Shane Greenstein:

I'm professor Shane Greenstein and you're listening to the Harvard Business School Digital Initiatives seminar, a premier seminar series that hosts thinkers and scholars who are pushing forward research on the digital transformation of the economy by conducting and connecting with cutting edge leaders, equipping leaders, and building community. The digital initiative seeks not just to study, but also to shape digital transformation. To learn more, check out digital.hbs.edu.

Shane Greenstein:

Okay. How are we doing? We're ready to get started. So it's a pleasure to have Simon DeDeo from Carnegie Mellon here to speak with us. As usual, we go around the room to introduce ourselves and our background and help our speaker appreciate-

Speaker 18:

I'm just going to double check it's working.

Shane Greenstein:

Oh, let's double check it's all working. Everything's working. Okay, and so he knows who we are and we know who we are as well, but nonetheless, we should do that. So I'm Shane Greenstein from the technology operations and management unit. Take it away.

Daniel Brown:

I'm Daniel Brown my first year here in the strategy unit.

Frank Gangland:

Frank gangland, I'm a professor in the strategy unit.

Josh S.:

Josh Schwartzstein, professor in the Negotiations, Organizations and Markets Unit.

Speaker 4:

[inaudible 00:01:26]

Speaker 5:

[inaudible 00:01:29] postdoc.

Diane Williams:

Diane Williams, I already introduced myself, a computer scientist and tech entrepreneurial alumni of MIT Sloan and Harvard.

Tommy:

Alright, Tommy [inaudible 00:01:40], third year student in the [inaudible 00:01:42] unit.

Becky Karp:

Hi, I'm Becky Karp. I'm a PhD student at BU in the Strategy and Innovation department.

Dave Homa:

I'm Dave Homa. I'm the Director of the Digital Initiative here at HBS.

Do Yoon Kim:

I'm Do Yoon Kim a student in the Strategy unit.

Cirrus Roofie:

Cirrus Roofie fourth year on the strategy unit.

Eric Mankins:

Eric Mankins, senior researcher in OB organization.

Mike Wells:

Mike Wells, third year Marketing student.

Speaker 14:

[inaudible 00:02:05], postdoc at HBS.

Simon DeDeo:

All right, very good. All right, well thank you for saying hello. I'm going to talk today... I come out of the quantitative sciences actually my PhD is in physics many, many years ago and I got bitten with the data science bug. I did animals and microbes for a while, but humans are a lot easier to study because you all keep data for us. I don't have to watch you, you're voluntarily giving it up all the time. So, well done. I begin with this picture, so these two things somebody asked me what these are. I shouldn't have these, these were stolen from the university of Michigan's archeological museum, I don't have them anymore. I had to give them back. These are called beveled rainbows. They're about 5,000 years old and they're our first archeological evidence of state formation.

Simon DeDeo:

So this was how you got paid when you were dragged in ancient Mesopotamia to build something. At the end of the week, you take your bowl, these are all uniform size, you take your bowl and they'd fill it up with barley to the brim, and then you take that home. If you broke your bowl for some reason, it was really easy to make a copy, you just borrow a friend's bowl, you dig a hole in the sand, you put some clay and then punch it out and smooth it around. What's nice is, that means that you can put your finger in the bottom of the bowl and you can feel the knuckle prints of this person from 5,000 years ago, which is a free, spooky experience. These are the junk mail of archeology.

Simon DeDeo:

There's like thousands, whenever you dig, you find thousands of them in the area. And I suppose the story is that it took only 5,000 years to get from here to here. And I'm going to try to tell you a story about how that happened, but more importantly, I'm going to tell you hopefully what's coming next. So this work is with a whole bunch of people at Carnegie Mellon and also Indiana where I used to be and I'll talk about the work of all these people and a little bit of the theory with a new student I have Zach at CMU. Theory and empiric, we have a very permeable boundary. What's nice is that our theories are

quite simple and the things that we say are the critical objects of our theories are actually pretty easy to measure directly.

Simon DeDeo:

So the particular numbers that we're measuring, we think actually have direct neural correlates, for example. The secret to how we got from Mesopotamia to Harvard business school is evolution, cultural evolution. You may know this book it's a lovely book by Joel Mokyr, I guess we have some people in innovation. So Joel is telling a story about what happened in the industrial revolution, particularly in Britain. What's nice about this book is an early chapter on a crash course in cultural evolution, which was first actually invented by a guy called Rob Boyd, I used to know at Santa Fe.

Simon DeDeo:

So there is a whole stack of things that tell you why somebody might observe some practice and pass it on. So we have frequency bias where you copy things, you spread things that other people are using. Velocity bias, which I guess is what you guys do, which is whatever's rising quickly you try and get it before it's common. Model bias you copy fancy people like I guess at Harvard, utility bias, it's rare, but you copy things that work. So fundamentally these all end up being biases that are generated by what you see and how you process what you see.

Simon DeDeo:

So a big chunk of cultural evolution actually ends up being about the human brain and how the human brain processes information. It's tricky to get across how powerful evolution is, it made us, it made our brains, we draw about a hundred Watts total to power our brains. They work very, very well we've never been able to replicate it, we may never do so. If you want a fun read, David Brin is a science fiction author, this book Existence is the whole story of existence, is the ways in which culture evolution powers life. So one of the things that happens in the digital era is rapid and massive acceleration of timescales. So this is the first issue of wired magazine, perhaps some of you have it. The time scale, which you would encounter information, and then pass it on, on the order of a month. Right? So you get wired, you might write in to wire write letter to identity or do you have a letter in wire you might write in.

Simon DeDeo:

So it's a timescale of a month, and that's in the modern era of digital printing and mass acceleration, I'm sorry of mass distribution. What am I going to tell you is that cultural evolution now has sped up so much in fact that in the current era, our timescale is now about a minute. Somebody posts something on the internet, you can see and respond to it roughly as fast as you can type. I think we underestimate this. So that acceleration factor is on the order of 10 or 100,000. If we had sped up biological evolution by the same timescales, we would have gone from dinosaurs to Jeff Goldblum, so things are going really, really quickly. We are experiencing a Renaissance of ideas, some for good and some for evil, I'll talk a little bit about both, We're experiencing renaissance of ideas that happen at timescales so quick that cultural evolution is now sufficiently fast that whatever we see tomorrow may be completely unpredictable from what we see today.

Simon DeDeo:

So biological evolution and cultural evolution they have one profound similarity. So here's a plant, if the plant wants to make a copy of itself, it can't jump into the three D printer and run a copy. It has to pass all of the information, all the algorithms, all the data necessary to make a copy of itself through what we

call the genetic bottleneck. So this goes through your gametes or the plants gametes, and out the other side, it makes a new plant. Now, what that means is that an enormous chunk of the structure of this plant is defined by how much we have to squeeze down through a very, very narrow channel. You can't understand evolution, at least evolution in the new synthesis era without understanding the basic properties of how this part happens. Darwin knew this and this, but it wasn't well Crick and Watson and Franklin that we understood the nature of this thing in the middle. The same thing is true of culture. We, like a cultural practice, like a copy of wired magazine can't again jump up on a photocopier and reach around and copy itself in order for this practice, this idea, this set of patterns of behavior in order for that to be replicated, in order for that to be passed along and pass this through a bottleneck as well, pass it through a cognitive bottleneck.

Simon DeDeo:

So everything that happens in cultural evolution is defined by what happens at the individual level. Every single thing that gets passed is passed by you typing, speaking, behaving, choosing. In the same way that everything that happens in biological evolution is defined by the copying and production of proteins by your ribosomes. So the cognitive bottleneck is weird, it's very, very strange. And I'm going to give you a sense of it, I'm going to play with your visual system first. One of the basic problems that we have in understanding cultural evolution is that we don't see things as they are, we see things as they expect to be fundamentally. We're fundamentally novelty resistant. We throw out differences and we replaced them with what we expect to see. So here's a really simple example, if you are a neuro-typical person, you will see square A and square B as very different shades, very different luminescence. This one to me looks lighter.

