Recall This Book

47 Gael McGill, Glimpsing COVID

John Plotz:

When scientists peer into the minutest crevices of the human body, building hypotheses, and vaccines, based on what they learn about our cells, our membranes, our folded proteins--how the hell do they do that? If you're like me, anything smaller than the head of a pin has always filled you with a mixture of awe and unease. Yet, every day theories are made and tested, inferences are drawn about the functioning and the disease of the animal body each of us inhabit.

So, what exactly is the history of our microscopic data visualization? And what are the latest developments that crack data scientists have unveiled in 2020 as the deadliest time bomb for a century ticks away? Well, let's find out.

From Brandeis University, welcome to Recall This Book, where we assemble scholars and writers from different disciplines to make sense of contemporary issues, problems, and events. I'm John Plotz, and my Brandeis co-host today is the esteemed neuroscientist, expert on synaptic scaling, and lately on sleep, Gina Turrigiano. RTB long haulers will recall her from our episode on addiction and from a scintillating interview with the author of *Circe*, Madeline Miller. Hey Gina.

Gina Turrigiano:

Hey, good to be here.

JP:

It's great to virtually be here with you. And our guest today is that promised crack data scientist, Gael McGill, who is the director of molecular visualization at the Center for Molecular and Cellular Dynamics at Harvard medical school. As a leading data visualizer (and I'm sorry Gael if that invented title doesn't please you, but that's how I think of you) as a leading data visualizer, he's the founder and CEO of *Digizyme*. He's also along the way the coauthor with EO Wilson of the amazing *Life on Earth* iBook. So Gael, welcome to our virtual Recall This Book studio.

Gael McGill:

Thanks, John and Gina, I'm very excited to be with you here today.

JP:

Great. Well, I'm incredibly excited to have you. Can you maybe begin, Gael, just by telling us about your current data visualization projects?

GMG:

Visualization is more than just resulting in beautiful images, memorable images that inform, it's also for me fundamentally a knowledge synthesis process. And it forces you to think about your data, even if you're the world expert on a particular dataset. I've come across many situations where in collaboration with such specialists, the process of creating these visualizations will shed new light on otherwise familiar data. So I just wanted to say that because for me, the visualizations live kind of on a continuum from explanatory to more exploratory as it relates to biological data.

So, what have we been up to? I think the first place I have to start is to maybe share a bit of our efforts with the relevant and current question of the Pandemic and the virus that's causing it, which is the SARS-CoV2 virus. It's been particularly exciting and challenging because the rate at which the scientific community has rallied and published materials over the last eight months is just this breakneck pace. I mean, every new issue of *Science, Nature, Cell,* there's something new--and it's relevant, and it's interesting. So just as this started we wanted to try and see what we could contribute, in terms of our visualization efforts, to our understanding of the virus. And both in terms of the general public, but also even within the scientific community.

So we've worked hard over the last eight months to not only keep up with, especially the information we have about the structure of the virus--and in particular, the spike protein on the virus. It's that protein which sticks out and actually gives the coronavirus family its iconic name, but it's also the protein that's responsible for initially making contact with our host cells through a receptor. And after that interaction happens, that same spike protein is a little bit the Trojan horse--once that contact happens, it will drive the virus to fuse its membrane with the host cell and deliver its genetic payload into the cell.

Without going into too many of the details, that process which really kicks off infection is a critical time in our fight to inhibit infection. And so we wanted to see if we could use all of our knowledge, but also our software development,

and I can go into that in more detail, to try and visualize that process as a continuous process. As opposed to the isolated snapshot pictures that you see very often in the media, and even in the scientific journals.

GT:

You use data from many, many sources to construct these models and put them into motion, right? But the dynamics are kind of an inference from a bunch of static moments in time that you can get with all of these approaches for actually visualizing the structure. So, I think this kind of data visualization really is model building also. And you're having to actually make inferences between points in time as to what's really going on. Right?

GMG:

Exactly. And that's a really good point. Because the model building, as you said, is not just about bringing together pieces of structure that different labs have solved in different ways. But also, once you have those models, and let's say that you have your favorite protein that exists in an on-state model and an off-state model, and we know that it transitions between them—but how exactly does that look like? How do you make that inference? I mean, I would say that it's not completely blind, so there are a different set of data, of course, that we use to guide those types of transitions.

JP:

So is the model you're envisioning then capable of sustaining multiple hypotheses to explain the data that are recorded within it?

