Hi, I'm your host, Sarah Webb. And this is Science in Parallel, a podcast about people and projects in computational science. In this third episode in our series on creativity in computing, I'm talking with Danilo Pérez, a Ph.D. student in computational neuroscience at New York University. He is currently supported by a Department of Energy Computational Science Graduate Fellowship. At NYU Danilo is studying value-based decision making, examining how both animals and humans weigh complex choices based on how much each option appears to be worth. He works with Christine Constantinople and Cristina Savin and is developing computational tools to help understand this process, which can be applied in economics, medicine and public policy.

But this is just one project among many that Danilo has taken on during the COVID 19 pandemic--when he was just 24 years old--he maintained Puerto Rico's COVID-19 public health data dashboard. For his CSGF practicum in 2022, he worked with an Argonne National Laboratory team led by Arvind Ramanathan on an artificial intelligence model that simulated the evolution of novel coronavirus variants. That team was awarded last year's Gordon Bell Special Prize for High-Performance Computing-Based COVID-19 Research. We discuss how compelling science problems have propelled his training, how music and family support him and his long-term goal to support citizen-facing science, especially in Puerto Rico.

Welcome, Danilo. It is great to have you on the podcast.

Hi, Sarah, it's a pleasure to be here with you and your listeners.

Before we start talking about value-based decision making. I want to get some background on how you got your start in science and where that came from for you because you've worked on a lot of really different, really interesting things.

So after like, being born in the Bronx, when I turned eight years old, I moved out to Puerto Rico. I grew up in a small town, Coamo, where my main passion when I went to school was really math. Science, I was actually pretty, pretty poor at it. And it wasn't until I had the privilege of going to a boarding school for my last year of high school: CROEM, Center for Education Opportunities of Mayagüez that I really, really got my hands on what science really meant and started to understand that science was a wider enterprise that I could see the rest of my life developing within.
Sarah Webb 02:39
And how did you get involved in research?

Danilo Pérez 02:42
So after graduating, I received an offer from the RISE program at UPR-Cayey. This is a great program. I know UPR-Cayey continues to have a RISE program that is led by Dr. Juan Santana. It has a strong track record of getting undergraduate students to pursue research opportunities in parallel with their undergraduate studies, particularly sending us on summer internship programs. So despite the fact that may be the exact research questions, I perhaps was developing an interest to weren't present on my university grounds right at my campus, RISE would actively pursue liaisons beyond our university system and send us out for different summer programs in collaboration with REUs, etc. My first summer I spent it over at the University of Missouri-Columbia, pursuing a project under Dr. Zezong Gu on traumatic brain injuries. So at that point in time, I was also amongst the group of undergraduates that really wanted to pursue medical school. And I was still in the process of defining my interest. So that research project had a strong clinical component, even though it was sponsored by computational neuroscience. I'm definitely closer to biology than many of our DOE fellows, but I am also very well defined on the opposite end of clinical research in the realm of neuroscience.

Sarah Webb 04:07
So was that the first computational project you worked on then? Was that the root of your interest in computing?

Danilo Pérez 04:14
No, so I had the opportunity-- and perhaps even the fateful blessing-- of arriving at the same time as the first biomathematician at the department of mathematics at the UPR-Cayey had recruited which is Dr. Maytéé Cruz. And upon my return from Missouri, I definitely was clear that neuroscience was something I was interested in, but I did want something that was more rigorous in mathematical detail. So I went straight to the department of mathematics. I spoke with my professors who had given me calculus the year before, and they mentioned that they had just recruited Maytéé, and she was interested in starting up her biomath lab. From there I ended up taking pretty much any course that she offered until I graduated, but I was also one of our first research assistants. There we pursued projects in epidemiology and my particular interest during my time at the university, diabetes, so developing a mathematical model of the interchange of insulin and glucose. And that's where I really got my hands on mathematical modeling. On the side, we pursued a bit of epidemiology that ended up coming a bit handy later on, even though I had no idea that we were heading into the pandemic.

Sarah Webb 05:21
Okay, you basically set me up to talk about COVID. And you will have a very interesting story and a unique perspective on COVID. Tell me about your work during the pandemic in Puerto Rico.