Simon DeDeo:

You know what's coming right? There are in fact exactly the same shape, So what's going on, your brain is getting photons on your retina and it's correctly registering these are in fact the same shades, as that gets propagated up your visual system, your brain says, that's not important, well we need to tell you is if this square is in shadow and it's the same luminescence as this square, it means that if you yanked that cylinder away, it would be lighter in the same amount of lanes. So your brain is constantly confabulating, it's giving you useful information or information that it thinks will be useful to you. We've learned this a little bit in the machine learning era, so I'll tell you, this is perhaps the reason that the machine learning works so well.

Simon DeDeo:

This is something called an auto encoder. So an auto encoder is one of the basic units that you would learn about, what you do with this auto encoder, I'll tell you what this architecture means in a second. Is that in the input here you might have like a thousand inputs, I guess you guys would call them independent variables, they get stuffed in here. That 1000 independent variables might get squeezed down to let's say a hundred at this second layer here. So you're tossing out 90% of the information, that hundred set of variables might then get shuffled down to something as small as 10. So again, you throw out 90% of the 90%. This image is a colleague of mine, Dave Feldman. So one of the things they like to do with this is those hundred variables at the input level will be, let's say pixels in an image.

Simon DeDeo:

The auto encoder is a clever thing because what it does is it takes a thousand down to a hundred down to 10, and then you say, "Okay, tough guy, take 10 up to a hundred, up to a thousand, and what I want

you to do is to match the original image as closely as possible." Now that should not be possible. You've thrown out nearly everything in that image, and yet it turns out, you say, "Look, I want you to match the original one." It turns out that it can do pretty well. And the reason it can do pretty well is that there's an enormous amount of structure in the world. Pictures of people are made out of lines and shadows and areas of color. You have eyes that tend to be in the same place, you have a nose, hopefully between your eyes. And what happens at this central layer here is that you've produced a reduced description of the real world.

Simon DeDeo:

Somewhere in the center here is, at least the machine language, something like "Dave is a white guy. He tilts his head a little bit, he tends to wear shirts, he has brown eyes." And what you can do then, and you've probably seen these fake faces as you can forget about this input here and just stick your finger in and twiddle the neurons and give it hallucinations. And so in fact you can create Daves that never existed. So you can go online and say, is this a real person? And we'll generate lots of fake Daves for your fake George Clooney's if that's your thing. Here's another example, this is the Ponzi illusion. Again, if you're a reasonably normal person, you will see this line here as wider than this line, you see what's coming, which is that they are in fact exactly the same length.

Simon DeDeo:

And just to reinforce that, why? Again, your brain is giving you useful information because it says, "Look, if you have convergent lines, that you think ought to be parallel, that means if you see something here, it probably means that it's further away. So we call this an ecologically valid prior, if you didn't see this line here as being longer than this line here, like you might get hit by a bus. So you see like this line here, You better remember that when that comes towards you, if you're here, you're in trouble. So your brain is again, giving you useful information as opposed to correct information. I'll do one last version of this. So you have your handouts. So start with this side here, has anyone seen their blind spots?

Simon DeDeo:

So start with this side here, put the X on the left, close your left eye and with your right eye look at the X and move this in and out. And for me, somewhere around here, up to about here, that dot on the right disappears. People see that? This is pretty weird, it's because, sorry your eyes' terribly designed. If you're in your iPhone, you have the light sensitive pixels and the wires come out the back. Unfortunately, we were built incorrectly the wires come out the front and sneak around and then you have to plug it in and there's this gap in your optic disc where there are no light-sensitive photons.

Simon DeDeo:

This disappears pretty well, you don't notice that it's gone. So if I'm looking directly here, Somewhere around, I can't see anything here, but my brain is like, "Don't worry about it, you've never noticed your blind spot." And just to give you a sense, I took a newspaper off the street in Harvard square. So do the same thing, close your left eye with your right, look at the X. At some point the dot will disappear, but behind the dot it's like, "Don't worry, it's some Chinese stuff, don't stress out." So your brain again is telling a story about what it expects to see, there is like in that dot is just fake Chinese. It's like some stereo, I mean unless you speak Chinese, in which case you'll fill it in with something else. But if you're not familiar with Chinese, it's like a bunch of lines and it's scattered in a certain pattern that Chinese often is. You can do this trick with anything, you can do this with a bureaucratic form or even if you've written a book, your own book.

Simon DeDeo:

So as should be clear, what I'm really interested in is this process here about your brain reconstructing the world. To me that is in fact the barrier for newness, that's the only thing that keeps our culture mildly inertial. The fact that we constantly discard weird things without even noticing it, so if you guys were to do something weird right now, it probably just keep going. Don't try it. That led me over the years to object to one of the main things that we do in, well in the social and decision sciences, certainly in economics, which is to reject rational choice theory. So I'll give you a crash course in rational choice theory, I think this is dualism, but I think it's fundamentally unscientific. But anyway, here's how it goes, What you know, plus what you want and, if you're advanced, what, you know you often represent in some Bayesian fashion, probabilities of things happening.

Simon DeDeo:

What you want is represented in terms of utility theory. I call this dualism because this means that there's two quantities that matter, [inaudible 00:17:29] whatever that is and bits. So this is epistemic, this is an epistemic quantity. This is a desiring or wanting quality. So if you know what you know and what you want, then out the other side you can predict what you do. So this is rational choice. It's also a big chunk of classical economics. Again, I think this is weird, I have no idea what this is. I don't know what it means to want something, when I introspect on this, I don't really have once here's another way to say it, like what do you want? Well, I guess I want what I want. It's a little bit like in biology you say I'm fit because I reproduce and I reproduce because I'm fit. There's something circular here, so I'll propose a slightly different model of the world. That's a fundamentally an epistemic drive.

Simon DeDeo:

So I'm going to throw out utils and I'm only going to give you bits. I'm going to describe you as an information seeking machine. So now, and this is okay, this is monism, so well done Spinoza, the beginning of science. So what you know in fact gives you what you want, right? And so now in fact, there's only one input to the prediction of what you do and willing to say that there were certain things that you probably don't want to do, like walk in front of the car.

Simon DeDeo:

So I'm willing to take the some external forcing, but that your desires are actually basic and shared by everybody around you. What makes us different is not our preferences but our States of knowledge. So and what we will do, in other words, it's work entirely with a representation of people's knowledge state and not try to stick in anything to do with what they actually want. You should feel free to object at any point. So I said, okay, we only have one quantity, it's the bit I will not tell you what a bit is. Does anybody know? Has anybody, I won't ask you to do this, but has anybody encountered information theory?

Simon DeDeo:

Okay, this will be fun. So Shane, I'm going to ask you to play a game with me here. I don't know if you've seen this. So here's the crash course in information theory. There's game called 20 questions, sometimes called animal vegetable mineral. The game of 20 questions is there's a child and there's a parent. Parent has something in mind and the child has to figure out what it is by asking questions that have only a yes or no answer. So okay. Shane, will you play this with me? Okay. So I have something in my mind.

Shane Greenstein:

And you want me to figure out what it is?

Simon DeDeo:

Yes.

Shane Greenstein:

Is it an animal?

Simon DeDeo:

Yes.

Shane Greenstein:

Is it bigger than a person?

Simon DeDeo:

Yes.

Shane Greenstein:

Is it a mammal?

Simon DeDeo:

Yes.

Shane Greenstein:

Boy, does it live in Africa?

Simon DeDeo:

Yes.

Shane Greenstein:

And bigger than a person, is it an elephant?

Simon DeDeo:

Yes, alright. You're the first person to get that. Wow. Yeah. Wow. This is amazing. No, I know. So what's funny actually is a couple of things I'll get to what you did in a moment. So this game is called 20 questions. Famously, this is a game on Polish, communist television in the eighties, so they bring in... It's like the plumber's union would come in it's like, "Okay comrade I'm thinking about something." And they do the yes, no question game, and the joke is that this had to be canceled when the Polish mathematicians union showed up. So they began by saying, "Is the first letter of the word you're thinking of, does it fall in the first half of the alphabet or the second?" " Okay, the first." "Does it fall on the first half of the first half?" "No." "Okay. Does it fall on the first half of the second half of the first half?" "Yes."

