The Interview Column

BY

Chen Avin and Stefan Schmid

Ben Gurion University, Israel and TU Berlin, Germany
{chenavin,schmiste}@gmail.com

Know the Person behind the Papers

Today: Antoine Amarilli

Bio: *Antoine Amarilli has an Advanced Research Position at Inria Lille in France, and is on leave from an associate professor position at Télécom Paris. He studied at École normale supérieure and received his PhD in 2016 from Télécom Paris, for which he was awarded a Beth Dissertation Award. His PhD focused on probabilistic databases and open-world query answering: his advisor was Pierre Senellart. He was the director of the ICPC SWERC programming contest from 2017 to 2020. He served as the inaugural managing editor for the TheoretiCS journal from 2021 to 2023, and was a local organiser in 2022 for Highlights of Logic, Games, and Automata, for which he is now environmental chair. His current research focuses on database theory and theoretical computer science; he received his habilitation from Institut Polytechnique de Paris in 2023. He co-authored works which received best paper awards at the ICDT conference in 2020 [2] and at the ICALP 2021 conference (track B) [3]. Since 2024, he is a vice-president of EATCS.*

EATCS: We ask all interviewees to share a photo with us. Can you please tell us a little bit more about the photo you shared?

Antoine: This is a selfie from a small hike near Lake Tahoe, which I did during my fall semester at the Simons Institute in 2023. This was my first overseas trip since COVID, so I wanted to make the most of it!

EATCS: Can you please tell us something about you that probably most of the readers of your papers don't know?

Antoine: I was passionate about computers as a kid – programming video games, hacking calculators, playing pranks on my classmates, etc. I was steered towards theory by the French education system, but for a long time I tried to resist – I was quite dismissive of teachers who called themselves computer scientists but couldn't even host their own mail server or compile their own kernel.

I eventually got interested in academic research during my internship with Pierre Senellart in 2012, but topic-wise I was still working on knowledge bases, in a futile attempt at practical relevance. It's only at the end of the internship that I embraced theory. I visited Fabian Suchanek at MPI Saarbrücken, and instead of knowledge bases I got nerdsniped by an abstract question of his. It was about tree labelings: if you label the vertices of a rooted tree with their preorder and postorder timestamp, then you can test in $O(1)$ if a node n_1 is a descendant of another node n_2, simply by retrieving their labels and checking if the interval formed by the timestamps of n_1 is included in that of n_2. Fabian observed that this approach could be generalized to certain DAGs, and I spent the visit trying to understand precisely for which DAGs this worked. Figuring it out was quite fun, and (I must admit) very rewarding in terms of ego. So I shifted towards database theory for my PhD with Pierre, and since then I have been continuously navigating towards the kind of theoretical questions that I like.

EATCS: Is there a paper which influenced you particularly, and which you recommend other community members to read?

Antoine: Let me instead confess an embarrassing truth: I really don't read many papers in detail (except as a reviewer). I look at dozens of papers per month but only to skim them: I look for related work on topics that I'm interested in, I search for answers when I ask myself questions, I stay informed about what people in my area are working on, etc. But typically I don't read papers from beginning to end, unless it's a life-or-death necessity for the stuff I already want to do. My exposure to outside ideas mostly comes from talks and from people explaining their work in discussions.

Maybe it means I'm missing out on some things, but I'm not sure – I already feel like I'm stretched too thin... I also think we should be bolder and reach out to other researchers (at conferences or online) to chat about their work even if we haven't read it. Further, I suspect our papers are read by far fewer people than we'd like – in the academic career game, you score points by writing papers, not reading them. But this doesn't mean that writing those papers isn't worthwhile!

EATCS: Is there a paper of your own you like to recommend the readers to study?

Antoine: It's hard to pick one, but I like our results in [4]. This paper is about enumerating the words of regular languages. Say we have a regular language L, and we want to write L as a (generally infinite) sequence of words: $w_1, w_2, \ldots$. We want to ensure that there is a uniform bound $c \in \mathbb{N}$ on the distance between any two consecutive words, i.e., each word w_{i+1} can be produced from the previous word w_i by changing at most c characters. For $L = a^*$ for instance you can simply write $\epsilon, a, aa, aaa, \ldots$. But for some languages it is impossible to construct such sequences, e.g., for $L = a^* + b^*$ (do you see why?). And sometimes we can build such sequences in non-obvious ways, e.g., for $L = (a + b)^*$ we can use a Gray code. Hence the question – for which regular languages do such sequences exist?