GMG:

Yes. The work, in fact, specifically that we're doing on the SARS-CoV2 spike, it's actually not just a single linear animation. And by linear, I mean like a single narrative where we just take the viewer by the hand and say, *this is how it works.*

What we're creating is an interactive visualization. Where at different steps along the way we will stop the visualization, or give the viewer the ability to interact with it and say, *Okay, right here there's actually three different models of how this might work.* And we don't want to skip over that. The whole point of the visualization is actually to visualize alternating or even competing hypotheses. And also to improve the visualizations in a way that gives the viewer a sense for the quality of the data too, right?

So not just the different methods, but some methods give you data that is more coarse-grained and messy and dirty, and other datasets are much higher quality. Many current visualizations don't do anything to tell the viewer, not only where the data comes from, but even the quality of the data and how it's being used. And I think that's what we're trying to do.

GT:

I'm thinking a little bit about, or trying to wrap my head around the similarities and differences between different kinds of scientific modeling. So, neuroscientists use modeling at lots and lots of levels of detail to try to come up with tests of whether their data have really adequately explained a phenomenon. And also to be able to perturb it and look for predictions that they might then go test with experiments. And so, pretty much most branches of science do that.

How do you see this sort of visualization? What's the visual element to it? How does that change the way you think about these models? Their efficacy and their, maybe ability to lead you astray sometimes…versus other things. I mean, we visualize outputs of our models, but it's not the same thing as this. Which of course, the listener can't see, but maybe… Are we going to post a link to this?

JP:

Yeah, we're going to definitely post links. Yeah. We encourage people to come to the site and watch Gael's models at work. For sure. Yeah.

GMG:

Yeah. I'm thinking of a couple of different parts to the answer. I think that the first thing I think about in response to your question, Gina, is that visual representations of data, as you said--and even when we hit the history part of it perhaps, John--We'll see that it's everything from the earliest sketches of Leonardo, to Galileo's paintings of the moons, to everything up to now. And so there's incredibly rich history of that and the role that it's had in scientific discovery. All the way through the DNA, the helix model by Francis and Crick And all of those things. The-

JP:

Franklin and Francis and Crick.

GMG:

I'm sorry, that's correct. That's correct. Yeah, the problem is what I visualize in my head is the *Nature* paper, that single-

JP:

Fair enough.

GMG:

Thank you for catching me on that. Absolutely. Couldn't have done it without the actual x-ray data. But to answer your question, the part where I think visualization, at least to me, so this is a personal answer, that I'm most excited about is to think of the use of visualization. And literally the visual output of the data, as you're asking me, in realms where the data or the representations are outside the realm of the human senses and of human intuition. Right?

So, if you're telling me about, let's take a synapse, right? We can write about it, we can talk about it, we can describe all the elements. But ultimately for someone who is not an expert in it, I would argue that there's incredible power in trying to create, as accurate as possible, a visual representation of that structure informed by everything we have. From microscopy on one end, all the way down to the calcium ions on the other, and everything in between. So that's the idea of spatial elements that are outside the realm of human experience. And incidentally, the laws that go along with it that go against our intuition.

So what I love to tell my students all the time, is this notion that gravity just doesn't matter at this scale. It's not that it doesn't exist, but it's just if you're looking at the forces of how, let's say one of those calcium ions is traveling around the synapse, it's like the environment is molasses to that calcium ion. As opposed to what we envision a little ball might be floating around a swimming pool.

So there's spatial scale, and the unique laws that happen at different scales and the way we can use those and visualize them, but there's also temporal scale. And whether it's, again, the femtoseconds of molecular vibrations and side-chain rotations, and all those things, all the way to the other end. Which is, you're faced with the same problem with students if you're describing geological timescales, or evolutionary timescales.

JP:

Can I just jump in here? I mean, Gina, I kind of want you to say more about your understanding of the distinction between visual modeling versus other kinds of modeling. Because I heard you drawing a distinction there, and I'm just not enough of a scientist to know what the modeling alternatives are that you see producing different qualitative... Like either conceptual pluses or different kinds of deception in visual as opposed to other sorts of modeling.

GT:

Yeah. Actually, I was just really interested in hearing what Gael had to say about it, because I think it's a continuum. It's not two kinds of models. That every kind of modeling that we do there has to be some visualization element to it. I think when you're talking about structures, that becomes the main thing that you're looking at, right? Where something like patterns of activity in neural networks, the structure doesn't matter, right? What you're thinking about are correlations, or some higher order thing that you then have to figure out how to visualize. So what it is you choose to visualize is actually a statement about what you think is important.

