



TRANSCRIPT

**Computational Antitrust First Annual Conference:
Exploring Antitrust 3.0**

On December 13, 14, and 15, 2021, The Stanford Center for Legal Informatics (CodeX) organized the world's first (online) conference dedicated to Computational Antitrust under Prof. Thibault Schrepel's direction and with the help of the entire Computational Antitrust project team. Here is the transcript. It has been cleaned but not edited. Please use it accordingly and refer to the videos published on our YouTube channel.

Day 1: Computational Antitrust Endgame

Transcript by Alessia Lercari Bueno

#1 – Introduction to Computational Antitrust

Roland Vogl

Welcome everyone to the Computational Antitrust: Exploring Antitrust 3.0 Conference. My name is Roland Vogl. I'm the Executive Director of Codex, the Stanford Center for Legal Informatics, and it's my pleasure to welcome you on behalf of Stanford Law School and of Codex to this unique three-day event.

Codex' mission is to advance the state of the art in computational law, which is the automation and mechanization of legal analysis. Earlier this year, we launched a project called the Computational Antitrust Project, which explores how legal informatics and computational law can foster automation of antitrust procedures and improve antitrust analysis more generally.

I would like to thank everyone who has helped in putting this event together, especially, of course, Professor Thibault Schrepel. Thibault is Codex' Affiliated Faculty. He's a Professor of Law at VU Amsterdam. He is the founder of this project. Not only does Thibault has a tremendous vision for bringing computational law approaches to the antitrust space, but he really is tremendously dedicated and shows tremendous grit in actually making things happen. We're really fortunate for Thibault to drive this initiative with the team he built with Stanford students and other collaborators. Thank you to all of you, thank you to the Stanford Law School Program Group for helping with the logistical aspects of this event. And, without further ado, I will turn it over to Thibault.

Thibault Schrepel

Thank you very much. It's funny because we didn't coordinate—which would be an antitrust infringement—but most of what we'll be discussing in just a minute actually ties to your very kind introduction.

Welcome everyone to Antitrust 3.0. I am thrilled to welcome you, to welcome our speakers, members of the Advisory Board, academics, practitioners, officials, computer scientists—I've seen a few in the room—lawyers, economists, philosophers, and who knows, maybe unicorns. We have quite a program for you with these three days of discussion about computational antitrust.

I thought I would be kicking us off with a little story. When I joined Codex, during the summer of 2020, in my first meeting with Roland, Professor Vogl, the center's Executive Director, gave me *carte blanche* to start a project, which came as a surprise. I'm not used to the Silicon Valley spirit. But, excited about the idea, I went to my favorite cafe down the street here in Amsterdam, and I thought about it. Codex is known throughout the world for its leadership in the field of computational law, and having a Ph.D. in Antitrust, I thought that I would be first

researching articles dealing with computational antitrust and that I would take it from there. But to my surprise, I couldn't find a single article in which the terms computational antitrust were used or discussed. So, immediately I thought that this was a bad idea, if not... a very bad idea. As you know, there are ten gazillion antitrust articles published every month. So, why would we be the first to explore Computational Law and Antitrust? But then I talked to antitrust agencies. And, here again, to my surprise, you showed an immediate and unanimous interest in the idea. So, the project was born—or at least the idea of it.

As an academic, I feel the challenge of leaving the field of ideas and moving to practicalities. And, as the French comedian Pierre Desproges once said, and I quote here: “One day, I will go to live in theory, because in theory, everything goes well.” So, computational antitrust, I thought, deserved more than being stuck in theory. The project has become a reality with the help of a fantastic team. I know that most of you are in the room right now. And I am not going to start mentioning names. But I can't help but mention Roland, Maria, Monique, Alex, Teodora and Maura, and a few others for helping and running the projects. To all of you in the room, thank you very much. I think we have already come a long way. We have been quoted in the New York Times three times this year, quoted in the academic literature, and by agencies—some of them are creating related programs, something to be discussed, especially on Wednesday.

All that to say, the Computational Antitrust Project is and was made possible only for the support we received from all places. And, importantly, I think it exists because of you in the room right now. We have received over 700 registrations for the conference. Your interest in why we publish and organize this event.

Now, back to today, December 13. As I was telling you, we have quite a program for you; I will first introduce computational antitrust with a 20-minute presentation, whose objective really is to tease you as to what is coming. I will be giving an overview of what we've been publishing and the kind of work that we do. I will then moderate a first discussion asking whether computational antitrust can succeed. As you can guess, the answer won't be negative. After that, Alessia Lercari Bueno will take the upper hand to moderate a discussion on Antitrust by Algorithm. And lastly, I will moderate the last discussion on the Legitimacy of Digital Market Surveillance.

So, without further ado, again, my objective here is to draw the portrait of what is computational antitrust. I have the ambition of discussing three different subjects here. The first one is *what* is computational antitrust. The second point is *how* do we implement computational antitrust. And the third is *when* do we do it.

What is computational antitrust?

The idea of computational law is actually very old. I think it's fair to say that it was kind of introduced, or at least the idea of it, by Leibniz, a German philosopher, who already in the 17th Century said that with the right tools, it would be possible to calculate legal outcomes, or in fact, all the social outcomes. Of course, this was a dream at the time. Leibniz never argued that it would be possible to achieve. And

what we see is that when we transpose that to the field of antitrust, there is indeed a mismatch between the time of market and the time for law. This creates a problem. The idea for the computational antitrust project is to see if we can calculate just a tiny fraction of antitrust, and automate procedures and substance so that we can close the gap.

So that's the very basic idea. To be more specific, and to give you some numbers that you may have seen. And if you are a practitioner, you're aware of this, the same if you work for an agency, we have a big problem when it comes to detecting anti-competitive behaviors. The empirical work is clear in the US, in Europe, and in fact, it's hard to find empirical work outside of those two continents. For Latin America and Asia, I believe that the numbers are quite similar. We detect about 10% to 20% of all infringements to antitrust. So, long story short, if you infringe antitrust, in all likelihood, you'll be fine. In fact, what we see is that, unfortunately, antitrust agencies do not have the means to be proactive. The OECD is clear cut on that: 80% of all cases in Europe and 90% of all cases before the two agencies in the US are coming from reactive methods. What it means is that they generally wait for a company to complain or another company to apply for leniency. But when it comes to going out there and identifying practices themselves, unfortunately, they can't conduct such studies, or, if they do, it represents a small percentage of their detection rate.

That is the picture. In this conference, we are going to discuss solutions and be on the positive side. But let me stay negative a little bit more. Unfortunately, I think that if we do not use computational antitrust, things are not about to get better. To just give you some numbers, in 2020, we produced 44 zettabytes of data. One zettabyte equals one trillion gigabytes. So, it's a lot of data already. What we produced last year is 40 times more bytes than there are stars in the universe. That's a lot, indeed. But if you look at the number, from 44 zettabytes, we're going to go to 175—or something in that order—in just four years from now, 600 in 2030, and over 2000 zettabytes of data in just 15 years from now. It seems that, again, without the right tool, it will become increasingly difficult to detect anti-competitive behaviors.

Another problem that we have is that, even after the detection of practices, we may have a problem when it comes to analyzing the quantity of all the data that we have. To be very specific, the European Commission told us that in the Google Shopping case, it had to analyze 1.7 billion search queries. Needless to say, you need to use some computational tools somewhere. Another example is the one of the CMA. They got access to Google and Bing search queries for just one week and ended up with between three and four billion entry points to analyze. So, here again, you see that quantity could be a challenge.