Danilo Pérez 05:33
Yeah, so Puerto Rico was kind of in the aftershocks of a series of data scandals in general, particularly after Hurricane Maria, people became very wary of the management of official statistics in Puerto Rico. So when the Puerto Rican government decided to be as transparent as they could be. That being recognizing the lack of infrastructure that was ready and in place, when the pandemic began, that they would be as transparent as possible. At the beginning that meant the data was really hard to track down. So especially if you're not a person that's familiar with this sort of data and how it's adequately presented, their social media was full with different graphs that offered different interpretations. And so I recognized, with a group of peers that would eventually coalesce as the Coalition for Community Health, COSACO in Spanish, that we could aggregate these different data sources, and present them as well as possible. This led to actively pursuing different relationships with the Department of Health
that helped eventually form the major contact tracing team, known as SMIRC and I was part of the team that put together the data and analytics component for that contact tracing program.

Sarah Webb 06:45
I was reading that there were, you know, a whole range of different people involved with different expertise, that there were artists involved. Can you talk about the people who came together as part of that process to do this?

Danilo Pérez 06:59
As a neuroscientist and first and foremost, a neuroscientist that speaks Spanish, right. And that's, that's my first language. I can't help but think of the importance of art with our science, right? If it weren't for Ramón y Cajal, if it weren't for those beautiful and mesmerizing drawings that helped coalesce what we understand nowadays, as that discrete theory of a neuron that was still so controversial for his era, then we wouldn't have that transparent understanding.

Danilo Pérez 07:23
So when we arrived at our first meetings, I showed some very beautiful graphs, from my perspective, the perspective of a computational neuroscientist that has extensive experience in mathematical modeling of epidemiological phenomena. And the lead of the program, the director, Dr. Fabiola Cruz López looked at me, she smiled and she laughed. She understood nothing. And if she understood nothing with a Ph.D. in microbiology and initiating medical studies, then there was really no use of this graph. But we really do need to bring people with diverging expertise to really help communicate these ideas that we have. And that's why we brought in a team of not just artists, but educators, they effectively became educators as they transform these very complex ideas into messages that could be effectively communicated both verbally but more importantly, visually, as we live in an age of social media. It was that visual presence that almost became synonymous with the contact tracing group.

Sarah Webb 08:29
At the time that Danilo was working on COVID data in Puerto Rico, he'd also applied to graduate school at NYU and moved to New York in early 2021 to start his Ph.D. I asked about how he managed this important public health work while pursuing his own scientific training.

Danilo Pérez 08:47
I mean, thankfully, I switched off gears after initiating the program. And we found ways where I could actively continue my academic preparation in line with different research interests we could have with the datasets that were being generated. For example, in my machine learning course, I had a final project that I had to contribute, and we wrote it up and I obtained permission directly, both from the Secretary of Health in that moment and my director Fabiola to understand the relationship between preexisting conditions and outcomes, be them hospitalization and death. And this was prior to the release of different treatment options, particularly the monoclonal treatment. And what we ended up noting is that conditions like hypertension and COPD were particular preexisting conditions that statistically predispose people even to a greater extent, right, than had been to that moment known. This is now, pretty much in hindsight, known pretty well across the scientific community, and there's been extensive papers on this. But being able to have the volume of data that that contact tracing program was able to build that was so detailed in terms of symptoms, and the preexisting conditions of the distinct patients really set us up to do great things in record time. Like I was mentioning, we went from having essentially zero infrastructure and data that really couldn't be trusted as well to eventually have a sophisticated program that could be malleable to the most sophisticated algorithms to discover patterns in time. And before they were widely known in record time, and run that all back right to the artists, to the messengers, to the people on the ground to really build up that trust that is so important for the practice of public health.
That's so cool. A theme of this coming out of this is that you managed to take something that was a process of serving your community, and it was also an educational process for you, too. There's some wonderful synergy that I think is really unusual and really interesting.

Danilo Pérez 10:44
I think it was also a sign of the times. I think as we claw back and try to normalize operations in the quote, unquote, post-pandemic world that we navigate, we're becoming less and less interested in taking those risks, and we want to come back to the standard playbook. But I really want to thank, right, not just the people at the Department of Health, but even the people at NYU, right, my professors during that course that said, huh, that seems pretty rigorous. I don't know if that's appropriate for a class project. But it was the sign of the times, right.