Simon DeDeo:

So this in fact, has got to be a really efficient way to solve the problem. If you mentally zoom in here, you'll see in fact that there's more stuff, there's more letters in the first half, than there's more in the

first half of the alphabet and the second. So you might actually want to not split it right in the middle, but say is it up to Q or something. That should give you a sense that there are better and worse ways to play this game. So it's so turns out, and this has to be the case mathematically, that there is a set of optimal questions that you can ask. Here's an example. So imagine you're playing this game against your dad. Your parental unit number one, and there's only three things in his head sorry, trees, bird and car. He thinks about cars half the time, trees a quarter of the time and birds a quarter of the time. If your goal is to get this game over as soon as possible, what is the question you should begin with.

Audience:

Is it a car?

Simon DeDeo:

Because half the time, the game is over, you won. If the answer is no, now you have to do another round. Is it a bird? And you can all say, is it a tree? And at that point, you've now split it. So half the time the game is over in one question. Half the time the game is over in two questions. So on average, it takes one and a half questions to get the information out of your father's head. It turns out you don't have to, in order to compute that number, all you need is your dad state of mind. What he tends to think about, what he tends to have in mind. You don't have to build this, you don't have to figure out the optimal script. You just plug it into this is the one formula of the talk that you just plug in the probabilities of the different things inside the person's head. And put this in this formula introduced in 1945 by Claude Shannon. It's sometimes called entropy or uncertainty or information.

Simon DeDeo:

So this is my favorite quantity, if you're a computer scientist, you may know this as Huffman encoding. So your brain, sorry, your computer, when you zip a file basically creates the optimal script for figuring out what's in the file and that's what gets transmitted. We'll need one more idea, which is the idea of sub-optimal question. So if you think of that optimal question tree for your father as some representation of what's going on in his head of some encoding. And in fact, we think that the way we encode the world has some of the same properties. If you now imagine you've taken your father's question script and you go to your mother and your mother has slightly different States of mind. So okay, your dad is an Uber driver, your mother is a theoretical ecologist.

Simon DeDeo:

She thinks about trees half the time, birds a quarter of the time, car quarter of the time. Now, if you were good, you could build an optimal question script for your mother that is as good as it would be for your father. In fact, the uncertainty in your mom is just about the same as the uncertainty in your dad. However, if you take the script, if you take the model of your father and bring it to your mother and try to use it, you will be inefficient. You will not be as good as processing that information. And so in fact, if you use that encoding you're a quarter of the question bad. So you're going to take, you have 12.5% more time on average to get the information out of your mother.

Simon DeDeo:

Again, you don't have to build the optimal question script. You compute something called kullback-Leibler divergence, where here this P is your mother and this Q is your father. So this was not Claude Shannon, Kullback and Leiber we're cracking codes in world War II for the allies. There are many different ways to think about Kullback-Leibler divergence, if you're computer scientists, it's coding

failure. If you're a cognitive scientist, it's cognitive surprise, expectation violation. We're able to do eye tracking experiments. So imagine that you see a video, you see a movie that's playing in front of you. You can build a model of all the different patches on that screen.

Simon DeDeo:

So you might say, what's the probability that the colors are this, that, or the other thing as that movie plays, the patches will change, some patches will be pretty stable. So your coding that worked at time T1 will work pretty well for time T2, in other words, the Kullback-Leibler divergences is small. Other parts of the film will have large Kullback-Leibler divergence from one time to the other, you track where people look and that's exactly where people look. We don't look at the thing that we think is cool or sexy, we don't have some preference, we're in fact our eyes are automatically drawn to the parts of the world that violate our fears in that case. All right, so that's the theoretical part now we'll do the empirical part. Are there any questions? Yes.

Audience:

So the thing is you said last, does that not go down to the thing you said earlier about us being against new information, against surprise.

Simon DeDeo:

Yeah, that's true. Well spotted. We are not entirely novelty rejection machines. In the case of movies, there's some entertainment value here and in fact, we do have a bi-modal response. So I'll tell you a little bit about how that works in a moment, and maybe 20 minutes. So, but you're well spotted, you're right to say it. So what we're going to do here is use Kullback-Leibler to study pattern making and pattern breaking. So Imagine this is the system at time T1, at time T2 time T3, this might be, for example, the probability distribution over the words in an online form. It could be the probability distribution over the colors that you expect to see in a painting or the complexity of the painting, something like that. What we're going to do is measure from one moment to the next, the spikes or the decrease in this pattern breaking measure.

Simon DeDeo:

So we're going to take Kullback-Leibler divergence and we're going to measure that. So it's short given the texts I've just encountered, how surprised am I by what happens next? So this is the text version of that cool movie eye tracking experiment. We had to give it a name, so we called it novelty. We introduced this actually, our first study of this was in something we call culture analytics or digital humanities. We use this to study Charles Darwin's reading habits. So we have... In fact, Darwin was like an early life hacker, so he kept records of all the books he read. And so we can study for example, he reads this book, how much does it break the patterns of the books that came before.

Simon DeDeo:

And that gives us a sense of whether Darwin is focusing down whether he's concentrating in some part of the scientific space and when in fact he makes leaps somewhere else. The short and long of it there is that Darwin begins very focused. He spends most of his time centered around a set of problems in geology. It's only later in his career that he actually starts jumping out the opposite of what we tell grad students to do. After the Darwin paper, I worked with Alexander Barron, also an ex physicist, and we had this challenge.

Simon DeDeo:

So we know, I hope I've given you a sense of what novelty is. So let's say again you have these cultural practices, like some distributions and probability distribution, some model of the world at time T, how much is it broken by what's happening now? And Alex did the obvious thing, which is just flip it around and we can also measure how surprising something is here given what's going to come next. This is a bit mysterious, like this direction is pretty easy to understand, but what's this direction? So we said, you ponder this for a moment and you say, I don't look like the future, if you go to your department chair and your department chair says you don't look like the future of this department, that's a problem. So we ended up calling it transient. If you look different from the future, it means that whatever you've done has died out. Whatever patterns you have tried to put in, whatever modifications you have made actually get rejected. So now we can measure two things, we can measure how new something is and we can measure how persistent it is or how transient it is. So-

Audience:

Is this just like an evolutionary experiment that doesn't persist, species.

Simon DeDeo:

Exactly. Yeah. So it's a mutation-

Audience:

A mutation that doesn't replicate.

Simon DeDeo:

It doesn't replicate. So if you have a dog and your dog has puppies and the puppies look like nothing you've ever seen, that's probably a bad thing for the puppies. So these tools are very simple, we measured this in a separate project to study the speeches of the French revolution. So I'll introduce you to that in a moment. Just as a warm up, we had fun, I used to study, to be a physicist, I used to work in string theory and theoretical physics. I took all of the papers that are published in high energy physics from 1992 until the present day. We're able to do that because physicists got into the open science movement really, really early. And so in fact, they're all publicly available, including the non-peer reviewed ones and including early drafts. So you would upload your paper to what's called the archive server run at the Cornell, called. You upload your paper when it got published, you would upload the final version, and it gives us a window into how people are understanding what it means to do physics. Here's just an example.

Simon DeDeo:

So this is actually the favorite John Luke and I were grad students together back in the day. So we take all the papers and now we have to build some model of each paper. We have to represent it epistemically many different ways to do it. So for example, a standard thing that people do is just do the word counts. So what's the probability that if you stick your hand in this paper and pull out a word, it's like, I don't know, singularity. To do this one, we use a slightly more sophisticated model called topic modeling, which instead of counting individual words, tries to extract co-occurring word patterns and then looks at the presence and absence of the co-occurring word patterns. All you really need to know is that we've turned each paper into a probability distribution so we can ask the same kind of... We can play 20 questions and we can also say, I build my 20 questions script for John Luke's paper and now, I don't know, Ed Witten writes a paper and to what extent did Ed Witten break with John Luke?

Simon DeDeo:

So here's the, here's the outcome on the X-axis here is novelty and on the Y-axis is transients, and each point here is a paper in our data, use your amazing art skills. There is a incredibly tight relationship between how new you are and how likely you are to be forgotten. I say this over and over again, everybody thinks that what people want is the new, almost all the new is rejected. This is how new you are, I emphasize this because actually it's a great example of how people reject information when I give this talk, and then the marketing people come up afterwards and they say, "Great, how do I find that optimal point of newness?" There's no optimal point of newness if your goal is to survive just be like everybody else. Our appetite for the same, yes.