And of course, the nature of the data is also a challenge. As we know, unfortunately, when it comes to antitrust, we focus very much on static variables because those are the ones that we know how to analyze. When it comes to an understanding the "movie" of markets, we are actually not so good. If you are an antitrust agency, you need to have something solid; you need to be able to make a decision on the basis of something that you can explain before courts. And so, that is why they tend to

analyze those static variables. The objective of computational antitrust could also be to move toward the dynamic variables.

Against this background, we have two solutions, really. Solution number one, which I frame here in a bit of a provocative way, we give up. If we give up, we could say that the market will take care of it, which is true most of the time, but it might need 50 years to take care of it for a monopoly to disappear. So, is that good enough? I don't think so. Solution number two, you do what I call the Minority Report. I don't know if you've seen the movie, spoiler alert, it doesn't end well. But the idea in the movie is to prevent infringements even before they happen. And this is kind of a tendency that we see in the field. The idea that you could do competition law by design and that no one will ever infringe it. We need here to discuss ex-ante regulations. But again, we know that companies will try to escape and argue that they are not within the boxes of those ex-ante regulations. So, you end up again with the need for computational antitrust, which I am definitely in favor of.

I know we have lawyers in the room. So, if you are in need of a definition, here's the one that I like to use: “the automation of antitrust procedures and the improvement of antitrust analysis,” knowing that what you do is an encounter between competition tools, which are computer-based problem-solving methods, and antitrust law.

How do we do it?

I think that the big question is, indeed, if antitrust without computer science is like Sherlock without Watson, the question becomes, which Watsons do we need? Should you work for a company, a firm, an antitrust agency, where do you go? What do you do? How do you design those computational tools? Here, what I want to do, as I told you, is just to tease you a little bit. To give you a flavor of some of the discussions that we're going to have during these three days, knowing that the only thing that I will be discussing right now are the articles that we published already, but not the ones that are already scheduled for this conference, but which are not being published yet.

Number one, when it comes to anti-competitive behaviors, we know, as I mentioned, we have a problem when it comes to detecting practices. An idea that has been explored in a few of our publications is to use market screening tools, something to be discussed around today. It could also be that you want to design APIs so that the data could actually be transmitted from agencies to companies and vice versa. Or that you use natural language processing, which is a machine learning-based method according to which you provide the machine with lots of documentation, and ask the machine to identify patterns, whether or not you have flagged those patterns before, something to be discussed today. So that's number one.

Number two, as I mentioned, the analysis could be improved. It could be that you want to use more than just competition documents. It could be that the documentation in the hands of older agencies could be helpful. Here I am referring to the 10-Ks, for instance, which are in the hands of financial agencies, and in fact,

in the public domain. Those could prove helpful when it comes to better understanding market dynamics. So here, this is just an overview of what could be possible when it comes to detecting and analyzing practices.

But something, where all antitrust agencies could start from is an idea which was mentioned in an article by William Kovacic, who is to intervene in this conference on Wednesday. All antitrust agencies could, in fact, start from their own case law.

And in the second article that we published by Massarotto and Ittoo, they trained a machine learning system based on the FTC case law and actually came up with some very interesting results. Here again, I'm just here to tease you and not to explain the methodology because they will actually be giving a presentation on the subject when it comes to the sanctions.

Another article that we published is co-authored by three gentlemen who worked on the Microsoft case. What they show is that when Microsoft was ordered to generate the technical documentation to ensure interoperability on the market, they first assigned the task to 10 employees, and they ended up with over 800 of them. And here again, the idea is, can't we just use compositional tools so that we can automate part of the remedies imposed on companies.

Now, something that we'll be discussing, especially tomorrow and on Wednesday, is merger control. What we see here is that, of course, the issue is not the same. Most of the companies do notify the mergers when they have to. So, there is no detection issue. But there is an issue of time and issue of trust. Can I trust that the data set is complete when I work for an antitrust agency? And, when you work for a company, it's actually really hard to provide the agency with a complete data set, especially when it comes to guessing what the turnover of your competitors is and their market shares. Consider these two examples of complaints by the European Commission. The agency complains it was provided with incomplete databases, whether or not this was on purpose. It seems to me that blockchain could help here by ensuring that all the companies have a secure database and could actually provide agencies with access to their private blockchains.

When it comes to merger control still, another article that we just published last week is actually exploring how we could use internal agency data to improve merger control analysis. Between the FTC and the DOJ in the last 20 years, they had to analyze over 37 different mergers. It seems that this kind of data could be helpful to better understand the future. It could be that the data is also coming from outside, meaning on the market, to better predict what will be the impact of a merger on that subject.

Another paper to be published and to be discussed on Wednesday by MIT scholars analyzes what they call the fitness of a company, which is the link between a company's growth and the amount of connection that this company has with other companies. They show that there is a correlation between those two. When it comes to addressing killer acquisitions, this is the kind of data that we may want to take into account, on top of considering the turnover and the market shares.

Final point, policies. In an article that we haven't published, it seems that one priority would be to conduct retrospective studies to better understand the impact of antitrust policies and antitrust decisions on markets. But something that we've been publishing already is the idea to use machine learning to better draft policy. Scholars in a paper we published have analyzed all the contributions made to the European Commission when it comes to the Digital Markets Act and shown that, depending on the size of a company, you do not understand the keywords “dominant,” “gatekeepers,” “monopolistic,” etc. in the exact same way. It seems that this kind of information could be very helpful when it comes to actually better drafting policies that are being understood the exact same ways by all the players on the market.

Another paper that we published refers to the use of machine learning to analyze past case law to see if there are trends. For example, the number of cases discussing tying is declining. This could be helpful information when it comes to enforcing more of such cases, or on the contrary, potentially putting more of your efforts on other types of practices, which might be more trending on the markets.

And the last point when it comes to policy will be the use of agent-based modeling, something that we'll be publishing next year. The idea here is that you are actually trying to create a computer simulation and assign value to all of those agents. You could see when you change the environment—for instance, a legal rule or the introduction of a new competitor on the market—you could see that those agents will go to the right or to the left, which will be the equivalent of exiting the market, or on the contrary, investing more on the markets.

During these three days, we will be discussing improvements to antitrust, such as it exists, and slowly but surely, we'll be discussing how we could change antitrust law—that is, the actual substance of what is antitrust.

So, when do we do all that?

As you can guess, the answer is now. You can access all of our articles, podcasts, and all the news on Codex website. We are right now engaging in publishing more articles and more podcasts, something that is soon coming up your way. Here, I just want to show you the team behind the project. Again, a big thank you to all of them; you will be seeing most of them as they will be moderating panels in the coming days. I also want to show you the overview of the advisory board. Thank you very much for all of you who are commenting on articles.

Now, I want to showcase a few of the key figures for the projects, most of which I mentioned already. Something which is clear is that the computational antitrust project is 100% Codex. The encounter between computational law and antitrust, I think, was only possible within the Codex Center. And, for that reason, again, Roland, a big, big thank you for everything.

Here, there is an overview of all the competition agencies we are working with. I believe that most of those agencies are in the room right now. And, as you know, if you have more questions, feel free to always come back to me. We will be

publishing a report in just a few weeks, in which most of those agencies will explain the use they make of computational tools. We also publish one article every month; then we publish a podcast episode that goes with it. And I can tell you already that the objective is to organize another annual conference in December of next year. So, we hope to see you then.

I couldn't leave this presentation without mentioning a few of the challenges to be discussed during those three days.

Challenge number one: which fields of antitrust law can we compute? It might be easy to compute predatory pricing. It might be much harder to compute innovation-related practices. The quality of a product is something that is really hard to compute because it is really hard to put in a system and to know according to which criteria to determine if a product is innovative or not. This is one of the big challenges that we are facing.