Sarah Webb 11:14
Danilo's public health work in Puerto Rico influenced how he approached his CSGF practicum research with Arvind Ramanathan at Argonne National Laboratory, a project that applied large language models, code similar to the underpinnings of ChatGPT and other tools, to tackling genetic information and predicting how coronaviruses evolve.

Danilo Pérez 11:37
As I mentioned, the director of the contact-tracing program was a Ph.D. researcher in microbiology and her previous work was on dengue virus, particularly their genotyping and characterizing those genome sequences. So prior to my arrival at NYU we had already fantasized like, well, if we're building this infrastructure that now is doing what many would call perhaps social epidemiology or community level epidemiology. What would it look like if we could develop infrastructure for genomic epidemiology? And, lo and behold, right, they built an entire international genomic database that spanned the Earth for COVID, something that was progressing very slowly, with a couple 100,000 genomes of influenza viruses. But the dengue virus, my director had collected perhaps 100, we now had millions of.

Danilo Pérez 12:25
So over at the practicum, there was, in fact, Arvind Ramanathan, thinking about these ideas, taking advantage of other advances that are occurring in the field of artificial intelligence, and the development of models, like transformers models, and all of these things, again, started to coalesce-- and, as a sign of the times, drawing in these kinds of disparate ideas, started to push us in a particular direction. And that's where right the Ramanathan group came up with this idea of modeling the evolution of genomic sequences, kind of how we model large languages right, and use it as sort of a generative model that could project predictions of sequences that we have not yet observed. And not just generate those sequences but also take advantage of other algorithmic advances that would translate those sequences and evaluate the fitness of the proteins, or, and the particularly functionally defining proteins that would give us variants, right.

Danilo Pérez 13:31
So this term that we're now very familiar with: the variants of a particular disease. We've heard of the alpha variant, the omicron variant that came pretty much to stay with all of us. So these distinct variants were characterized by mutations in their sequence that conferred proteins, particularly focused on the spike protein that was on the surface of that COVID molecule. And by generating new sequences and in evaluating fitnesses, we could potentially, and eventually, confirm that using our model. So this kind of sophisticated idea kind of followed me all the way from that idea that we could start grouping together different data sources, kind of build up a small infrastructure in Puerto Rico, and eventually have the opportunity of seeing what having that infrastructure and having the compute resources could offer us. And it was truly a magical time for me. In the pandemic, you mentioned the prospect of how
long days were. Well, they were very long. But you know, the days are long, but the years are short. At the end of the day it was only three years ago. It feels like a lifetime ago. So I would say that it's definitely not a time I hope to repeat in many aspects. I think there was much tragedy that could have been avoided, but it was a time of great learning. And for those of us that had the opportunity to serve others, we also hold that with a bit of gratitude.

Sarah Webb 15:03
We would also be remiss in not mentioning that that work was recognized by the computing community with the, you know, the Gordon Bell Special Prize for COVID-19. It's kind of mindboggling to think, on the computing side, that we now have algorithms that are powerful enough to really take a wealth of data and start to think of how a virus evolves. I'd like you to reflect on that a little bit. What are your takeaways from the computing side?

Danilo Pérez 15:30
That happened just before, right, this popular revolution of ChatGPT and then the other AI models that came in. I, perhaps, didn't have a full appreciation prior to what that would mean to all of us. I thought that as we used it at the national labs, which was with perhaps the level of care ethics and expectation that others would have expected us, right. Scientific rigor: We tested the particular hypotheses we were interested in. It seemed to me that it was very much something that was to be used in a controlled environment on supercomputers. What has marveled me even further is how rapidly the public has adopted these tools and has wanted to play with them themselves. And that's provoked other questions that I think are going to be relevant to all of us.

Sarah Webb 16:17
What sorts of questions do you think we need to be thinking about and addressing?