Audience:

Have you ever seen the John's [Musil 00:32:59] paper on novelty, which says also that the best... It's very similar, that you can't actually be too new, that most successful research is this combination of using and citing-

Simon DeDeo:

My computer's a little funny here.

Audience:

I should be able to do this is a repeated occurrence that our IT team spent two hours and three minutes-

Simon DeDeo:

Ah, there we go, okay.

Audience:

John [Musil 00:33:32] have this finding that the influential science has enormous number, it's slightly different, but it has enormous foundation and tons of things at sites and has just a tiny bit of novelty and what we regard as foundational almost never is born, in high novelty, actually it does get forgotten, that is consistent with it-

Simon DeDeo:

Yeah. I mean we'll do a couple of examples here.

Audience:

So do you know their work [inaudible 00:34:04].

Simon DeDeo:

So it was up at Northwestern a while ago and so this obviously we battled this one out. I'll give you some, we disagree on some points and we'll see what happens, right? The people say like, "Are you sure that people want the same thing over and over. Like don't we get bored?" And the answer is no, we listen to the same Taylor Swift song over and over again. We see this on online forums. People will sometimes take a comment that's really popular and just copy it so they get like the kudos or the praise for it and people instead of encounter the first one, read a bit more encounter the second one, they just like it just as much. They liked it just as much as they did the first time. So like the first law of surprise, what is new is quickly forgotten what hasn't been seen before won't be seen much again. If your dog

gives birth to puppies that the vet has never seen before, this is a problem for the puppies. John Locke said this, "People are not so easily got out of their old forms as some are up to suggest." Biology has talked about there's something called the hopeful monster theory... Yes ma'am.

Audience:

Is there a question about this notion of surprise.

Simon DeDeo:

Mm-hmm (affirmative)

Audience:

Couldn't surprise be something that isn't new, but is somewhere I don't expect it to be?

Simon DeDeo:

Yes, one of our hidden in this model, is the idea of what you expect is based off of watching what you've seen before. In that case, what's new is also unexpected. So as a physicist you grow up, you read lots of papers. So it's not just that it's broken with what's come before, but also with your own mental model. It could certainly be the case that if you've never seen physics before, and you open a physics paper, you might have preconceptions about what happens in physics. And even though what you're reading is a totally normal physics paper, it might be very surprising to you. Like you think it's going to be about black holes and cats and actually it turns out to be something else. So okay, so obviously this can't be the whole story or life would be pretty boring culture does move. There's an enormous inertia to it, So even though, we have this first law of surprise now let's look right at these outliers. So let's look at the people who are high in novelty, but anomalously low in transients.

Simon DeDeo:

So this may take this point here, so let me just advance the slide here, take this point here and click. What is it? It's this paper that I'm sure you'd be excited to read it. It's one of the top 5% of the citations in the database, so the gap, it's so turns out the gap from here to here is a very good measure of the social power you accumulate, the extent to which you're able to put new things into the system. And to have those new things stick is the extent to which you accumulate the standard measures that are useful for getting tenure or getting the praise of others. We call that gap resonance. So novelty minus transients is resonance. The extent to which you're off that angle there. We can look at this, how do you use this? The joke, how do you use this to get tenure? This is the top 1% of papers and when I plotted here is the difference in surprise between the normal papers and that top 1% of the papers.

Simon DeDeo:

So if you look backwards, these papers are surprising given what's come before. And if you look forward, they're unsurprising given what's happened next. They're low transients, high novelty and you can even look at this curve you can look at the timescale. So if you want to really make it, you look backwards in time and break with whatever was three or four years before and your reward comes maybe two years later. So now you can time it perfectly. So first law of surprise, what is new is quickly forgotten. Second law of surprise is that the system honors the new. This is a high risk strategy, but if you want to succeed, at least in physics, and I'll give you some other examples in a moment. If you want to succeed, you have to be out here. So I think the reason people often talk about this zone, they often say this zone of optimal surprise, right here is surprise and here's when. They often think like, "Oh, there's some... Don't

be like this, be somewhere in the middle here. I think what they're secretly doing is convolving two things, surprise and risk tolerance. In fact, it's like this and it's just like, how much do you want to gamble?

Simon DeDeo:

So again, like the marketing people already say, this is inverted U curve. They may in fact be wrong, and so in fact, we find not an inverted U, but a U, and I'll give you evidence for that in a moment. I mean, that's high energy physics as string theory, let's just do this again with, it's so turns out the French revolution. So the French revolution, it was like the American revolution, but it went drastically wrong. And this is the early part of the French revolution. The King wants to raise taxes, he needs to convene, he's trying to figure out how to do it. He can't get the consensus among his advisors to do it. So what he does is convene something called the estates general. Estates general has three groups the out of the clergy, the aristocracy, and third estate. No one really knows what the third estate is, it's probably like you guys, doctors and lawyers and professors. So he convenes this thing, everybody from all over France sends a representative, they get to Paris no one really knows what to do.

Simon DeDeo:

The last time it was convened was 160 years beforehand. They get there and they decided to hold a revolution. This is what's called the tennis court oath, the King locks the third estate out of the seminar room, and they convene in the tennis courts, or I guess the squash courts nearby, and they pledge to uphold the principles of the revolution. So what we're going to do is we're going to look at these individuals and look at their novelty and transients and resonance curves. So how do we do this? Well fortunately, the French government and a fit of nationalistic pride digitize the records of all of the speeches made in the parliament itself. So in fact, we have a transcript going from the beginning of the revolution, the fall of the Basti, all the way through to the terror when everybody starts killing each other.

Simon DeDeo:

How did they do this, did they have [inaudible 00:40:20]? Actually even have shorthand very well. These were written down in different places. So, for example, some of these speeches were pulled from newspaper reports at the time. Everybody wanted to know what was happening, nobody could write it all down at once. And so in the 19th century, French historians got all the newspapers and transcripts that they could and recombined this. So this is a reconstruction, but it's an incredibly detailed reconstruction. So let's zoom in here, here's like three speeches in a row. The French revel, it's boring, most of it's incredibly boring. So the first thing they do is open with a discussion of the minutes and they argue about the minutes for about an hour or so. So it's like faculty meeting. So I don't know if you speak French, it's like this incredibly fussy debate about exactly who was like, "I like to call to your mind that your committee is recalled because you've decreed the alternative..."

Simon DeDeo:

There's some fun parts, so this is a battle that happens. [inaudible 00:41:18] says, "I want to speak on this topic.", And a bunch of people say, "Order, order." And somebody else says, "You don't have the right to speak right now." A third person says, "Mr president," the guy organizing it, "Mr president, you're really bad about this, so I want to object to how you're running things." The first guy says, "I just have four things to say," And it goes on and on and at some point people start yelling at each other, so we even have the yelling. And at the end, the president says, "If discussion will continue, Robespierre

has the floor." So Robespierre turns out to be a key feature in the revolution, like this is a terrible idea. Don't give Rosepierre the floor. So we do the same thing we do with the string theory and now the units are not scientific papers, but speeches.

Simon DeDeo:

And so for each speech we're going to say, how much does it break with the past and how much does it break with the future, we're going to measure novelty and transients. And here we go. Sorry. Here we go. It looks like string theory. So this is the density plot of the 40,000 speeches in our database. And again the newer you are, the newer your speech is, the more likely it is to fade away. It's again the same result as we get in the original case. And now of course, we do the same thing, like who sits in that high resonance space? If you're a string theorist, you sit in that high resonance space, you're able to dominate the system. What we can do is roughly the same thing.

Simon DeDeo:

So we'll look here, for example, so these are people doing things, giving speeches that are very new and yet don't disappear as fast as they ought to. That turns out to be the radical left, so Robespierre is in fact, top of the list. He's next to a guy called [inaudible 00:42:59] who is like his frenemy, they both battle it out. Barnoff disappears in the end, but both of them end up being the Kings of the parliament after the constitution is first written. And here's another interesting spot, this is the low novelty people. They say things that are very similar to what's come before and it's also low transient. So you can think of these people as the guys who keep things on track. They're the inertial dampener of the system and in fact they turn out to be the right wings.