Our paper [4] gives a characterization of these languages based on a criterion on deterministic automata, which (much to our surprise) can be tested in PTIME. What I like about this paper is that I find the problem very easy to state, and the proofs are non-obvious but also do not use any sophisticated techniques. This makes me feel that theoretical CS is still a young field where you don't need many prerequisites to find fun questions to explore.

EATCS: When (or where) is your most productive working time (or place)?

Antoine: I wish I had a recipe to be consistently productive! My ability to get the right things done varies a lot from one week to the next. My working hours have also changed a lot during my career. E.g., during my PhD I was a night owl – I would frantically write until around 2-3 AM, and on mornings I'd barely make it to the lab in time for lunch. Nowadays I have a more standard schedule.

That said, one thing I now understand is that good company is key for me to be productive and to have fun: I'm much more reliably focused and efficient when I work with others. This goes for research discussions, but also for writing – in a "pair programming" kind of way.

EATCS: What do you do when you get stuck with a research problem? How do you deal with failures?

Antoine: Another confession: I spend very little time stuck on problems, because I'm very impatient and lose interest very quickly if I'm not making progress. At any given moment I have a lot of open directions on which it's feasible to move forward. So if I start thinking about a new question and run out of ideas to try, then I'll quickly move on to something else. The main exception is when I convinced myself (or, worse, convinced others) that a question should be solvable, but then the approach breaks – at this point I'll get very annoyed and work hard to save it.

I'm not sure my way of doing things is optimal, because it biases me towards easier questions – maybe I should learn to embrace failure more and spend more time on high-risk endeavors. Yet, I think it's reasonable to allocate one's efforts wisely, and work in priority on projects where you're in a good position to bring new techniques or new ideas and can reasonably hope to find a solution. It's true that you can sometimes break problems by relentless frontal attacks, but also it's often the case that you're just doing it wrong and the solution will have to come from elsewhere, e.g., using techniques from another field.

EATCS: How do you choose what to work on and what kind of impact are you hoping your work will have?

Antoine: Do we choose problems, or do problems choose us? I find there are some problems that immediately fascinate me [1] – they feel "natural". And I get more and more picky with time, and find it harder and harder to get interested in research unless I can see how it relates to the kind of problems I care about.

That said, in the past I have often been seduced by problems that seemed natural but turned out with hindsight to only make sense in a narrow community and at a specific time. So I try to shield my sense of naturalness from the appeal of trendy topics, by checking a few things: Can I explain in elementary terms what the question is and how we ended up there? Would the problem appeal to a more general audience? (Maybe not to random people on the street, but to a motivated undergrad, or to my friends with a CS/math background?) Is this problem the first question we should solve, or is there something simpler that we already don't understand?

In terms of impact, my hope is that I can contribute a little bit to the progress of theoretical research. I don't hope that my work will ever directly influence practical computing or even practical research, and I'm fine with that. Theory matters for its own sake – and I also believe it helps practice, but in indirect and global ways that are often hard to attribute to individual papers or researchers. I also think it's valuable and important to develop the use of advanced theory to solve practical problems, but personally I don't claim to do this – in my opinion this job must be done in collaboration with people who are actually getting their hands dirty to achieve something in the real world. I don't believe in papers that target hypothetical practical needs but didn't start as an attempt to solve a concrete problem – I find they are usually neither theoretically interesting nor practically relevant.

EATCS: What research topics or approaches do you wish the community did more of?

Antoine: I don't have specific topics in mind, but in terms of approaches I think TCS would benefit from a more open culture: we should be sharing our ideas earlier and more broadly, instead of keeping them private until they can be posted as finished papers.

Here's how research currently seems to be done around me. You first find new directions to explore via discussions – with existing collaborators or with new people that you met at a conference or other event. Once the collaboration is started, everything happens in private, except maybe for some in-person gossiping at conferences or with colleagues, plus perhaps a few targeted questions sent by email to specific people. Nothing is visible to outside observers until the paper in finished and posted on arXiv.

