basic probabilistic facts. Adapting this construction to ASP raises the question of how to deal with multiple answer sets for a single given program (rather than the unique definition an ordinary logic program provides).

Several proposals have been made in the past. In order to return a single probability distribution, one can simply give every answer set equal likelihood, corresponding to a maximum entropy assumption. Alternatively, the LP-MLN language moves beyond the distribution semantics by adding weights to clauses and weighting structures depending on how many rules they satisfy.

One could also give up on the idea of returning a single distribution by admitting any combination of answer sets for a given interpretation of the probabilistic facts (the *credal semantics*). Queries then return probability intervals, providing the minimum and maximum probability that can arise from choices of answer sets.

These foundations give rise to many interesting extensions and reasoning systems proposed at the conference, starting with *PASTA*, a language which incorporates statistical statements of the form "Between 80% and 95% of birds fly" and compiles them into probabilistic answer set programs under the credal semantics, via *plingo*, an inference module for LP-MLN programs incorporated into the state-of-the art answer set solver clingo, to the *L-stable semantics* which extends the credal semantics to allow for cases where some interpretation of the probabilistic facts renders the logic program inconsistent.

Of course, all this wonderful expressivity comes at a cost; while inference in ordinary probabilistic logic programs is already intractable, probabilistic answer set programming has considerably higher complexity. This makes research into approximate answer set counters such as *ApproxASP* or into knowledge compilation for higher-order inference tasks (*2AMC*) particularly valuable.

The beauty of FLoC is precisely that contributions like these can come together at a single place — in fact, the five innovations just described have been presented at four different venues; PASTA as recently published research at the Probablistic Logic Programming workshop (originally accepted at LPNMR 2022), plingo at the Answer Set Programming workshop, the L-Stable semantics at KR and ApproxASP and 2AMC at the ICLP.

Felix Weitkämper
Computer Science, LMU Munich

Uncertain Reasoning

Argumentation theory finds its motivation in one of the main human activities: reasoning.

Human reasoning, which goes from arguing about something, and defending our own opinion to performing scientific reasoning and, as we will see, statistical reasoning, can be fruitfully formalised by Dung's argumentative paradigm as introduced in (Dung, "On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games". *Artificial Intelligence*, 77(2):321–357, 1995). The investigation of the intrinsic connection between abstract argumentation theory and different logical aspects has been a fertile field of research. In particular, by instantiating the arguments with Gentzen's sequents as done by Arieli and Straßer in, e.g., (Arieli and Straßer, "Sequent-based logical argumentation" *Argument & Computation*, 6(1):73–99, 2015) *dynamic derivations* have been introduced and crucial aspects of argumentation such as the generation of new arguments from existing ones and the exclusion of some other that were previously derived can be performed in an automated way.

The rules of a specific calculus, e.g. **LK** (the sequent calculus sound and complete with classical logic), can be used for the definition of new arguments. For example, if $\Gamma \vdash \alpha$ is an argument (the support of this argument is Γ and its claim is α) and $\Gamma \vdash \beta$ is another argument, then also $\Gamma \vdash \alpha \wedge \beta$ is an argument. The generation of the argument $\Gamma \vdash \alpha \wedge \beta$ is the result of the application of the rule $(\wedge, r)$of **LK**. Additional rules, known as sequent elimination rules, have also been defined and allow to discard of a derived argument. Sequent elimination rules have a form similar to the inference rules, except that the conclusion is the elimination of one of its premises. In general, the premise of a sequent elimination rule has three parts: on the left side, there is the attacking argument (a sequent), on the right the attacked argument (still a sequent) and between the two the attacking condition. As conclusion, there is the attacked argument, i.e. the argument that has been eliminated through the attack. By performing a dynamic derivation, that consists of the application of inference rules of some fixed calculus and sequent elimination rules, some arguments are accepted and some are not. Dynamic derivations illustrate the non-monotonicity of the reasoning process in argumentation theory. In these derivations arguments can be challenged by counter-arguments. Thus, a specific sequent may be considered not derived at a particular stage of the derivation even if in an earlier stage of the derivation it was considered derived and vice versa. An eliminated (sequent-)argument may be restored if its attacking (sequent-)argument is counter-attacked through a new elimination rule. At the end of the derivation, only some arguments are considered acceptable.

Statistical analysis has been introduced to extract information from data and —obviously—human reasoning plays a central role in the way data are processed. Different theories might be applied and consequently different results might be inferred. The non-monotonicity of statistical inference derives from the acquisition of additional data or a change in the statistical assumptions. Seen as a specific kind of human reasoning, the statistical analysis of a given dataset could be interpreted in a sequent-based argumentation frame where the data are the arguments and the statistical inference method considered are encoded in the rules allowed to be used in the dynamic derivations. Thus, the acquisition of new data can be seen as an update of the given argumentative framework where new arguments are introduced. Different statistical analyses would result in different rules and consequently different calculi. In this interpretation, the non-monotonicity of statistical inference is preserved and by considering different calculi, i.e. different statistical models, some arguments might be easier to justify than others. The interest in establishing such a connection is twofold. On one side we might use sequent-based argumentation theory to make explicit the statistical theory and assumptions used in the analysis of a given dataset, on the other side, we could have a new perspective on the interpretation of the conjunction of two arguments. In statistical analysis, the aggregation of two different datasets might motivate the need for the use of methods that account for possible lack of homogeneity in the process of data compilation. Thus, e.g., the conjunction of the arguments $\Gamma_1 \vdash \gamma_1$ and $\Gamma_2 \vdash \gamma_2$, $(\Gamma_1 \vdash \gamma_1)\&(\Gamma_2 \vdash \gamma_2)$, shall not be simply defined as $\Gamma_1, \Gamma_2 \vdash \gamma_1 \wedge \gamma_2$. The use of the symbol & to denote the conjunction of two arguments is due to enhance the different levels the logic operates: between arguments and inside the arguments. The analysis of data aggregation might help in the investigation of the connections between these two levels the logic operates. The introduction of *weighted* dynamic derivations (dynamic derivations that operates on sequent-based argumentation frames where the attack relations are grated) might be necessary to establish this novel connection between argumentation theory and statistical analysis.

Esther Anna Corsi
LUCI Group, University of Milan

Events

September

'True': The Meaning(s) of 'True,' University of Bristol, 5–6 September.

MOSAIC 2022: Modalities In Substructural Logics: Theory, Methods And Applications 2022) Paestum, Italy, 4–11 September.

PFN: 'Philosophy & Finance Network', Rome, 8–9 September.

FoCoRe: Formal and Cognitive Reasoning, Trier, Germany, 20 September.

ME& CS: Model Evaluation and Causal Search: Empirical and Experimental Approaches, Pisa, Italy, 25–27 September.

E&EiS: Engaging Ethics and Epistemology in Science, Leibnizhaus, Hannover, Germany, 29–30 September.

Courses and Programmes

Courses

LAIS: Logic for the AI Spring, 12–16 September.

Programmes

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.

Jobs and Studentships

Studentships

12 PhD grants: in Mind, Brain and ReasoningThe Human Mind and Its Explanations: Language, Brain and Reasoning , University of Milan, deadline 2 September.

Jobs

Associate Professorships: in Philosophy of Language/Logic and Epistemology, University of Oslo, deadline 31 August.