GMG:

That's a really powerful statement. Because even within visual modeling, which again it's a continuum, it's not just a little category. But the endless decisions we make literally on a daily basis is exactly what you just said, that there is no such thing as a visualization or visual model without considering at least two things. The target audience, *who's going to be looking at this*? It's a totally different thing if it's a middle school student versus a professor down the hall.

And also, what are your objectives in what you're depicting? It's kind of like, in some ways good teaching, which is knowing what to leave out is just as important as what you're communicating. So that applies to models. And maybe to frame the whole, at least my attitude towards models, which maybe I should have said earlier. I love the phrase, "all models are wrong, some are useful." That's very much the approach. It's not, let me put together everything I know and make the best possible visualization for you that's going to make... What's almost more important than the model itself is the discussion around the model, and the conversations that are sparked by the model. Whether it be correct or not, it's understanding that the model is a means by which to externalize, typically, a mental model in some cases.

But I think to go back to what I didn't hit yet in what I wanted to answer to your earlier question, Gina, is there are things that I think closer to what I would think of as quantitative models, or maybe what we see in areas like systems biology.

Just to take an example. So let's say you're trying to model the behavior of a E. coli cell, or any cell. And what that might mean, that word model, to a systems biologist is more about...Imagine you move the cell from 100 millimolar salt to 150 milli... In other words, change its environment. Can you predict, quantitatively, and over time, what's going to happen? What genes get turned on and off? Is it going to start moving? Is it going to die? And those types of predictions, those types of questions, don't necessarily have to have a visual output to be useful answers to that kind of modeling question. So I don't know if that helps address what you were asking, because I started answering more in the realm of, again, the part that we tend to focus on. Which is, can we use visual depictions to help scientists and non-scientists grasp things that are inherently difficult because they're outside of natural human intuition? Because of issues of spatial or temporal scale.

JP:

Gael, can I just pick up an implication in Gina's question though? Which is sort of related to this phrase that was banging around in my head from Wittgenstein, "a picture held us captive." So Gina, I think you were asking, and I know you were thinking before about moments when the visualization was the problem rather than solution. In other words, moments when we're trapped in a prior model and that ossifies and actually leads us astray. Thoughts about that?

GT:

Yeah. Well, I mean, I was actually trying to think of a really good example from neuroscience for this. And I'm sure they're there, but I don't have it in my head. But I guess in a general way, we do get conceptually trapped by our models all the time. I'm using *model* very loosely here, not as a beautiful data-driven visualization, but just our kind of stick-and-ball conceptualization of causal interactions between things, right? So scientists are always drawing out these linear pathways and trying to come up with causal explanations. And-

JP:

Wasn't there something called the "plum-pudding model" of the atom? I feel like I remember that--in which the negatives were suspended. There was a proton, there's a positively charged object that was with chunks of negative in it. And people were committed to the plum-pudding model for a really long time before they realized-

GT:

I think you made that up, John.

JP:

No. I didn't make up the phrase, I love plum pudding.

GMG:

Sounds delicious, but I haven't heard that one before.

JP:

I mean, that goes along with your stick and ball point, Gina. You're just saying that diagramicity is inevitable. But then sometimes, what do we do when we get the wrong kind of diagram? And it just stops us seeing what's actually going on.

GT:

Yeah, but I also wonder, I guess where I was thinking about this, as you start to build something as beautiful and visually detailed as the visualizations of the data you have for viral fusion, are there places where that beautiful detail lulls you into a sense that you understand a process, and you've really got it wrong?

JP:

I really want to find just a moment to talk about the history stuff before we pivot back to COVID. And so can I do it this way? Maybe, can I ask each of you to just talk about, if it comes to mind, a historical model or a historical visualization that was an important past moment? I mean, I heard you mention Leonardo before, Gael. But maybe since Leonardo, if there's a nearer term instance of a visualization that performed some of the transformational work that you're describing the visualization being able to do.

GMG:

Yeah. Should I go first?

JP:

Or Gina, whichever one of you guys has one on the tip of your tongue.

GT:

I guess I could think of two things. One is going back into the beginnings of neuroscience, so something every neuroscientist would understand. And that's the drawings of Ramon y Cajal.

JP:

Yes. I was hoping you would mention him....