Danilo Pérez 16:20
So there's this huge issue, right, of hallucinations. Plenty of people have already pointed out that this doesn't work, as perhaps most people think it does. Some people might think that they ask a question, and they're receiving a discrete answer that exists. And essentially, that's not what's happening. It produces the answer that is most likely. So I had a great example of this that I developed. I asked the model a simple question, what's the name of the Pokémon that rhymes with cobra? And to my knowledge, I don't think it exists. But there was a particular Pokémon that's well associated with the word cobra. It's Arbok. Right? It's cobra spelled backwards. And the model would insist, ah, yes, the poker man whose name rhymes with cobra is Arbok. And I'm like, No, that doesn't rhyme. Oh, I'm sorry. It's Ekans. I'm like, No, that does not rhyme. And right, yeah, would insist that it can find a solution. Right? Why? Because my original question kind of assumed implicitly, that there was Pokémon that in fact, its name existed with Cobra. So I tried to ask ChatGPT. And to my surprise, it just kept hallucinating. And there was no way I could engineer this prompt for it to give me a satisfactory answer. So even this idea, right of the fact that we have the engineer prompts, and how we're interacting with these models, becomes necessary to consider, right. Is what we're doing at the level of complexity to compare it with engineering, when we execute a prompt? I think that's a question that that's going to be decided soon. It was truly, truly a revolutionary moment, I believe it to be a forward-facing enterprise that will move us forward. But that does sometimes involve taking three steps forward, but two steps back.

Sarah Webb 18:19
Let's go back to computational neuroscience and decision making, I want to come back to the things you've been working on most recently, and the things you've been studying and where you are with that work.
So I think the fact that we were able to spend so much time outside is reflective of how I approach the work that I'm doing inside NYU. And it's that we're developing the computational ability to characterize complex problems. And the complex problem we're interested in here is value-based decision making. It's how particular areas of the brain may perceive and compute and compare value during task execution in ways that explains the behavior that we're observing. So we have a particular task paradigm that allows us not just to evaluate the ability, or the valuation of a particular reward offer, on face value, but also within its context.

And I'd like to right, peel that onion back. So if I were to offer you $10, $10 has a discrete unitary value, it's $10. However, if you're particularly broke, those $10 come to represent something much more significant than if you had $10,000 in your back pocket at that point in time. So context is always important when we're assigning the actual value that we perceive, and economists have spent plenty of time in this discussion. I just want to underline that that's the concept we're dealing with. It's the idea that objects or offers can have a distinct value depending on the context in which they're presented. One particular area of the brain that's been associated with right value estimation and decision making has been the orbital frontal cortex, right? It's one of these sub-areas of our frontal region that many people hold on to when they're really thinking hard about a problem. And I have no idea if that kind of psychophysical interaction that we have, with the front of our brain really has anything to do with with us thinking. But that's kind of the area in which we're operating in where we record from in rodents.

So we have this paradigm that's able not just to communicate a discrete value to rodents, but by studying the time that a rodent is willing to wait for each value, we start to develop an understanding of the value within the context. And to continue to peel that onion, if I were to tell you that the animal has to wait for the offer that they're being presented. And that one offer was worth five microliters of water. So they're motivated by water in this task. And 80 microliters of water is the second ask. And 80 microliters of water is the second offer for which of these two, would you expect the animal to wait more? The distinction between 80 and five is clear as patent. And effectively, in the task paradigm that we have, the animals will wait more for 80 than for five. However, we're able to manipulate the context. So that medium offers like, say 20 microliters. Sometimes they'll wait more for it, or sometimes they'll wait less for it. And that's dependent on the history of rewards that they've received. So if we were to prime this animal, by historically giving them 5, 10, 5, 10, and then magically 20, appeared, when out that 20 is perceived as the most locally best option they've ever received. And therefore they're motivated to wait even more for it. Opposite. If we were in a context, we're at 80, 40, 100., and was that trial history and a 20 appeared in the animal would be less incentivized they would wait less for it.

I assume at that point that the mouse is going, "What? You're only giving me 20?!!"

Exactly right. If we were to anthropomorphize in that way, we'd find that there is a contextual adaptation at the behavioral level just by waiting, just by measuring how long these rodents are willing to wait for these distinct offers, we could identify neurons that are consistent with reports in the literature where the discrete value, so 5, 10, 20, 40, 80, these would scale the neural response of particular neurons. But we're also able to identify neurons where that take 20 microliter offer, we would separate into blocks where they were receiving low offers or high offers. And the neurons would also be responsive to those changes. We're taking advantage of those simultaneous recordings. And we're trying to build a more concrete representation of the process that really generates this neural activity. So as I mentioned, for this task, it only really seems to be context and offer: these two variables. So why would
I need hundreds of neurons to communicate those two variables that are being controlled? Essentially, what that does, right, is enforced two different scales of interaction. There's a fast latent that evolves at the same timescale as the observations, but also a slower component that are varied trial to trial to try. And it's on that trial to trial-to-trial variation, where the rodent, we expect to be updating their evaluation of the context, whilst within the trials is the actual management of, okay, this has been offered, this is how much I'm waiting, Okay, I'm going to opt out, or I'm going to wait for the reward. That's where the intermediate lead would come in.