Simon DeDeo:

So these are the conservative group that's trying to argue people towards a constitutional monarchy if they hadn't written. Here's the tennis court oath, I can show you what these look like. So here's a terrible place to be, low novelty, high transient. Low novelty means you're saying everything that other people are saying, and as soon as you say it, they're like, "Great, let's move on." So this is Barnoff, here he is. This is a guy, this is low novelty, high transients. He's in the back and here is, sorry, here is Robespierre. So this is a visual depiction of what these different parts of the space look like. Any questions on the French revolution? Yes.

Audience:

So to make sure I understand the methodology here, you're looking at texts, not at the authors of those texts and the authors come out afterwards.

Simon DeDeo:

So we pull all the speeches and we'd say in this little quadrant here, who's making the speeches in that quadrant?

Audience:

That's the result of [crosstalk 00:44:41].

Simon DeDeo:

Exactly. Yeah.

Audience:

But when this is happening, maybe I pay more attention because it's coming from all those peer.

Simon DeDeo:

Exactly. Yeah. So we don't know if it's the ideas, we don't know if it's social power, we don't know. Rebecca Spanx. So we have a French historian on this paper, so never do a project without a domain expert. And like this room is really large, so if you're a quiet person, no matter how good it is, no one's going to hear you. So like residents also could be literally like, "Oh," how you make it across the hall. Yes.

Audience:

Can I bring up the graph for a second? So is there a sense of what you don't want to be too novel, I mean your other graph had this characteristic also that the highest resonance was somewhere in the middle or well not at the extremes, maybe I'll say that because the middle is perhaps not the characterization. But so you don't want to be too novel relative to everyone else. So that's defining novelty and in a social sense, in terms of a population in which you're participating.

Simon DeDeo:

Yeah. So your novelty is relative to the practices-

Audience:

The practices going on around you, are you going there and moment [crosstalk 00:45:56].

Simon DeDeo:

Yeah. I mean, I think what you're is like, this is better than maybe being out here.

Audience:

Yeah. But had there been a lot of people out there, the better place might've been somewhere else.

Simon DeDeo:

So if there are people all the way out of here, so this is really a good question. So if everybody's novel, there's no coherence to the system. There's no actual inertial center to what's going on. So in some sense there is no revolution without these people, these are the people who are producing a consistent baseline on which others can innovate. So you couldn't actually have everybody out here, what you could have, and we tried to find this, you could have this curvature.

Simon DeDeo:

So like novelty transients up, up, up and then it could curve down. So if you're really novel, you're even lower transients than you'd expect, we've never seen that curvature. We've tried to find it like five different systems, we always see this really strictly linear relationship. That curve can tilt, it doesn't have to be one to one, the slope can be slightly lower, which means that it's novelty accepting slightly higher, which means it's novelty rejecting. But that linear fit works pretty well. So, okay. So what do we have so far? We have string theory, the French revolution. Okay. Fan-Fiction. Does anybody know what Fan-Fiction is? Someone read Fan-Fiction? Has anyone written Fan-Fiction?

Audience:

This is like the Harry Potter things that take off from the story.

Simon DeDeo:

Exactly. So this sometimes called transformative works, Fan-Fiction is like the older less pretentious way to say it, sometimes called transformative works. So what happens is that people encounter a movie series, television series, like, I don't know, Firefly or buffing. They become enamored of the characters, the characters are both structured. They learn how plots work and what they do is they write stories that expand upon the universe that authors created. So Harry Potter is a great example. JK Rowling is really positive about this phenomenon. It technically violates copyrights. So it used to be some suppressed, if it was ever discovered online, it was suppressed. JK Rowling now permits it. Some of the older fantasy writers have come along for the ride as well-

Audience:

[crosstalk 00:48:07] community and the star Wars world [crosstalk 00:48:09] around this and then there's a big fight with the star Wars community.

Simon DeDeo:

Lucas hates it. Star Trek loves it. And so I did, actually years ago, I did a qualitative study of Fan-Fiction. People tend to think of it as this very conservative genre where George Martin says no one should do this. If you look at actually what people are doing, it turns out to be quite radical. So they are often a very strong erotic component to these stories. So for example, early fiction used to be called slash fiction because the title on the bulletin board would be a continuation of star Trek, Kirk slash/Spock, which means Kirk and Spock fall in love. I worked with Elise Jing at Indiana on this project and, [Y.Y. 00:48:55] her other advisor?

Simon DeDeo:

So I said, "Great, just scrape all this from a cycled archive of our own." That's a big collection of it. And I said, "Great, what do you got?" And she said, "Well, I scraped a bunch of stuff, Sherlock and star Trek, all these things. And I have 90,000 short stories." "That's great, that's enough for a database." And she says, "No, I have 90,000 short stories solely extending the particular universe of this particular BBC dramatization of shock." This is incredible, this is like the dark matter of human creativity.

Simon DeDeo:

The stuff that's produced on this system dwarfs every single trade paperback, every single movie script. This is where people are telling stories in the world. The stories are also strange, so people use this, in fact, there's a lot of innovations that comes out of it. You may have heard of trigger warnings? Like trigger warning microeconomics. Trigger warnings and more seriously trigger warnings are things that are appended to the beginning of something that's say, warning this will discuss sexual assault for example, or this will discuss addiction. So if there's a trigger for you, don't read it. People make fun of trigger warnings, I don't think they should. It was invented in the '90s in this community. One of the reasons is that people use these stories to work out some of the most profound events in their lives.

Simon DeDeo:

So people will write stories about sexual assault in these worlds that are working out some of their own stories in their own lives. They write stories about addiction, they write stories about depression, mental illness. There also, as I mentioned, people who write erotic works. And so one of the most common

things that people are attracted to when they write this stuff is erotic parents between two male characters in a story. Just odd. Most of this fiction is written by heterosexual women. So what's going on? There's a wonderful article in a feminist journal whose name I have briefly forgotten that says, "Okay, why do this?" We'll take Kirk and Spock, this is really interesting. Kirk and Spock are very different characters, Kirk is the wild, passionate one, Spock is the intellectual, rational, mildly autistic one. So these are different, the characters are different, so there's the erotics of difference, but they're also equal.

Simon DeDeo:

You can't have the enterprise without Kirk and you can't have the enterprise without Spock. So this is a way for particular women to tell stories, to tell erotic stories without the standard power dynamics that you get where there's an older man or a younger woman, a powerful man and the weaker woman. So this is a different story. When I interview people who write these, one of them said, she just said, "Look, no one's writing pornography that I want to read. No one's writing these stories, the standard romance novels don't fit this pattern." You can see a couple of things in fact. So Sherlock and Watson are obviously another great pairing that people eroticize, and they've invented something called male pregnancy. So one of the characters magically becomes pregnant and carries a baby to term. Again, these are characters who have some equality to them. Like Watson isn't forced to do the dishes.

Simon DeDeo:

So we can do the same great string theory, French revolutionist through to Fan-Fiction. So we look within these communities and we look at the stories they write and we ask exactly the same question. Someone writes a story within a particular fandom. What happens to it? And in particular what happens to it as a function of the novelty that's happening? So here we go, Elise does her graph the different way, unfortunately, but you can't tell if the graph means anything. Now what we have here is, we also have a measure of how much people like it. So we don't just have how new it is and how transient it is, but we also have within this community they call kudos. So you read a story, you click this button that says, great story. We also have the number of reads- Yes, sir.

Audience:

Question on that then. So is transients in this context, like whether or not the topic sticks as opposed to... And so there's an inherent measure of people liking it in the transients. So if it sticks around, people like it, if it doesn't, people didn't like it. This is another dimension of social approval or social-

Simon DeDeo:

Like I might not like it, but replicated anyway. So for example, here's a joke story in the string theory case. So who's out in the high novelty low transient spot? It's Princeton University Institute for advanced study, very high social power group. What's in the conservative spot, so low novelty but low transients, that's, for example, Indiana University. Safe but slow. They're doing stuff that's pretty much like what's come before, but they're judging the field correctly. They're picking subjects that are going to be sticking around. Then you have who's high novelty high transients, one of the groups is at Penn State, and then there's this awful chunk which is the high novelty high transients, and I won't say who that university is. So Indiana doesn't get a lot of citations, but they also have low transients. So you could be part of the inertial center and nobody respects you.