Challenge number two: which tools? Who should be developing those tools? If they are developed by agencies, according to which standard should they actually share those tools? If the tools are being developed by the private sector, should agencies use those tools? Should they be open access?

Challenge number three relates to the role of computational antitrust, here, quoting a Nobel Prize winner George Akerlof: "there are important topics and problems which are difficult to approach in the hard way," meaning, using a mathematical formula. I think a big mistake would be to say everything that cannot be computed should be excluded from antitrust. On the contrary, we may want to have some standard according to which you could create a balance between computable versus non-computable elements.

Challenge number four is procedural fairness, something that Daryl actually discusses in his paper. A big question might be, of course, if one company has a lot of computational power, and if that company is facing a company with no such power, what do we do? This is, of course, a challenge to be solved.

And the last one is to know the limits. In my view, it would be a mistake to think that we can predict the future using computational tools. It might be possible one day; if you're interested in the idea, there is a great TV show happening in Silicon Valley called *Devs*. But this is not something for today. So, in my view, again, it remains very important to know the limits of computational antitrust.

All of that to say that there are challenges, and we are facing the limits of computational antitrust. But what we see is that the same is true for weather forecasts. And yet, according to empirical research, the quality of weather forecasts is actually improving over the last 30 years, especially because machine learning and other AI techniques are being used. And, so, the idea for the project is that, if we do not want to be stuck, living the same day over and over again—meaning here not being able to improve antitrust because we are making no use of the data that we have right now—a good solution is to use computational antitrust. The

challenge is to create an encounter between the substance of what we see on the market and the method which is being used by companies, law firms, and antitrust agencies. That is what I call antitrust 3.0. Antitrust 1.0 is antitrust of the Sherman Act. Antitrust 2.0 is the one with economic theory, and antitrust 3.0 is the one with economic theory and computer science.

So, this was, again, the overview of all the discussions that we were going to have during those three days. I am delighted that you are joining us in such a high number. A few housekeeping rules. I will be moderating the first panel in just 10 minutes. You can, of course, put your questions in the Q&A. I will first give the floor to our speakers that I will introduce, and I will be making sure to ask your questions. Now, let me return to my normal screen. Roland told me that it looks like I'm in a James Bond movie. So, let me go back to my James Bond outlets and actually introduce the first speakers discussing whether or not computational antitrust can succeed.

#2 – Can Computational Antitrust Succeed?

Thibault Schrepel

I welcome two speakers. The first is Daryl Lim, Professor of Law and the Director of the Center for Intellectual Property, Information and Privacy Law at the University of Illinois, Chicago School of Law. The second is welcome Margaret Levenstein, who is the Director of the Inter-University Consortium for Political and Social Research and Research Professor at the Institute for Social Research and the School of Information at the University of Michigan.

Daryl Lim

Thibault, good to see you and to join you in your ski lodge, wherever you are.

Thibault Schrepel

Yes, well, let us organize a computational conference there. The floor is yours. Thank you very much again for agreeing to present your paper.

Daryl Lim

Greetings, everyone. And thank you to Roland, Thibault, and the Computational Antitrust team for convening this conference and more broadly for their leadership in this project.

AI lies, I think, at the heart of computational antitrust, as you heard from Thibault. AI, in turn, is the result of the convergence of big data, computing power to harness that data, open-source culture that shares that data and rivers of cash to finance artificial intelligence. AI has delivered antitrust an unprecedented opportunity to reinvent itself, to dream big, and imagine together what an AI-powered antitrust can look like, should look like, and what we can do together to achieve it.

By starting the conference with the question “Can Computational Antitrust succeed?” Thibault and his team invite us to ask what criteria computational antitrust needs to be. The word “can” invites us to consider what do we expect computational Antitrust to do for us? And what do we mean by computational antitrust in the first place? So let me make three points this morning. First, I’ll describe what computational antitrust is—I’ll take a slightly different track from Thibault. Second, what do we expect from computational antitrust? And third, how will we know if computational antitrust has met challenges?

First, what is computational antitrust?

Well, if you’re like me, and you first heard the phrase “Computational Antitrust,” you would probably Google it, find some definition, give it context. But the first thing you discover is that Westlaw mentions it only three times. Google has more references. But both Westlaw and Google consistently point us right back here, to Stanford and to Thibault.

So, it’s safe to say that the computational antitrust website is the primary source of all things computational antitrust, ground zero. And if you visit the website, you’ll see its goal is to explore how legal informatics can foster the automation of antitrust procedures and improve antitrust analysis. The board features leading academics from antitrust and non-antitrust, legal and non-legal. The team is staffed by age energetic Gen Z’s and millennials probably more used to seeing floppy disks in museums. And remarkably, computational antitrust has caught the attention of the New York Times. It gathers probably every competition agency on every continent, except Antarctica. Then you get a sense of the scale of ambition, promise, and hope that lies on its importance and its success.

All well and good. But as with concepts like artificial intelligence, the summit of democracies, computational antitrust is basically a made-up concept, a Terra Nova. Rather than saying what it is, we need to see what ideas populate it to understand what it is all about. So, we go to the publications section, starting with Thibault planting the computational antitrust flag, making a great pitch for how agencies, policymakers, and market participants can benefit from it. We heard him speak eloquently about it just a few minutes ago. We find Anthony Casey and Anthony Niblett’s piece on micro directives and computational merger review, where they talk about how AI can improve merger notification and review using supervised machine learning. And here, it’s worth pausing to knock their concern. It’s not about feasibility, technological limitations, or the availability of data the thing. Rather, the greatest hurdle is the difficulty in pinning down a precise and translatable objective that such an algorithm would optimize.

In other words, the key to this whole enterprise called computational antitrust is to understand what we want computational antitrust to do. It is important to define the computational antitrust project so that when it enters its teenage years—whatever that means in computational antitrust years—the growth spurt that comes with it, the judges, academics, agencies, practitioners, will be able to look back at the breadcrumb trail and cost correct from time to time.

That brings me to my second point. What do we expect from computational antitrust?

Unlike many other areas of the law, judges—not Congress or legislatures, perhaps in your own countries—write the rules, and sometimes, in surprisingly capricious ways. Here in the United States, a number of five-four decisions from the Supreme Court over the years best testimony to it. As to radical shifts in how courts have treated the exercise of intellectual property rights over the years, *per se* illegal, *per se* legal. And with Qualcomm and the Ninth Circuit recently, we see a seeming shift towards *per se* illegal again.

As a result, antitrust itself, in a sense, operates like an algorithm. Its lodestar, the rule of reason, can be a black box. So, precedent can be unhelpful when allegedly anticompetitive practices are new or precedents are inconsistent with each other. Courts routinely lean on ideology too—and perhaps even more so than antitrust law as a heuristic—when they need to interpret the rule of reason in the light of economic theory and evidence. As judges struggle to apply the rule of reason, they may end up confusing the doctrine even further.

Over the past half-century or so, Chicago School conservatism has reigned in some of the excesses of earlier populist and structuralist movements. But it has also hamstrung antitrust enforcement raised the threshold for plaintiffs and everything from pleading standards to thresholds for liability, particularly with respect to single firm behavior. And, for a while now, we have heard—you’re probably familiar with this, scholars and policymakers have talked about this—that Chicago and antitrust has overshot its mark and has failed to protect consumers against corporate hegemony. Whether you’re talking about drug prices or digital platforms, the widespread anxiety of excessive private power has beckoned a return to antitrust law’s populist past or a future where issues like censoring political speech create more antitrust concerns than high prices. At least for now, under the Biden administration, the appointing of Lina Khan, Tim Wu, and most recently, John Kanter are signs that what is old is new again. The charge against antitrust law is a plausible one. Whether you’re talking about Neoclassical Antitrust, Chicago, Post Chicago, all the economic tools that we have relied on have failed, and we need something new and better.