GT:

What's astonishing about these, first of all, their beauty. They're absolutely gorgeous.

The man was just a consummate artist. But more astonishing was the inferences he could draw from these images that he created. And just example after example of extracting potential principles from these images that could then be... Then people have been testing for the last century. And many of them, some were wrong, but many of them turned out to really be remarkably in line with what people have found about the way information flows in circuits and things like that. I think that was an incredible example where those images really drove a lot of research that—people wouldn't have thought to ask the questions they were asking if they hadn't actually seen those images.

JP:
Yeah.

GMG:

Can I ask you a quick follow-up question? Almost to give you back some of the types of questions you were giving me earlier. You mentioned the beauty, which is undeniable. Do you feel that the beauty helped or impinged? Or, what is the role of beauty in the example you're giving us?

GT:

I love that question. I think it undeniably helps. I mean, you look at these images and you want to understand the structure, right? So the beauty does somehow bring you into... Yeah, it-

GMG:

It's an engagement.

GT:

Yeah. I mean, why? I don't know.

GMG:

Well, but-

GT:

I think in time we can see something that we have predicted, and it turns out to be a beautiful image. It just enhances the sense of... I'm not quite sure how to phrase this. The importance, in some way. I mean, I don't know that there's intellectual validity to that, but in terms of bringing humans to engage with the problem, it's undeniably powerful.

GMG:

Well, and we might be having a different conversation if the beauty had turned out to be associated with completely wrong... In other words, I think we're talking about it the way we are, because it turns out that he did have the impact, and it was real. It was beauty in the service of drawing engagement towards something that turned out to be very powerful and scientifically correct.

Whereas, if that hadn't been the case, I think we'd be having a conversation about kind of where we started. Which is, let's be aware of how visualization can mislead. But I think there is this... And in another grant that we're having at the moment, we're interviewing teachers to try and better understand how they use visualization in their classroom.

And what we find is that probably the most common answer has more to do with the engagement factor rather than the more pure and clinical, instructional mechanism that visualization can bring. I like to think of it almost as they're buying attention credits. You're going to start a lecture on ribosomes and translation. But show them an animation at the beginning of

class and everyone wakes up and is, wow, that was so cool. And you've bought yourself 10, 15 minutes of attention, maybe. Something like that.

JP:

Right. If this were a conventional episode of Recall This Book, we would be allowed to name a recallable book. And I think I will use my recallable book credit and say, there's an amazing book called *Science in the Marketplace*, which is a 19th century book about 19th century science and how people used, for example, magic lanterns. People would perform experiments between two plates of glass and a magic lantern. And it's totally that thing you're describing, Gael. Which is that the engagement of an audience then becomes like an incentive for people to think further about what's going on. When I think it's hydrogen and oxygen are interacting between those two plates. So the act of being projected and being seen is actually part of the scientific experience itself.

GMG:

Yeah. I mean, I think we would probably agree that there's nothing wrong with the engagement factor, and we strive for aesthetically memorable output in our work. There's no question about it. For the reason that you just described, which is that we are in settings where we can't just assume that people are coming here because they're already 100% motivated to absorb and think about. So that has a role to play.

But I think we can't talk about any of that without acknowledging that what rides underneath it, which is our attempt at getting the science as right as we can possibly get it. But even more than that, again, what I tried to say at the beginning is that even with our very best attempts of taking all the data, all the viewpoints, pack it all in there. We still, I think, have a long ways to go in our field at thinking about and improving the design of our visualizations so that we do a better job of mapping data provenance, data quality, design decisions into the visualization itself.

Which is not how it comes across at the moment. At the moment, even the leading, most beautiful, scientifically accurate visualizations, they are very much handed over to you as the viewer in the mode of, this is it. We figured it out, here it is. It's beautiful. It's got a soundtrack, it's got sound design. You're going to be transported. And there's nothing about, *how did we build it?*

And another mode just to throw this in there. One thing I'm very interested in, and trying to figure out how to engage teachers to try this with me, is this notion that, what about engaging students into a conversation about dueling visualizations? Like even just having more than one would be, in my mind, a major improvement. So let's say we're doing the cell cycle today. Well, I'm not going to find the best animation on the cell cycle I can possibly find it, I'm going to find three as different as possible.