Simon DeDeo:

In fact, we find the opposite of string theory, if you are, sorry, the opposite of Fan-Fiction, so on the X-axis here is Kullback-leibler divergence see score, on the Y axis here is the log base 10 of the kudos how many people click like. And all these curves essentially say that the newer you are, the less people like it. And it's a huge effect so you go a couple standard deviations out from the most inertial stuff and you drop by maybe one or two orders of magnitude and the number of likes you get. So there is a huge penalty to people moving away from what other people are doing. Makes sense, actually. So, question, why is this different for string theory?

Simon DeDeo:

So if you look at the highest cited papers in string theory are way out of the novelty curve, many different reasons. One thing that people have talked about is that physics has an institutional structure. So you can do physics and break the rules, and you still have a physics department. This is a self constituted culture, and so without that inertia, there is no discussion. There is nothing that other people know to go to. We see this, so everybody I work with does their graph a different way. So here's, and this is physics, here's the citation percentile and here's the residence. So that's novelty minus transient and you can see it goes way up. So this is the opposite of that. Yes.

Audience:

So what's the notion of canon here? So canon is the high transient, low novelty here. It's the base around which everyone can agree and is always there.

Simon DeDeo:

So low novelty, low transit. Yeah, exactly. So that's what people reward in fan-fiction. So canon, like male pregnancy, Holmes and Watson love canon, so very different from the original story.

Audience:

Partially because the Star Wars the idea of background and user generated content-

Simon DeDeo:

It's great. Okay, great.

Audience:

Part of what happened to the star Wars community is this thing called canon defined, and it anchors the whole darn thing that you see on the discussion groups that they say, "Well, according to canon and that's over and over again, according to canon. And it's a way of dismissing a lot of the novel Fan-Fiction as well.

Simon DeDeo:

So I say it's funny, I'm not even sure if star Wars is on the archive of our own. There's no attention to canon, at least in the stuff that I see, but at the same time, it's not this crazy high novelty stuff, they've just centered it around a different point. But yeah, I know what you mean, so certainly when they hire people to write the new books... So I don't know, you should look, I'm sorry. Elise has the... I think she might have the Star Wars data be interesting to see the extent to which that's really rejecting the new stuff as much.

Audience:

Wikipedia would be the other place I'd look.

Simon DeDeo:

Getting there?

Audience:

Self appointed experts, [crosstalk 00:57:31] guard the canon.

Simon DeDeo:

Right. Exactly, this is the verified version. Anyway, so Elise, Y.Y and I wrote this paper together. Our title was saying Sameness Attracts Novelty Disturbs. One thing you'll notice here is that this variance is going up and in fact that's not variance due to measurement error. That's actual variance in terms of the scatter of the points getting larger and larger. Not only is the scatter getting larger, it's a non Gaussian distribution. And so in fact the highest ranked stuff, the stuff that even blows away, tend to the four, tend to the five comments comes from the extreme. So people like this, thank you for helping, thank you for keeping this system going, thank you for giving me what I've come here for.

Simon DeDeo:

This is death zone, the mild variation. This reminds me a little bit of academia, don't do a slight twist on the last thing because people will neglect you. They don't want to deal with it. And this all the extreme events come from here. So we titled the paper Sameness Attracts Novelty Disturbs, but Outliers Flourish. What's out here, so we said if you're going to pull this and just take a look. So one of the most, in fact the most liked thing on the entire archive is a story in the guardians of the galaxy universe told from the point of view of the tree Groot.

Simon DeDeo:

And it's just the phrase, I am Groot repeated 60,000 times. So that's an odd one, it's almost a joke. Of course it's a complete outlier no one's seen such a low interview story. But otherwise for example, one of the highest ranked is a crossover between Dr. Who and Captain America. So these are incredibly novel combinations that people haven't seen before that are the sources not necessarily have changed. So we don't know if after that point Dr. Who and captain Marvel stick together to follow your point. We don't know if anyone moves over there, but it's certainly what people love. Yes.

Audience:

Just refresh my memory I'll just set up here, so it's, you can only like things, you're not giving it a ranking, it's just a thumbs up.

Simon DeDeo:

Just yeah, thumbs up.

Audience:

There's no thumbs down.

Simon DeDeo:

There's no thumbs down. We can also measure it, so there's a couple of number of reads, number of bookmarks and people leave bookmarks and they come back to it, and number of comments. Comments, kudos... Comments, likes, oh sorry, likes, reads and bookmarks all go together. Comments actually have a different phenomenon. Comments structure [crosstalk 01:00:00].

Audience:

That's the place you can express negative [crosstalk 01:00:01].

Simon DeDeo:

Exactly, I think that's right. So in fact we get a different curve there, but this is death song. So Mark, don't do this, don't make a slight, don't do new coke. One more if we have time, there's so much fun, you take a student and you say, "What do you want, what do you care about?" So this is Jenny, now a Rhodes scholar. So Jenny was Indiana with me and she's a double major in English and mathematics. So she's okay to study poetry magazine. Poetry magazine is from 1912 to is still going today. When we're writing this paper fortunately, poetry critics will say things like, "The history of Anglo American poetry is the history of poetry magazine. It's where all of the major poems first appeared." So the love song of J Alfred Prufrock, Ts Elliot is in here, Silvia Plath is in here, Allen Ginsberg is...like everybody is in here. What's amazing is that most of the stuff in poetry magazine is just terrible, it's the worst part you've ever read. It's embarrassingly bad. So this is the first poem in poetry magazine.

Simon DeDeo:

It's a poem about poetry, which is a... Never do that. And if you can read this, it's embarrassing. It feels like a pretentious, inauthentic whatever you like. So we have 26,000 poems in poetry magazine from 1912 to 2000, which is when the system kicked us off from scraping it. 72 poems make it into the Norton Anthology, so this is the canonical, speaking of canon, so the canonical thing that you get taught in school, no one is even close to this number, the next poetry journal after Poetry magazine has like six. So this is, in some sense the big story, and of course, like 26,000, 212 minus 72 was a very large number.

Simon DeDeo:

So we have a sense not only of who wins, who gets the most kudos from the faculty, but also what's going on with the conception of poetry at each point. So here's the graph that Jenny made on the X-axis here is year, on the Y-axis here, everyone does axes differently, on Y-axis axis here is, this is novelty going down. So think of this as similarity. The blue line is the average of all the poems. So take every poem in an issue, look at its similarity to the past, and this window is six months. So take all the poem by poem and average it, that's the blue line. The red dots are the 72 poems that make it to the anthology. So what's happening here, this is actually a system that flips. If you look in the early point, the majority of red dots are below this line.

Simon DeDeo:

The poems that get canonized from the early part of the century are the ones that look nothing like anything around them. Love song of J Alfred Prufrock is like, what? There's something really fun that happens in the 1950s, '60s, two things you'll notice, one is that the blue line goes down, which means that the whole magazine is going nuts. The poems here are much less similar to what's come before, and in fact, they're even much less similar to themselves. So it's not just that they've moved somewhere, but they've also expanded the range. Most of the poems in poetry magazine actually come from that era, or most of the canonized poems come from that era.

Audience:

Is that the Beatnik era or is it the Vietnam war era?

Simon DeDeo:

That's everything. That's Beatnik. I'll show you a couple examples. Yeah. It's like everything, if it's incredibly diverse. And then I think there's a new editor, so editorial switches that happen here. And then you look in the next half, and in fact the red dots are above the line by and large. So we go from a novelty accepting system to a novelty rejecting system. Of course the Norton anthologies published here. So it's in part dictated by the people who are here, and they like the things that look like the stuff. It's partly due I think to the MFA system. So in an MFA you're taught to write poetry by sitting down, reading a bunch of famous poems and writing poems that look like it. The further back you go, people were perhaps more willing to take stuff from here. This is of course, a value neutral thing. I'm not saying these are bad or good, I'm just looking at which poems accumulate social power.

Simon DeDeo:

Let's look at a pair of examples. So these are two poems published in the same year of the same issue. So one is WB8 near the end of his life, it's the last volume he ever published in 1934 poems called Mayroo, it's a beautiful poem. Civilization is hooped together brought under a rule under the semblance of peace by manifold Ellucian, actually turns out it's a talk. The illusion of sameness, the semblance of peace. Next to [Mayroo 01:05:05] is Robert Ross, you've never heard of Robert Ross.