Computational Antitrust brings together the maximum precision of artificial intelligence to address Chicago and fears of judicial ineptitude and false positives. The age-old tension between dynamic and static efficiency that informs everything from killer acquisitions in the not-too-distant past to antitrust in the Metaverse in the not-too-distant future is where I think computational antitrust can do its role in everything from detection, to assessment, to investigation, to recommending whether to launch an investigation, to prosecution, to adjudication, to the implementation of remedies. And let me just give you four quick examples.

First, with detection, natural language processing can analyze consumer complaints and media reports with antitrust potential. To use a facetious example, say you saw tens of thousands of people amassing together, and you’re trying to

figure out what was going on. So, you go to their leader. He denies there's anything untoward, but your algorithm picks up that he had lamented about the disintegration of a hegemonic Empire he was part of and describes how he was affected personally by the tough economic times that followed the collapse of when his country suffered double-digit inflation. He says to you, "sometimes I had to moonlight and drive a taxi." It's unpleasant to talk about it. But unfortunately, this also took place, and then you realize this is probably more than a group of enthusiasts.

In the same way, artificial intelligence can learn and analyze language patterns and tones of informants that applied for antitrust leniency.

With respect to adjudication, computational antitrust can scour reported cases to assess how past courts weigh competitive effects and identify influential factors. Some factors may be conventional; others may be previously unobserved. For example, algorithms could scour cases and match them against depositions and other preprocessed evidence to provide a more quick and consistent analysis. So, whether you're talking about big data, deep learning, data mining, that can help to identify relevant market variables, even in the absence of an established theory. More broadly, you can detect connections without the current legal significance that parties do not know or have no capacity to examine. Algorithms can also account for interactions about the indicators that escape expert witnesses, contextualize and associate information with the familiar and provide predictions based on entering parameters.

Eventually, this will coalesce to create per se rules of illegality or legality or find more appropriate cases for burden-shifting under the quick look approach. Courts have devised per se rules, for example, by using the damning presence of an agreement to fix prices as a sign of market inefficiency. There's no need then to engage in a counterfactual exercise giving defendants no opportunity to prove the value of those restraints. A judge's unfamiliarity of the industry at issue will then become of less importance, less of an impediment as they will be able to apply precedent across industries and, in fact, in the Supreme Court words, establish one uniform rule applicable to all industries alike. So indeed, far from a reason not to apply the per se rule, a judge's lack of experience in the industry is precisely the reason why they should do so, and computational antitrust can provide a reasoned, evidence-based approach.

To do that, of course, the opposite could happen, and you could eliminate per se rules if computational antitrust determines that these are not manifestly anticompetitive or procompetitive effects. To build up a database, algorithms can creep and mine reports on Westlaw. In fact, if you look at Westlaw, this year, they have unveiled a quick check document analysis tool that identifies patterns and connections that users do not detect. And it enables judges and attorneys to determine the merits of a case efficiently. Importantly, the results from the AI recommendations are there to challenge prior assumptions to provide a check against what behavioral scientists called coherence-based reasoning. Because

research shows that confronting people with the merits of the opposite side reduces the effects of these coherence biases by about 50%.

Third, with recommendations, AI can also predict the impact of things like killer acquisitions by running simulations to determine the optimal contestability conditions and better map synergies that help innovation pathways by tracing the user adoption of technology and Thibault mentioned my paper, where you can look at that for more detail and discussion.

Fourth, in exploring counterfactuals and crafting remedies, AI-enabled antitrust can also maximize a preset reward. And in the paper, I also talked about that.

Now, the point is, how do we know if computational antitrust has met the great?

Well, one is you should have a robust data set. And with data set, there's always that concern about bias. But my response is that algorithms do operate in a black box, but so do judges. And even that general trend of judges engaging in post hoc reasoning is well documented and discussed. In contrast, AI provides a more objective anchor to the rule or rule of reason or some other formulation in the face of coherence-based reasoning. And if you are seeking transparency, you might find accountability instead to be a more realistic and helpful goal.

And recognize that case law only provides a starting point because the precedent may or may not be based on sound economic analysis and, as I mentioned, ideology. But what happens is judges that are adjudicating life cases can compare the model's prediction with the ground truth and adjust the model's parameters, minimizing the error between those two values over time, and eventually, the parameters of the model will be sufficiently fine-tuned. And remember that perfect is the enemy of the good. It is progress, not perfection that we seek.

There are limits to computational antitrust as well. I'll just list a few. One is black swans. Thibault talked about this earlier. All data cannot predict some new events. Humans are the other limit. Of course, there's bias. But ultimately, there is also the potential for technological failure. And therefore, it's important that the human decision-maker can and must be able to intervene, especially when the algorithm has insufficient data to make a good prediction. And the third is relevance. Well, you've seen just the last few weeks a slew of decisions from the British UK authorities and from Italy. They seem to have no problems taking robust action against the market. So, the question is whether or not computational antitrust adds a meaningful incentive for them to proceed in a way that they already seem to be going in. And fourth, thought leadership. China is a voice that is notably absent from the agenda of this conference, which concerns me a little if it represents a further sign that we're not just bifurcating our technological systems but also our legal systems as well. Antitrust has been linked to the AI game, but anyone who has bought shares in companies like Alibaba will no doubt the global impact of China's antitrust model based on its notion of common prosperity. And I think that big tech conferences like these have an important role to play in placing guardrails to reduce risk, and the chance of miscalculation, as more and more of what we do becomes automated, including any antitrust space.

In conclusion, let me say that computation antitrust comes to us at a time where courts and agencies are underfunded, they are overwhelmed, and they have to apply indeterminate rules to massive amounts of information and fast-moving markets. Thibault argues that this automation, with computational antitrust, offers a way to fight fire with fire. And for the first time in antitrust history, things will get much easier. Over the next three days, the speakers I think you'll hear will reflect on the depth and breadth of skills and experiences that we will need to bring to bear to make computational and trusted success, the spear enterprise of developing a coherent understanding of computational antitrust, and this Terra Nova, and its implications on the status quo have only just begun. In this, and in all our endeavors, we can be grateful to Thibault, and to their team for providing a focal point for collective efforts. So, with that, thank you very much. I'm happy to hear from people, from Margaret and from the rest of you.

Thibault Schrepel

Well, Daryl, thank you very much. Just a few things before I give the floor to Margaret. First, that's a perfect start. Second, I did indeed use fighting fire with fire, which is a quote from a band that I'm not a particular fan of. I just wanted to mention that. But I think this is a nice way of putting it indeed. And I just want to remind everyone that if you do have questions, please feel free to put your questions in the Q&A. You can ask questions to both Daryl and Margaret. So, feel free to just drop your questions there. Without further ado, Margaret, the floor is yours.

Margaret Levenstein

Thank you very much, Thibault. And thank you, Daryl. Those were two really interesting and important presentations. And, I have to say I thought your presentation to kick us off, Thibault, was really quite inspiring. I hope as people join the conference over the next couple days, come back and listen to us, because I think it really kind of gives that overview of all the potential of computational antitrust, which like you I'm really excited about. I think that the things that I'm excited about computational antitrust are kind of different from what Daryl is was just talking about. So, I'm going to talk about two things that I think are our differences in what I would want to emphasize. One is about the theory, economic and legal theory. And the second is computational antitrust potential.