Danilo Pérez 24:07
So with this idea, right in this hypothesis, that by modeling the data this way, we would obtain separable traces of these distinct processes, right, context and offer, we went ahead and applied it to the hundreds of sessions that we currently have available. The issue here is right, that's a very, very large amount of data. So thankfully, we have access to the high performance computing resources of Greene, the Greene cluster over at NYU, and that's been instrumental in helping us actually get these fits up and running and be able to observe and in fact that they're early traces, and we hope to continue to track down and perhaps optimize our model to evidence that the these neural components are in fact separable, at least in rodents that quote unquote, know what they're doing, that actually present this contextual modulation. Because on any given day, right These are rodents, sometimes they just do what they want, right? This is this is what biological research is about. You don't always get the result, the same result every day on each experiment. But as we build these databases, and we start to scale them up, taking advantage of these HPC resources really helps us underline and extract those sessions, where, where, in fact, it is occurring and underline them.

Sarah Webb 25:24
I think about all the things different things that you've worked on, I don't know in what way computing is the hammer or the toolbox, or whatever else, if we're extending that analogy that you've already put in there.

Danilo Pérez 25:34
I think that's a great question: Is computing the toolbox or the hammer? I guess I usually think of it as as the hammer, right, it's really what's going to leverage the work. But also, being a computational scientist, which is what this fellowship is about, isn't just access to supercomputing resources. It's great that we achieve that through this fellowship. And it's great that we pursue it, but it also entails so many other things about the developing a conceptualization of data curation of evaluating your data, of thinking from a perspective of scale, that computing itself ends up becoming a hammer in a very complex toolbox that a computational scientist actually has at their disposition. And I think the big superpower that any computational scientists could have is the willingness to collaborate because that's where they're really going to be able to come in and do all their work.

Sarah Webb 26:31
I want to ask you what you have fun with when you are not thinking about computational neuroscience, or scientific problems or computing and what those passions mean to you.

Danilo Pérez 26:43
Yeah, I have the privilege of pursuing my graduate studies, not necessarily at home in Puerto Rico, but with family in New York. I have an immediate family; I live with my family. And something that's really big in our in our household is music. And it always has been. So my uncle has a studio that he set up, a home studio. [Music plays in background.] And we just after a long day of each of us having our day jobs that we hold on to because that's what puts like the food on the table, we sit down and we listen to what's going on. And sometimes it's just you know, the news cycle and what these artists are doing on their Instagram. But nowadays, it is increasingly about music, there's an under appreciation that there's
just so much access to different sounds that every artist is putting us out towards. So we love to listen to what other artists are creating, that's first and foremost.

Danilo Pérez 27:36
Then we like to sit down and improvise and see what we can bring together and what isn't out there. What do we miss? That's a big thing. With how rapidly the music cycle moves nowadays, and how radio stations will give the next track and you only get a week of play, just because of the level of competition. And if not, people have streaming services right there and they curate their own playlists. We look for the music that we want to hear. Right. And that's what we want to create every time we stepp in that studio is music that perhaps isn't in the public eye anymore. We make a lot of particular reggaeton and rap that perhaps isn't as pop in this moment. But it's what we like to listen to. And it allows us to bring in context everything: the updates, what New York City feels like, what's missing, our country is like. I particularly for Puerto Rico him for the Dominican Republic. But it both tied back to the Caribbean very strongly. So I think that's what I look for in music. It's kind of a yearning to what's in the past.

Danilo Pérez 27:37
This song, "Tu Dices" by Danilo and his uncle Yendi Espinal, who performs as Yei Flow, chronicles both love and heartbreak. It also comes back to neuroscience, in particular, Danilo pointed me to the following clip.

Sarah Webb 28:43
[Louder clip from Tu Dices, 13 seconds.]

Sarah Webb 29:11
Roughly translated from Spanish, it says, I look at you and see how much you lie. You say you have a blast, but you don't seem so fine. Even when you're with someone else I come to mind, because the memory stuck in your neurons is always mine.