Simon DeDeo:

This too stands up poem is a breakup poem, it's embarrassingly bad. Let's produce the curves for these two guys. So this black line here is when they're both published. Same issue. This is the year and this is the similarity. So the red line is [Mayroo 01:05:27] so it looks like this. And the blue line as Robert Ross goes like this. So what can we see? If you look backwards in time, the canonized poem is much less similar to the past, no one's writing like [Mayroo 01:05:44] 10 or 15 years before, it's rising, so the question is, was it Yeats or was he surfing the wave?

Simon DeDeo:

Meanwhile, Ross is very similar, he's writing a poem that structurally looks like all the poems that have come before. I should say Jenny's not looking at content here, she's looking at stylistic signals, so rhyme scheme, line lens, word density, things like that. Afterwards what happens while you follow the blue line, Roster style dies away, in fact, Yeats' style peaks. So he writes at the beginning of this wave and 10, 20 years later peaks in fact to go forward and both disappears such as life. If you look a little bit later, it seems like Yeats is coming back in the late nineties, two thousands. So this is a way to track how influence works. This is Yeats in that pre-1960s era when people are selecting for novelty, perhaps they found a secret that [Mayroo 01:06:39] is making it into the Norton anthology. At the same time, it looks a little bit like the string theory curve. Yeats is a high novelty low transients and also makes it in to the Canon.

Simon DeDeo:

Story's more complicated though, so I said, you inverted you. When you go into, when you look at the individual poems, you find a whole bunch of different patterns. So this is a poem by Robert Creeley, the blue is the average of poems around him, the red is Creeley himself. So you can see here he's got the Yeats' curve. He's breaking with stuff from let's say 1940. He peaks maybe 1970, his style peaks in 1970,

but you'll notice actually goes up really early on what's happening that you can think of this as Creeley reviving this older pattern. In fact, he's reviving in some sense counts. He's writing in a compressed crystalline style that we hadn't seen for many years before.

Simon DeDeo:

The most interesting curve for me is this one, it's a poem, The bean eaters by Gwendolyn Brooks which you might've read in high school. It's one of the most famous poems, it's beautiful poem and her curve is actually the exact opposite of Yeats Brooks is writing like things that look backwards and when she writes, she writes like nothing that comes next. So what's going on? There's a number of different stories to tell, one is Gwendolyn Brooks is the first black woman poet to make it in a major way within Anglo American poetry, at least the elite in Anglo American poetry. And one thing you might say is in order for her to make it and she has to pass a gatekeeper and that gatekeeper is going to take poems from an unknown source or from a source that doesn't look like other people that look like poems. If you don't look like a poet, you have to write like a poet. Of course, what we don't see here is the subject matter.

Simon DeDeo:

So Brooks is taking this older style, but at first what she's doing is stuffing something incredibly different into it. So she's bringing in the black experience, the American South, and something that we've never seen before. But stylistically her curve is the opposite of Yeats. Yeats can innovate and Brooks has to in some sense, ratify what's happened in order for that poem to be accepted. All right, so summary so far, I guess, do people know Calvinball? The Calvin and Hobbes thing where Calvin it's like the game, the rules constantly change. The reason that we reject novelty in part is that if everybody accepts it, if we're all super novel all the time, it's Calvinball. We have no stability around the system. Somehow we need to learn the expectations to bring it to your question, to learn those expectations in order for us to have a community sufficiently cohesive that we can do something together.

Simon DeDeo:

The title this talk was, where does novelty come from or where does newness come from and what do we do when we get it? I told you what we do when we get it. So I'll now go the first part, where does newness come from? If you are an academic you think that great ideas come from a cooperation, we all sit around this is great conference in Brussels where we have put straight our, I put dot quantum mechanics together, they're all friendly nobody hates anybody else in that room. We work together, we boost each other's ideas, we're polite, we're friendly, we help the grad students. That's our general image of the world. Let me dig a little deeper. Has anyone gone through peer review? So here's peer review. This is a paper I wrote in physics when I was, actually my first year of my postdoc. And I wrote it with [inaudible 01:10:23] amazing scientists now also faculty. So we did this thing, we took on this big theory, we show that there were these mathematical inconsistencies.

Simon DeDeo:

I should have expected that they would send it to the guy whose theory it was. So here, this is a quote from my referee report. "This manuscript has several instances, provides observations that are better suited for a coffee break, I could give an ultra long list of examples." He never did, but in the end actually we ended up publishing it. And in fact I met him, it was obvious who it was. And so I happened to be in Rome and I rang him up and he said, "Simon, I confess, I wrote that report." So we have a sense that may be this conflict is part of what we are. This is a healthy field, high energy physics is a reasonably

healthy field. We have a sense that this conflict and in fact with the institutions designed to protect people from the... So I didn't know who this was. I could tell just by use of definite articles that he was Italian. But I couldn't really tell who it was and when, in fact, when I met him, he was very surprised. "I thought you were some old important guy and you're nothing, I feel terrible."

Simon DeDeo:

So what this system has done is in fact insulate, as from any kind of social context, any kind of instinct towards altruism or cooperation and enabled us to just bang it out. I guess you guys are familiar with this, this is like the free market, this is adversarial legal processes where you're never in a position, if you're a defense lawyer, you're never in a position to say, "You know what? You have a good point." Like my client did it, like it's a conflict all the way. And in fact, we think that these kinds of conflicts, if we set them up with the correct institutional structure, can produce stuff that we don't see otherwise.

Simon DeDeo:

I would say peer review is one of the great secrets of modern science was in fact it goes all the way back to the founding of the Royal society. I would say when peer review gets broken in a scientific subject, the subject is not long for this world. String theory today is in a bit of a crisis. It's also completely abandoned peer review. So you can get tenure entirely on non-peer reviewed articles today. So something has broken. And in fact, what happens is if you have enough social power, you don't have to publish everyone thinks you're a genius anyway.

Simon DeDeo:

So let's do this. Let's do Wikipedia. What I'm going to do is I'm going to give this demonstration and so we'll turn out that creation, novelty is in fact associated with conflict. Conflict is generative. So I mentioned at the beginning, I don't know if I made this point well enough, like most of the behavior that I see in the world is not motivated fundamentally, cannot be fundamentally described by a set of preference functions. I think we obviously have some weak ones. So what is Wikipedia's purpose?what's it meant to... Like no one knows. Like no one has any sense of what the preference function is, people have different ideas about it. It's fundamentally one of these playful star Trek things, like let's just create something that no one's seen before and let's do it together. Here is the George W. Bush page, this is the most edited page on Wikipedia. There's 45,000 edits. I started working on this stuff in 2012, which is nice because we just keep getting more, so I can test my theories as I go along.

Simon DeDeo:

This is page to page surprise, so this is Kullback-Leibler on the Y-axis and this is time on the X-axis. So here's that particular version of the page. How surprising is it given the previous version? It looks a bit like a heartbeat, if we smoothen out, this is what it looks like. You should be able to see chop, chop. You see these moments where the content and structure of the page changes radically. This is Koon, everyone loves Koon he may be right on this. Yes.

Audience:

Sir, what's the unit of time here. Is it per edit or is it per day?

Simon DeDeo:

It's per edit, if you were to look at the 45,000 points on the screen, you would see they're not evenly spaced, so there's some points where it accelerates and decelerates in fact we'll talk a little bit about

that in a moment. So you see these innovation bursts, so I'll prove to you that they're real, you've got to check even if you believe Kullback-Leibler. So I'll show you the page, right before that first one, and right after that first one. Here's the page, this is 2004. So here's the page right before, and here's the page right after, most of the contents are actually pretty similar. The big thing that changes is they start inserting all these controversial things about Bush. So before it looks a little bit more like an encyclopedia Britannica article neutral, but also avoiding things that might anger people. The switch, after that innovation spike is a battle out of which results in controversies about Bush. These are so mild compared to the controversies we have today, it's like a different era, like he was arrested for DWI in 1976. It's nostalgic now.