So, one of the central claims that were being made, particularly in the paper, but also in the presentation is that computational approaches to antitrust can be helpful for pushing back on the Chicago school approach and improving the evidence base for civil cases as well as others. So, while I'm very optimistic about the potential benefits of computational approaches to antitrust, I think it's important to remember two things first, about Antitrust, and second about computational analyses.

First, we talk about the Chicago school approach to antitrust, and things like the concern that judges might make certain kinds of errors because they don't have evidence, and therefore, computational antitrust can help us to sort of push back or put into context the Chicago school approach. I think that that's actually not very likely.

The Chicago School Antitrust is not data-driven. It's not based on empirical reasoning or analysis. It's based on economic theory. So, it's not a testable hypothesis, right? It is about living in the best of all possible worlds. So, if we want to change that, we need to address theory and ideology, and politics and power, because I think that's really what these things are about. If we think that the market produces the best of all possible outcomes, anything that non-market actors do, like competition authorities, is going to reduce social welfare. And you can always kind of turn that argument on itself. I don't think AI or computational antitrust is going to change that.

In fact, I think relying on that undermines our ability to address some of the real challenges of the Chicago School. The way that that gets challenged requires narratives, narratives that articulate how individuals, organizations, innovation, economy, how they exercise market power, and how that creates harm and inequity in both the short and the long run, and what's necessary to create the space for more proactive antitrust. We need those narratives about the harm caused by the exercise of market power, in order to create narratives that will allow for judges and policymakers to enact more proactive antitrust. Those narratives can be informed by economic theory and by empirical evidence of different sorts, including computational, but the narrative itself is important. And I don't think we should pretend that it's not.

Second, I think it's really, really important—and I say this as somebody who has done a lot of computational work in different spaces—that we recognize that computational antitrust, like all artificial intelligence, cannot be theory-free. There is this idea that we can have big data that answers all questions. And I think that that's a mistake. I think that we need to understand what we want to use big data, AI and computational approaches for, and what the limitations are.

To start off, and this gets back to the Black Swan point that Daryl was making, I think, very generally, that big data is very big. There's a ton of data out there. But it tends to be very, very short, in terms of the duration or the time series of data. And that's because the underlying processes that generate big data are constantly changing. When we use data to analyze a question we need to think about, what was the data generation process? Is it representative of the population being studied? Is it representative of the firms? Is it representative of the consumers? When we use design social science data, we know a lot about the data generation process. When we use big data that's naturally occurring, we pay much less attention to this. That's a mistake because we don't know what it is we're learning from the data unless we understand the processes that create it.

One great example of this is that lots of people use Twitter data and they use these Twitter feeds. But the Twitter feeds are 10% or 5% of the Twitter hoses, and those hoses are not a random sample of all the tweets. In fact, we know very little about what the 5% did we get. We also know that in the processing of things like tweets, people make lots and lots of decisions about selection. Are we using English language? Or are we using other languages besides US English? Those decisions that are made in creating datasets for analysis are actually very fundamental in terms of determining what we learn from the data.

So big data can be large in terms of quantity, it is, by definition, small, short, in terms of its duration or time series because the underlying processes that generate big data are constantly changing. The technological platforms that create the data are changing. And therefore, we're looking at the current period, and maybe a little bit of a look back, but not very far. This creates an inherent challenge to its usefulness for prediction; we are constantly trying to use data to predict behavior and outcomes in situations that are fundamentally different from the ones in which the data were generated. It doesn't mean that data are useless. But it means that the data need to be placed in the context of a theoretical model that says how does the period in which this data was generated compare to what came before and what is likely to come after. We can learn a lot about employment by looking at employment data, but during a pandemic, whenever nobody wants to walk outside their homes, we have a supply constraint that's very different, and, therefore, gives a different interpretation to the employment numbers that we see. I'm not saying that using this kind of computational method is wrong. I'm just saying that it can't be without theory, to put it into context.

Again, AI is not magic. In order to use AI, for example, to predict the impact of a merger or a merger decision, we have to have a model, either explicitly or implicitly, of how the merger affects structure and behavior and how it is in the past. If we train the model on an underlying truth deck that's based on, for example, previous merger decisions, we are building into those assumptions about what the underlying model is of behavior. We have to do that, but it's really important to make those assumptions explicit. Burying them in the truth doesn't make the assumptions go away. So, again, it's just really important that we think about the models that are that we're using, that we do not treat AI as model or value free. I think it's really important to focus on transparency, if we make decisions using AI, with little transparency, we create problems we didn't anticipate, and we undermine the ability to create the narrative about the value of competition, which I think is actually what's really critical to getting better, more pro-competitive policies going forward. If we use past decisions as ground truth, we reify and obscure past policies. I think, for example, this is why I wouldn't say we want to shift to per se rules based on AI that makes it seem like AI is sort of making the decisions for us. We actually really need coherence-based reasoning because AI is not in and of itself unbiased and relying on AI doesn't eliminate bias. It obscures it and gives privilege to people who have a certain set of skills to manipulate data, but it doesn't eliminate the bias either in the underlying truth deck or in the models themselves. While I think we can learn a lot from AI, I think we do not want to

create automatic decision-making driven by AI because we often end up ending up reifying the biases that are in the decisions that we’ve made in the past.

I do think that computational antitrust has enormous potential for things like improving evidence. I think that the usefulness of computational antitrust for detecting anticompetitive behavior or declines in competitive intensity is extremely important. I think that we can look at this, using both text data like Twitter or the language that firms are using to make their public announcements and investments, looking at price data, looking at 10-Ks, good. All of those things would allow competition authorities to detect changes and how firms are interacting with one another that would then allow more direct investigation into anti-competitive behavior. I think that this is extremely important, given the scarce resources that competition authorities have available to them.

I really liked the idea that Thibault mentioned about using computational antitrust to implement remedies as a way of monitoring what firms are doing, to make sure that they’re not responding to remedies in ways that still managed to undermine antitrust. Creating new metrics for evaluating mergers has enormous potential. Things that can be automated in terms of taking advantage of new kinds of data has enormous potential. The idea of using secure multi-party computing to allow competition authorities to integrate data from different organizations that cannot directly share data to improve the set of information that competition authorities have available to them and merger decisions have an enormous potential benefit.

One challenge that we haven’t mentioned so far, but I think is going to be really important, is that we’re trying to use this because we know that competition authorities have limited resources. For competition agencies to use computational antitrust, they require to develop appropriate skills and appropriate capacity. This kind of capacity can be expensive, particularly in today’s world, where IT resources are scarce. You can easily imagine that private firms and private law firms may well be able to expand their employment of appropriate expertise much more quickly than competition agencies. So, I think that it this highlights importance in terms of developing the capacity of agencies to be able to do this themselves and to develop the kinds of tools and metrics that I think we’d like to see so that they can make use of this.

But I would, I would urge us to try not to mystify computational antitrust or artificial intelligence. It is not going to magically solve problems for us or allow us to stop thinking about some of the big questions. For example, what it is we value in terms of the short-run versus the long-run, in terms of static versus dynamic efficiencies, consumers versus competitors, those things. Computational antitrust isn’t going to do that for us. We still have to address those things, because they are they are choices about who benefits from Antitrust and who doesn’t. And I think that the potential of computational approaches will really allow us to make better decisions, but it can’t make those decisions for us. Thank you.

Thibault Schrepel

Well, thank you very much.

There are a few questions in the chat. One of the two questions will be answered in one of the presentations later today as to which computer languages are being used. So please stick around and ask the question Jerry, because, of course, it depends on the application and the kind of what you're trying to achieve. But more on that later.