Some may be intentionally wrong. And I'm going to make that the assignment. And the learning moment is not, *can you remember what Cyclin A through D does?* But rather, *tell me what differences you noticed. What do you think it means?* And start there. Because the key for me with that, is that I would venture to say that leads to a learner, a student who becomes a better consumer of multimedia out there, where you have no control over what they're going to see. You can project things in the classroom, you can give assignments, but you know full well that the minute they go on YouTube or Google and do a search they're going to be exposed to all kinds of stuff. Are we training our students to know how to sort through that material? Is a separate question that needs to be part of...

And so I'm going to give a ten second really short answer to your question John. And I fear it's going to be boring, but it's very close to what I described before. Which is, in our field of molecular vis., I want to go back to David Goodsell. Who is this incredibly wonderful scientist, incredibly kind human being, brilliant human being. Who is trained as a structural biologist, but who kind of on the side ironically uses the least technologically advanced medium you can possibly think of—watercolor painting. So here we are talking about 3D Hollywood software, whatever.... David Goodsell does a mountain of work. It's kind of the iceberg below the tip of the iceberg. Which is, the amount of research that goes into understanding what you're trying to show. So he will go, and for a particular part of a cell, he'll figure out the entire parts list.

Like, what do we know about the hundreds of proteins that are there? Okay, fine, we've got that. What are their relative amounts? Okay. What is their relative geography in different states? So I won't go on, but he really... I mean, if there's anyone who does his homework, from what I've seen it's someone like David Goodsell. Then what to me was a unique moment that shifted people's thinking, and maybe not the experts who already knew this and their mind's eye. But the rest of us who maybe were not structural experts. He was

one of the first people to actually try to depict cross sections of cells in the full crowded complexity of what we know they are.

And it's one of those 20/20 hindsight things, where once you've seen one, you're changed forever and it becomes kind of obvious almost. But it's David Goodsell who brought that to the forefront. And speaking of beauty, I think part of the power of what made his stuff work is that it's gorgeous. It draws you in. And we've been endlessly inspired by his work, so that's one example. There are others I can think of, but I wanted to get you to at least one.

JP:

Actually, that relates to something that I learned when I visited one of these caves with prehistoric drawings. I visited a cave that had drawings on it of animals in flight, from about 22,000 years ago. And there was, the signage, if I understood the French, which I may not have, basically said that the multiple animals depicted (I think it might've been aurochs) it wasn't so much that they were depicting multiple animals, they were depicting the same animal as it looked in different stages as it ran.

GMG:

That is the first example of animation.

JP:

Yeah. Right.

GMG:

In the world. It's in Chauvet right, the Werner Herzog documentary?

JP:

Yeah. Well, this was a different place. But yeah, I think it's called Peche Merle. But in any case, it goes along with your point about the multiplicity of needing to visualize it at different stages in order to be able to understand it at, at one moment.

GMG:

And it's a problem with the vaccine development itself. Because it turns out that if you just take some spikes and inject it into bunnies and say, "Well we're going to raise an antibody." Well, what shape of the spike? One shape will induce one kind of epitope. So I think it leads to very serious issues of, what

immunogen are we using for our vaccines? If your vision, if your mental model of the spike is too simplistic, you're going to miss the boat potentially.

JP:

Gina, any last questions here, or last thoughts?

GT:

No, it's been really fun talking about all of this.

JP:

Totally. Well, my mind is blown. I want to go back and watch that movie, is it called *The Incredible Journey*, or *Fantastic Voyage*? The one where Raquel Welch is down the bloodstream. That's how I feel talking to you guys.

Okay, so I'm just going to say that Recall This Book was devised by John Plotz and Elizabeth Ferry, and is sponsored by the Brandeis Mandel Humanities Center and the School of Arts and Sciences. Our music comes from Eric Cheslow and Barbara Cassidy. Sound editing by Claire Ogden. Website design and social media, this semester comes from our newest Recall This Book graduate intern, Nai Kim of the English Department. We always want to hear from you with your comments, criticisms, or suggestions for future episodes.

You can email us directly or contact us via social media and our website. And if you enjoyed today's show, as well as sharing it with your friends, we would ask that you write a review or rate us on iTunes, Stitcher, or wherever you get your podcast. You may be interested in checking out our Books in Dark Times series, including conversations with the science-fiction novelist Kim Stanley Robinson, historian of science, Lorraine Daston, and poet Elizabeth Bradfield. So Gina, thanks so much for co-hosting with me. And Gael, thank you so much for this fabulous conversation.

GMG

Thank you.

JP:

And thank you all for listening.