Simon DeDeo:

We can measure conflict, so obviously, I was in conflict with Giovanni about that paper. It's conflict and Wikipedia is actually quite public and one of the best ways to track it, some good qualitative evidence for this from other researchers, is something called a revert. So here is, obviously this page is not going to be uncontroversial. We can look at the history of this page, so the edits people have made over time. So here we are, this is the edit history, it goes backwards in, sorry, forward in time is this way, going up the page. So I'll zoom in here. This edit here and this edit here are called a revert. So what a revert is, it turns out there was an accident of the software, that you can hit control Z or undo on what somebody has just done. So I come in, I edit like the Harvard business school page, and then the Dean comes along and says, "Nope." and he does undo.

Simon DeDeo:

Reverts are considered an aggressive act within Wikipedia, there's strong laws against them. To give you a sense of why revert is a bad thing, imagine you're working with your student, he sent you his paper, you make a lot of edits, you send it back and he says, "Thanks, I've rejected all your edits." So this is past conflict, the sociologists say, and task conflict leads to relationship conflict. You can look, so where is this, three minutes before midnight on the 28th of November, the user smooth comes in and he says, "I added a small paragraph explaining what the conflict is actually about as the current article fails in this most basic of tacts.

Simon DeDeo:

So this is not going to be an uncontroversial edit." And in fact, three minutes later this guy shows up and hits control Z. The Wikipedia system, we can take a particular article and generate a time series. So now we don't care about the content, we just care about whether it was a creative or contributive or changing edit. So an edit that changes the page to something new or revert edit that takes it back to a previous version.

Simon DeDeo:

So this is actually the time series with the George Bush page, but you'll notice is the first thing that the guy created something created the page and someone edited it. You'll notice that the third edit it's actually a revert. So it's not starting well, if you look at this there are in fact patterns in this system, this is not just a random probability of a revert. So we'll ask the first question, what's the relationship between novelty, I guess innovation, and distance to that previous revert? How far am I away from a conflict in the past and how does that co-vary with the novelty of the edit that I make. And if you do that, you find a negative relationship. So the closer you are to a point of conflict, the lower the surprise of the page and changes by about half of it, sort of half no question. Maybe five or six steps away then,

it flattens out, so this is maybe the timescale at which conflict appears to suppress creation and the timescale for people getting over it. Yes.

Audience:

You keep calling some of these changes innovation, which I'm just curious how much of the surprising stuff is in response to new information and how much of it is as in that 1W edit you show reframing, well known information. So the reframing sounds a lot more like innovation than responding to information that's [crosstalk 01:19:12] it seems like these would lead to different patterns.

Simon DeDeo:

So yes, that's coming. One of the things that's surprising is that Wikipedia is actually not particularly responsive to news. People will put news events and like a plane crash for example, obviously this gets created in response to an event. When we look for spikes of innovation in a page, we can't really tie them to any event in the news. So like the Argentina page had the spike, what caused the spike well, there was no election, there was basically nothing. It wasn't even a football match. It looks like somebody came in and was like the way in which the role of the Jews in Argentina has been portrayed as inaccurate or insufficient, so then they come in and write it. So there's a lot of endogeneity to how the system is creating.

Audience:

Yes. Could you control with Google searches?

Simon DeDeo:

Yeah. So what we look for is well, so we did Google news.[crosstalk 01:20:10] Yeah, we did news. And so I'll give you a little bit on this. In the aggregate, conflict is bad, that's maybe the friendly version of the world, like what your middle school teacher tells you. The story is slightly more complicated because we ran this time series through what we call a hidden Markov model. So what we did was try to model the different patterns of interaction. You'll notice in this blue zone here you get these long strings of cooperation with occasional interruptions by a revert and the red zone, you get much more of this RCRCRCRC pattern, and if we stick this end, so hidden Markov models are fun.

Simon DeDeo:

I won't go too much into them. But you can imagine the system can be in a bunch of these different states. When it's in one of these states, the system tends to, let's say cooperate, those are the yellow States. When it's in a blue state, it tends to revert. And what happens at this point dictates what happens next. So you can imagine the systems like happy, happy, happy, Uh, happy, it moves around. You'll notice I've drawn the arrows here, the heavier the arrow, the more likely the transition. And what we discovered in fact was the system tend to partition into two very different conflict patterns.

Simon DeDeo:

So you're in this state here, you might be in this state for years, and then at some point you flip over to this state here. This represents one pattern of interaction, this is a second pattern. This was surprising to us, this is a two phase model and it shows up in every single Wikipedia page we can find. These are the top 10 of them, so the Bush page, Wikipedia itself, Michael Jackson, Islam. And each of these cases the system can demonstrate one of two interaction patterns and it can persist for hundreds of edits, thousands of edits in some cases, years and even in more cases decade. So you had a question or no?

Very good. Couple of things about these patterns of interaction, when you're in one state, things happen quickly, when you're in another state, they happen slowly. We'll call them type one and type two conflict. So when you're in type one conflict or when you're in type one conflict pattern things can be tend to the five, tend to the six seconds on the order of months.

Simon DeDeo:

When you're in type two conflict, things happen rapidly on the order of minutes and seconds click, click, click. So what are these look like I said they're these two modes of the system, can be and have different patterns. If you're a game theorist you'll love it. So type one has some of these long strings of cooperation, but also these long strings of retaliation. So this is like, yes, yes, yes, yes, no, get out, get out, get out, get out a little bit like your canon.

Simon DeDeo:

Over here you can think of this as rejected proposal. So in this case you get these alterations and you can think of this as how about this? No, how about this? No. So type one conflict, which is this long strings of C and then long strings of R type one conflict and type two conflict are in fact very different. They have different statistical patterns. They're slower, type one is slower than type two, type two is very rapid. Let's break it out. It will turn out to be what the economists call an ecological fallacy because in fact, type one conflict has much higher innovation rates than type two conflict.

Simon DeDeo:

So you can think of when the system is in that type one state fights, it's so turns out they're actually good. So the aggregate statistics, the further you are away from a fight, the more novel the page can be. But when you discover the existence of these two kinds of patterns, you now find that when you're in the type one conflict, the novelty is higher and the type two conflict it actually doesn't really matter how far away you are from the flashpoint. The thing is, is that the conflict here is so high, that when you average these together, you get this flip. But breaking it out, you might say, look, there's a better way to fight or there's a better and a worse way to fight in the type one case, conflict predicts innovation. Innovation also predicts later conflicts, so you can think about this. You have the fights, you're doing something new, after the fight is over, it tends to stick around. So you have high novelty, low transients. In fact, the resonant page changing things happen in the flashpoint. As long as those flashpoints are a certain kind of pattern of interaction-

Audience:

So that's the healthy academic situation.

Simon DeDeo:

That's exactly, hopefully. . .

Audience:

Is that the tit for tat?

Simon DeDeo:

That's tit for tat, it's like you write something, I revert and you're like, you know what? I revert you, and I'm like, no, no, no, no I revert you.

Audience:

That's a healthy-

Simon DeDeo:

That turns out to be healthy. Yeah.

Audience:

Too much conflict is your second one, where it gets in the way of actual innovation.

Simon DeDeo:

I'm like, how about this? No, how about this? No. That's like the powerful adviser, the grad student keeps bringing new things that keep getting rejected.

Audience:

Or mistrust or something's gotten in the way of-

Audience:

Because even when it's healthy, it's healthy because there's long periods of cooperation-

Simon DeDeo:

Exactly, long periods of cooperation and then these flashpoints. But the thing is the flashpoints they create the novelty. That [inaudible 01:25:28] created cooperation is peaceful but also less interesting, than what happens at those flash points-

Audience:

Or within that period.

Simon DeDeo:

Within that, exactly within that time. So information is present, it's created, it's not more likely to disappear, so we think of this perhaps as this adversarial innovation. Data science it's hard to do an intervention, so the question is, which direction is the cause. Does innovation cause this kind of type one conflict or just type one conflict enable it.

Audience:

I don't have to do cause.

Simon DeDeo:

All right, great. So great. That solves one of the open questions. So I'll just finish here. It's two things, one is the social question, when do systems accept new stuff? How do systems treat new staff? What are the institutional constraints on the acceptance of novelty? And then of course, the flip side is what are the institutions that promote it? What are the things that make it possible for people to create? What are the micro events that enable novelty to be created? So thank you very much for coming.