One of the questions ties to what you were discussing, Margaret, and it is about the use of computational antitrust to train systems. So, based on past case law to better understand the future, Diana asked, what do you take into consideration, which books, which case law. And the other question we have in the chat is about the time factor, which of course is very important. So, my question to you, Daryl, is which time period do you take into consideration? Because you mentioned the Chicago School. So, do you start for just the last five years? And that's something which is very interesting that Giovanna and Ashwin will be discussing tomorrow. They trained the machine learning based on the FTC. And it's very interesting to see which case they were able to use or not to use. And as you mentioned, I mean, if you do unsupervised machine learning, you will ask the machine to come up with a number of clusters and that's a human decision. And, of course, five clusters may be in your best interest when six will be a disaster for the case you're trying to make. So, which body do you take into account? How do you envision of that which data we put into the machine?

Daryl Lim

Thanks, Thibault. And thank you, Margaret. I would say amen to the last thing that you said, especially, Margaret. I think computational antitrust and AI, in particular, are there for us to make decisions better. I think it's important also to recognize what limitations of AI, it's not going to be the Messiah.

Now, if I might just respond to Margaret's comments first, which I think, as you said Thibault, ties into the question that Diana asked, I would say, first of all, with theory, I think it's not either-or. When you talk about the Chicago School and the importance for normative theoretical background, I think that's always going to be the case when you have a human agent in the equation somewhere. And that's a feature, not a bug. I think it's important to recognize, again, going back to what I said, just at the start of my comments, that you need to have a check. And it should not be an automated system that runs from beginning to end by itself. To some extent, whatever theory you want to use is an act of faith. This is something that Copernicus and Galileo had to both confront in their time. What they offered the Roman Catholic Church was empirical evidence. I'm hopeful, based on what I've seen so far, that Computational Antitrust can have agencies, judges, practitioners, policymakers, a perspective that's evidence-based in a way that the market-based faith of Chicago School and, if you look at Europe, perhaps more socialist variations of that might lack. That's really what moves antitrust from version 2.0 to version 3.0.

If you look at the question of, how do you train that data set, well, I think I would approach it the same way that a judge would approach a case. They're not going to say “well, is this case from five years ago or fifty years ago”, and therefore decide that they're going to use one or the other. Unless and until, in a common law system, something has been expressly overruled, it's still good law. So, you put it into the big data, proposition comes out, and you ask yourself, is that something that the courts with where I see the law is and the law should be going? Of course, I'm taking off my blinders and recognizing that judges do, in fact, make the law, and they don't just interpret the law. It might be that it's completely stuck in the 19th century or the early 20th century, and that's not where you want to be. Then you can flag that. I'm sorry, I'm bound by precedent. That's what it says. And then it'll be up to Congress to change it. But it surfaces bubbled-up things which otherwise might not be. They cast light on into the shadows. So, what ultimately are we seeking here? We are seeking progress. We're not seeking perfection. And I think as we move forward into the brave new world, where more and more is decided by machines—and to be very fair to Margaret, the limitations she talks about with big data, I think are absolutely true, that they are large in terms of quantity, but short in terms of duration—you recognize what you can do and what you can't do. And then you take two steps forward and be prepared to take from time to time one step back, as we have done in the pandemic, but eventually do we lose hope that we'll get there in the end? I certainly hope not.

Thibault Schrepel

Well, thank you very much. Something that came to my mind is that when you labeled a decision, you may want to label a way of thinking. But it may also be that what you are trying to label is just the type of practices, the duration and the court to try to come up with a sort of objective data. I mean, at this objective, according to other cases, which cases the agency has chosen to bring forward, which already is biased. It may be that what you see is that there are a lot of tying cases or a few tying cases, not because they are more on the market, but because agencies feel comfortable with the type of practices.

So, that's always something to keep in mind. Margaret, would you have two minutes to close this discussion? Is there anything else you want to say?

Margaret Levenstein

Well, so I am going to react a little bit to what you're saying. Part of the challenge of this conversation is there are so many things that we could be talking about that we're sometimes talking about things that are inconsistent. So, you could imagine trying to make FTC decisions more efficient by training algorithms on past decisions, and trying to say, okay, based on how we've done things in the past, this is what we're going to do in the future. That might give you faster decisions. I don't know whether it would give you better decisions. And, of course, you're only going to be able to train them based on, as I said, a small period, because you can't go back and code, all that stuff. It's going to be limited to what you can do.

The other thing I think about more—and this is maybe because I’m an economist, not a lawyer—is looking at what were the decisions and what was the impact of those decisions on the markets going forward, on prices, on product diversity, on all the outcomes that we care about going forward. But then that requires an economic model of how those firms respond to decisions, and firms respond to decisions based on a lot of things. In order to understand how to use computational antitrust, in all of its possibilities, we’re going to need as well to continue to refine our economic models to understand how firms respond to antitrust policy, how firms might respond to the use of computational antitrust. If they understand that we’re monitoring certain things, certain kinds of things that are easy to scrape, then they’re going to engage in behavior that manipulates those observables. So, we want to make sure that we just are thinking about behavior in order to keep our goals in mind as we use these neutrals.

Thibault Schrepel

Well, thank you very much. What I did not mention so far is that the conference has been structured in the following way. Day one is about the end game. So, what can we do, what are the possibilities? Day two is about more of the approaches, and here you will be hearing from very practical approaches and what they’ve done. And they will be discussing the issue of adversaries: what if a company detects the use of computational antitrust by agencies? And how do we react to all that? And day three will be about actions. How do you change the substance of antitrust? Do you implement a new antitrust program within your agency or not?

And so, Margaret, and Daryl, thank you so much for this first discussion. It has been wonderful. I see your question in the chat, which will be better answered again on day two about the use of smart contracts. Of course, every time I hear about blockchain, I get excited. So please stick around.

#3 – Antitrust by Algorithm

Thibault Schrepel

I will just now give the floor to Alessia, who will be introducing our speakers. For now, we’ll just say that Alessia is one of our editors within the computational antitrust project. And as I said in the chat, all of our publications are open source. You just go computationalantitrust.com. It will lead you to the Stanford website where you can access everything. So, Margaret, Daryl, thanks. Thanks a million. Alessia, the floor is yours.

Alessia Lercari Bueno

Thank you, Thibault. Thank you so much. Welcome, everyone. It’s a pleasure to be here today to moderate the panel, Antitrust by Algorithm. Today we have two amazing guests. We have Professor Cary Coglianese. Professor Coglianese is a Professor of Law and a Professor of Political Science at the University of

Pennsylvania, in the US, where he is also the Director of the Penn Program on Regulation. And we also have Professor Michal Gal, who’s a Professor of Law at the Faculty of Law of Haifa University in Israel. She’s also the Director of the Forum on Law and Markets there. Our panel will be structured in the following way: Professor Coglianesi will start by giving us a 20-minute presentation on the topic of Antitrust by Algorithms. And later, Professor Gal will give a brief commentary on the topic. And, of course, we will hear questions from the audience. So, please write them in the chat or in the Q&A tool. So, without further ado, Professor Coglianesi, the floor is yours.

Cary Coglianesi

Thank you very much, Alessia, for the introduction. Thibault, thank you for the invitation to be here. Today, I’m going to talk about a paper that I’ve co-authored with Alicia Lai, who is joining me on the paper, but not on this presentation, unfortunately. But I want to give credit to her for everything you like about the paper, and I’ll take the blame for everything that you don’t like about the presentation of the paper.