All of this may have made sense in the pre-Internet age, but today it feels inefficient. Why aren't we publishing more things online than finished papers? To be clear: I'm not saying everything we do ought to be online, or that in-person discussions aren't valuable, or that major results shouldn't be written up as polished papers. But if only final results see the light of day, I think we're missing out, for several reasons. First, most projects just get postponed indefinitely. Lots of the time I spend on research ends up having zero tangible outcome because there's just not enough time to finish everything. So, if our research notes and drafts were in the open the entire time, then at least some usable trace of the effort would be available for others. Second, research discussions typically lead to many more open questions and ideas than what eventually appears in papers: why not post these questions (e.g., on `cstheory.stackexchange.com`) and see what it leads to? Third, online discussions are more inclusive than in-person conference gossip, and they scale better: they are open to everyone with a search engine, which includes researchers in other communities, people who don't often attend conferences, curious students you haven't met yet, etc. I think it's crucial that we reach out to people outside the core TCS community (not just the few students we directly supervise). So we should be making it easier for everyone to see what we are doing and get involved – and expensive onsite conferences and finished papers behind paywalls are not the best way to attract newcomers.

Overall, I think that much progress in research is made possible when ideas from different places can meet – and it's essentially impossible to anticipate what will have long-term impact and who it will inspire. So we should be throwing lots of our ideas around, to see what sticks. But currently we are blindly investing huge amounts of effort to write a small number of finished papers – faster and broader iteration only happens with our direct collaborators.

So why are we so hesitant about putting things online? Mostly for superficial reasons: we are worried about attracting parasitic co-authors, getting our ideas stolen, or embarrassing ourselves by posting something stupid or wrong. That said, I think it's mostly a question of social norms, given that so many successful Internet-native communities are open by default (e.g., Wikipedia, open-source projects...).

EATCS: Do you see a main challenge or opportunity for theoretical computer scientists for the near future?

Antoine: I think our main challenge is that we rely too much on conferences – to validate results, and also to hand out community recognition. Conferences can be great, but we should attend them by choice, not by obligation. As it turns out, many of the historical functions of conferences have already been taken over by other channels: e.g., preprint servers (to make papers available), Google Scholar recommendations (to advertise them to relevant people), etc. But to get street cred as a TCS community member (and to make a living doing TCS), for now there is no alternative: you need papers in the big conferences. (Yes, I know about journals, but they operate on unpredictable timelines and are often less prestigious than conferences. So they are a not a good first option – at least in the common case when some co-authors are still looking for a permanent research position.)

The problem with most conferences is that they require in-person attendance. And sure, meeting people in-person is a fantastic opportunity when possible – in particular for new students. But I see two main problems with this culture of mandatory travel.

The first reason is inclusivity. Not everyone can easily travel to conferences, for various reasons: visa restrictions, lack of funding, health issues, childcare responsibilities... So, currently, we don't even see the many people that the conference publication system leaves out. (EATCS itself is a good example of this: people mostly join via in-person attendance to member conferences like ICALP.) I think our community should strive to include everyone who does TCS, not just the privileged subset of people who travel regularly.

The second reason is climate change. To reduce greenhouse gas emissions to sustainable levels, we must fly less – our initiative TCS4F [5] is a way to advocate for this. I'm passionate about TCS research and how it can focus on questions that will only bear fruit in the very long term; but I am also worried about our climate trajectory in the short term. Honestly, I'm not sure our research is worth the tons of CO_2 which are essentially required nowadays to build up an internationally competitive CV.

So here is what I think is our big challenge: decoupling community recognition and conference travel. I think it is also an opportunity: if we modernize our practices and make our community more accessible, we can stay relevant and attract talent that we would otherwise miss.

Please complete the following sentences?

- *Being a researcher is ... a job first and foremost. (At some point I thought it was a passion, but looking more closely: I do not see many hobbyists participating to the kind of research I do, and I see that we all spend a large part of our time on tasks that we would never do if we weren't professionally pressured to do them.)*
- *My first research discovery was ... complexity bounds on frequent itemset mining in taxonomies by posing queries on crowdsourcing platforms. Yes, my tastes have changed a bit since then.*
- *Enjoying research is ... not the easiest kind of pleasure in life!*
- *Theoretical computer scientists 100 years from now will ... possibly not exist anymore – unless we manage to do what's needed to steer our civilization towards a more sustainable trajectory (geopolitically, environmentally...)*

References

[1] A. Amarilli. List of open questions, 2025. `https://a3nm.net/work/research/questions/`.

[2] A. Amarilli and I. I. Ceylan. A dichotomy for homomorphism-closed queries on probabilistic graphs. In *ICDT*, 2020.

[3] A. Amarilli, L. Jachiet, and C. Paperman. Dynamic membership for regular languages. In *ICALP*, 2021.

[4] A. Amarilli and M. Monet. Enumerating regular languages with bounded delay. In *STACS*, 2023.

[5] TCS4F. Pledge for sustainable research in theoretical computer science, 2025. `https://tcs4f.org/`.