Sarah Webb 29:27
In what ways does that time feed you as a, you know, as a scientist? How does it feed you as a person?

Danilo Pérez 29:35
It offers me a support system in which to contextualize everything that I do. It's to understand that today we're, we're, we're working and we're focusing on this point in time on the work we have to do, because later on, we're going to be able to decompress appropriately. There isn't just dinner and going back to sleep after this. There's fun. There's smiles. There's happiness. And a happy scientist as a good scientist. Don't let anybody convince you the contrary.

Sarah Webb 30:02
We've been talking about creativity over the course of this conversation in many, many different ways: scientifically, you talked about it musically. And we've talked about interdisciplinary science just as an institution. What does creativity mean to you, Danilo?

Danilo Pérez 30:20
Versatility. I think that's what it is, I think, definitely there is a specialization that we come with. There's a badge that we do want to put on. In my case, it's neuroscience; it's computational science. But being creative also means being flexible. It's not over specializing, it's not building out that great muscle. You have to also be able to work with it. And when I think of creativity, it comes straight back to the idea of versatility and amenability to the circumstances and problems that you're posed.
**Sarah Webb 30:50**
One of the things that has been a theme for you and continues to be is citizen-facing science. And I’d like for you to talk a little bit about that, and what that means to you, and how you see that shaping your future as a computational scientist.

**Danilo Pérez 31:06**
Yeah, I mean, especially as a computational scientist, there’s a temptation to engage in black box science withholding or relying on knowledge that you really don’t understand. That doesn't really end up being a productive enterprise. I think citizen-facing science takes that a step further. It says not only right, as a scientist, am I undertaking the responsibility of using tools with a full appreciation and understanding of what I'm doing. But it's also ensuring that the public is well-informed and can play its role in that process. And I think as right, the current scientific infrastructure and foundation is based on public trust, and the funds that the public understands are necessary to invest in our enterprise. We need to continue to build and reinforce that trust. We've seen glimpses of how dangerous it could be, to lose it, for whatever reason, be it political, be it for any other reason. The fact of the matter is that building that trust, and ensuring and protecting it is a project we all have to keep on our shoulders as scientists and as members of this community. So I take it on as great responsibility. But I also take it on, as fun, right? I think that getting us out of the lab and reminding us that our work is human. Being a better scientist, does it make me a better human? Sometimes I don't know, I've had this discussion with other students. But fundamentally, I do think that just being a better human being will leak over and make us better scientists, though. Getting us out and doing our science in the public and making sure the public has an appreciation for our work to me is fundamental.

**Sarah Webb 33:01**
And if you could pass along a piece of advice or to other researchers, what are some kernels of advice that you'd like to pass along and ask for it?

**Danilo Pérez 33:11**
As a as a young and upcoming researcher, I can be sometimes overlooked by other researchers to ask for advice. But people that come to me and asked me for advice, I always try to tell them to be true to who they are, right? It can be tempting to join the most successful project that's going on, or the most successful lab or just chase what seems to be trending in your particular field. But at the end of the day, your work, you're gonna have to do it. So let that guide you. Let that really resonate with you as a scientist at any particular time. Because no one wants a boring job. And the science is such an exciting enterprise. And there's so many great things happening every day that the field isn't going to suffer because you decided to go in a direction that is better fit for you. If we explore this wonderful world, this wonderful variety of options that science has to offer us, we can find a little corner where we're more than happy to continue to chip away.

**Sarah Webb 34:15**
Danilo. Thank you. It's been a pleasure talking with you.

**Danilo Pérez 34:18**
As well, as well. Take care.

**Sarah Webb 34:21**
To learn more about Danilo Perez, his work on value-based decision making and COVID-19, and the music he creates with his uncle, please check out our show notes at scienceinparallel.org. Science in Parallel is produced by the Krell Institute and is a media project of the Department of Energy Computational Science Graduate Fellowship program. Any opinions expressed are those of the speaker and not those of their employers, the Krell Institute, or the U.S. Department of Energy. Our
theme music is by Steve O'Reilly. This episode was written, produced and edited by me Sarah Webb. We'll play out the episode with more of "Tu Dices" by Danilo and Yei Flow.

[Music in Spanish]