So, the paper is Antitrust by Algorithm. What I want to do is just provide at the outset a brief overview of the argument that the paper makes. First, that business is changing, especially in this digital economy, where the private sector is relying increasingly on artificial intelligence tools. And that development in the private sector means that to keep pace antitrust authorities are going to need to increasingly rely on the same digital tools. And that reliance on artificial intelligence or machine learning algorithms, is what I’m referring to as Antitrust by Algorithm. The paper is somewhat agnostic about substantive legal standards and whether they need to change, although I think we’re likely to see that happening as well. But, in some sense, irrespective of the legal standards, anywhere around the world, we think that these changes in the private sector will mean that public antitrust authorities will need to rely on similar tools to enhance their monitoring and analysis of the private market behavior. In making that shift to an increased reliance on algorithmic tools, antitrust authorities need to meet a series of institutional challenges, which is a principle of focus of the paper. I’m going to talk about building organizational capacity within an antitrust authority and avoiding legal pitfalls while ensuring public trust. As much controversy as there might be around antitrust these days, there’s considerable controversy over the use of artificial intelligence tools by public sector organizations. And I think antitrust authorities won’t be exempt from that, that controversy if they rely on Antitrust by Algorithm.

Let me just highlight what is certainly evident to I’m sure everyone here, the new business environment relies on increased complexity and automation by private sector algorithms. Whether it’s differential pricing algorithms, algorithms for salaries, self-preferencing algorithms. There may be just even more complexity in terms of the digital economy being new and harder to understand and monitor than more traditional well studied sectors of in parts of the economy. This also means that private actors will have not only new ways of transacting that can be for the public good and for increasing efficiency in markets, but also possibilities for

colluding innovating. And that means new challenges than for antitrust authorities in detecting illegal anticompetitive behavior.

Now, we don't mean to suggest at all that all of algorithms used by the private sector are inherently anticompetitive. They can't be productive and advancing consumer welfare. But I do think that these new developments in the marketplace can make it easier for firms to evade detection or may create new kinds of opportunities for anticompetitive outcomes and behavior.

That means that I think that antitrust authorities should see algorithms not only as potential problems when used in the private sector, but also opportunities for themselves in performing their missions. To some extent, algorithms can also help with solving persistent challenges facing antitrust authorities, and any regulator faces this. The fact is that there are many, many more market actors and transactions than there are auditors, inspectors and regulators to oversee those markets. And, when they're changing quickly, that makes these challenges even more difficult. But algorithms may be able to help overcome some of these more persistent or inherent challenges, as well as to keep pace with the changes that are happening in the private sector with industry's use of algorithmic tools.

We see that other governmental authorities and regulators are using artificial intelligence for these very purposes. There have been a number of studies in recent years. One excellent work out of Stanford by Daniel Engstrom and Daniel E. Ho, and some collaborators, about how the Federal Government in the US is using artificial intelligence tools in a wide range of areas. One of the most frequent uses at the federal level that they document is in enforcement. So, we see, for example, the use of artificial intelligence by tax and revenue-raising authorities seeking to identify taxpayers who may be evading taxes and then to be able to target the auditing resources to those taxpayers who are more likely to be evading taxes.

Securities regulators are using artificial intelligence to identify fraud and insider trading. Again, in the current uses, these are being used to guide the discretionary authority about which firms to target for greater human oversight and auditing. They can really be quite helpful in that regard. We also see various health care funding authorities using these tools and having them used to identify medical fraud as well. In all of these instances here, this is used for targeting. But that can be obviously a very important and useful tool for regulators of any kind. In fact, what we see is that in these contexts, artificial intelligence does achieve demonstrable improvements. There's research about the evidence-based demonstration of the possibilities with artificial intelligence.

In other contexts, environmental regulators, in one very good study, show that, by relying on machine learning algorithms to identify potential violators of water pollution regulations, they could identify six times as many violators with the same level of inspection resources. As before, there's a very nice recent study in Greece, where border officials have used artificial intelligence machine learning algorithms to try to identify at airports those passengers who may be entering the country who are asymptomatic but positive for COVID. This demonstrates that the use of the

machine learning algorithm is able to increase the efficiency of identifying those asymptomatic COVID individuals in two to four times the rate of what they could normally do. Other studies have shown that regulators that are enforcing the bycatch restrictions for fishers who go out into the ocean and have to limit the number of non-approved fish can improve their detection through artificial intelligence.

There's there is sound research behind this move toward reliance on algorithmic tools. And that gives rise I think, to optimism that this would also be promising in the antitrust area, obviously, we're seeing a growing interest by antitrust scholars here at this particular conference in applying these tools to antitrust. I think there's growing recognition of the role of machine learning analysis in economic research, in estimating counterfactuals, for example, and the need for more dynamic modeling in general.

There is some early movement by various antitrust regulators around the world to at least begin to start to study AI tools within industry. The UK has initiated an office to start studying this, the US Department of Justice has indicated that they're moving in this direction. I'm sure many of you are here at this conference because of that movement and that interest.

So, if we're going to be shifting toward a world more of Antitrust by Algorithm, what does it mean? How do we get there? How do we how do it? I think the key question before going forward with any use of machine learning algorithms is simply “can the machine learning algorithms do better than the status quo?” We know that human judgment and conventional analyses driven by humans have their limitations. And in some situations, machine learning algorithms can do better as the evidence that I've reviewed already indicates, but that's not going to be always the case for every possible use of algorithms. So, we always have to keep that key question in mind.

It helps to also think about what two key preconditions are for relying on machine learning algorithms in any context, but certainly in the antitrust area as well. These algorithms, by necessity, need mathematically precise goals that are built into the objective functions of the algorithm. And they also rely on big data.

If the case that you're contemplating doesn't really have a very well-specified goal, if it is what Cass Sunstein once called incompletely theorized agreements that we want to try to identify unreasonable build behavior, for example, or undue restraints on trade, if that's the degree of specificity, then it's not likely to work. We've got to really get this down into some mathematical form. And then we also need available data that's current. That will yield accurate predictions in the end that will answer the key question that the algorithms would do better than the status quo.

Assuming that that is the case in some domains for a non-trivial number of questions or uses within antitrust authorities, then then I think there are three institutional challenges that the authorities will face.

The first is to build organizational capacity for undertaking this. The GAO here in the United States, the Government Accountability Office, has a very nice framework for the development of AI tools by government authorities and it highlights the process of designing, developing, deploying, and ultimately, also monitoring and use continuously algorithms. To engage in that kind of process, it's important, again, to have data availability, and to have the capacity to store and analyze data. A lot of public sector organizations are, unfortunately, not always fully equipped with the kind of computing infrastructure that's needed. So that's certainly going to be part of building an organizational capacity for this. If this data is sensitive as market data, then you're going to need to deal with privacy and cybersecurity protections.

You're going to need expert human capital, not only data scientists, but data scientists who are aware of and sensitive to the particular challenges of applying machine learning algorithms in the governmental context. Sometimes that human capital might be obtained from private sector firms who contract with the government. The procurement process there needs to be well thought out and anticipate the possibility down the road that there may be a need for some kind of disclosure to the public or to courts about how the algorithm was designed. That will be one strategy to avoid some legal pitfalls.

Now, I think that there are not inherent legal pitfalls from machine learning algorithms. In fact, for a lot of the uses that an antitrust authority might rely on, for basic research, there really aren't going to be legal risks there. But even with respect to enforcement targeting, as long as humans are following up, probably the legal risks are low, but if we look forward and we consider the possibility of greater automation, human potentially out of the loop systems, we can anticipate that some litigation will arise. It has in other contexts when government authorities have used machine learning algorithms.

Two major sets of arguments that come up are centered on transparency. These are considered black box algorithms. It's not as intuitively obvious what explains the outcomes that they arrive at. So, there's some transparency issues, and there's concerns about possible discrimination or bias, particularly when the when data themselves have those bias. I do think there's solutions to these best practices in validating, documenting and disclosing what can be disclosed can address a lot of the transparency concerns. There are ways of engaging in algorithmic impact assessment to try to identify possible biases, and eliminate them.

This is an area where we have to recognize that there are going to be public concerns in many contexts, both about antitrust and its future and about the use of artificial intelligence by the government. But there are ways to kind of minimize conflicts and controversies. It's very important to validate outcomes. So, the last panel's point about evidence-based is really important. And to try to anticipate the impacts that the algorithms would have.

There are best practices. So, if you're not already familiar with, for example, the OECD principles on artificial intelligence would be one place to look. Committing

to as much transparency as possible is important, and making sure that these algorithms are developed with the involvement of others, whether it's through peer review, or, if possible public consultation.

Let me wrap up by just highlighting again that we're in a world which is rapidly changing and increasingly digitized. That means that Antitrust by Algorithm can be part of that market oversight in that world. Obviously, that's going to need data and computing capability. But more than just technology and data, Antitrust by Algorithm also requires responsible human decision-making about when to and how to use these tools. If we do so, it can improve the performance of antitrust authorities as it has with other regulators. And it can also avoid legal pitfalls and maintain public trust. Thank you very much. I look forward to the conversation.

Alessia Lercari Bueno

Thank you so much, Professor. Now, we will hear the comments from Professor Gal.

Michal Gal

Thank you for inviting me and for this wonderful project. Thank you to Thibault for leading the field of computational antitrust in such an inspiring way.

So, I enjoyed the paper very much. Thank you so much. And I completely agree with the three premises that it makes. The first one is that algorithms create complex competition problems. The second that there is a need to use tools that are fit to the task of use of algorithms by regulators. And third premise is that using algorithms poses institutional challenges. I will add some thoughts to each, but I will also take upon myself the fourth task and I would also relate to some limitations of the use of algorithmic regulate.

Let me just say a word first about the motivation from statistics, the use of algorithms by firms in the market, what we are facing. More than 50% of retailers in the US are using pricing algorithms. 67% of EU firms who track their competitors on a daily basis, did so by using algorithms, and 35% of them used automatic repricing algorithms. More than 1/3 of the best-selling items on amazon.com were priced by a pricing bot in 2014 and 2015. This is only with regard to pricing algorithms, there are many other types of algorithms as well, but I just wanted to give you a flavor to add to the importance of using computational antitrust.

The first premise of the paper was that algorithms create complex competition issues. And let me add to that, that, first of all, I think we should relate to the whole ecosystem. It's the conductivity, it's the data, it's the algorithm, it's big data. For example, natural comparative advantages, like network effects. All of these contribute to the complexity. And let me offer this typology, which I hope you will find unhelpful here. I think that some uses of algorithms, as you can see, simply increase competition concerns. And one example is the use of algorithms in order to carry out traditional collusion. In a very famous case, sellers of posters online engaged in a cartel. They carried it out via an algorithm. This case is also mentioned

in the paper. These are relatively simple cases. Some other cases are those in which the algorithm increases the harm. Basically, you have the exact same competition offense. However, once you use an algorithm, the harm is increased. For example, algorithms might create, under some market conditions, more stable cartels. Another use of algorithms by markets are those in which the welfare effects are debated. So, everybody agrees that not all uses of algorithms, even those which harm consumers in the short run, necessarily affect competition, and, therefore should come under competition laws. And one example is personalized pricing. And the last example here refers to the use of some algorithms that may create new competition concerns, and I will get to this at the end when I talk about autonomous algorithms.

The second premise that the paper makes is the need for algorithmic regulation. And I completely agree that this is extremely important. We are already playing catch up with technology. We have the technology, or it can be developed. This would require a lot of investment of resources. But many of the tools are out there. Sometimes you need an algorithm to go against another algorithm, because if one side enjoys speed and sophistication, you might need speed and sophistication on your side, as well as the regulator. This type of regulation takes advantage of the nature of the algorithm which is used by firms. Because such algorithms create transparency of decision making, once you can see what the algorithm is, regulation is sometimes made easier. The algorithm can also be used to design and supervise remedies, as has been rightly said before.

The third premise—and once again, I agree very much with it—is that there are institutional challenges here. But the first question that I think we should ask is, how different is this in the antitrust sphere from other areas of law or from other regulatory tools that were introduced in the past and were new to the antitrust sphere. Because there's a lot to be learned from the introduction of new methodologies in the past and from the introduction of algorithms and other spheres. We don't actually need to argue every time we have these institutional challenges, but rather, we can learn from past experiences.

Just a few pointers here. One is with regard to building capacities. From my experience of writing on some of these subjects, it's really important to create a common language for lawyers, economists and data scientists. Sometimes the definitions, the expectations and the requirements are different when you come from different points of view, or from different fields. They need to talk the same language so that we don't create an internal Tower of Babel. I think it might be useful to build cases from the bottom up, from easy ones to more difficult ones. It might be useful to have crossovers with other antitrust agencies. In the UK, the CMA is already experiencing this. And it has already created some kind of algorithms to detect some kind of behavior. They won't necessarily only work on the UK market. It might work on other markets as well. That might be a way of overcoming some of the resource problems here by joining forces. Maybe even with other sectors specific regulations. Every use should be based, at least, on a medium-run cost-benefit analysis. It doesn't make sense to use algorithms and create them for everything and anything. I think that this already came up in discussions. And

one more thing is standardization. I think that Margaret has so rightly said that sometimes the data is collected under different standards by different firms, and it doesn't talk with each other, or the regulator might not understand it, or might not have sufficient information on what exactly the data points are. So sometimes, we might want to think about data standardization. Avoiding legal pitfalls for reliance, I think it's important to think of how important it is to need a human in the loop.

It might be important, and it actually connects to the third point of building public trust as well. And I think that the question of explaining is an extremely important one. What exactly does the regulator need to explain? At what level does he need to explain? And, if the regulator needs to explain how the algorithm reached the decision that the firms in the market do not, doesn't this create some kind of imbalance? For example, I would say that providing explanations might not be needed when you are deciding which cases to bring, or you seek divergences. As opposed to cases where you use the algorithm in order to prove the existence of, for example, an agreement in restraint of trade. With regard to building public trust, I think one question we might ask here is, should it necessarily be prior to the first use of the technology? Or maybe we can build some good cases, and then build trust based on them?

Finally, I want to say a few words about the limitations of algorithmic regulation. Because one question I think that we should ask ourselves is, will we lose what we have gained by the speed and sophistication of using algorithmic regulation once cases come to court? The agency is many times only a very, very important stock, but not the final stock. Judges would need to understand the complexity, which might be based, for example, on statistical inferences, or machine learning, or even deep learning analysis. How do we create a situation where judges are open to such understanding? One way to go about it is maybe to create specialized tribunals, or to create new methodological presumptions, with regard to these new tools. For speed and sophistication, we can use more interim measures once the algorithmic regulators indicate that there is a serious antitrust problem.

As a final problem here is that some conducts are not captured under competition laws, even if it is harmful to competition and to consumers. And it is carried out by algorithms. The example is autonomous algorithmic pricing, where you have an algorithm that simply increases or maximizes profits. And the algorithm determines this by itself. And the question is, of course, can algorithms facilitate coordination? And the answer was given by the OECD in 2017. Theoretically, yes. And where we are now is that there are experiments that were performed that have found that algorithms can actually learn to coordinate by themselves. And you see that algorithms increase the price by themselves. And there are many other studies that are now accumulating, which are showing that this is happening in the real world. This cannot be captured by our agreements in restraint of trade because there's no agreement. So, what we need here is to think beyond computational antitrust and in changing our laws. Thank you.

Alessia Lercari Bueno

Thank you so Professor Gal for the presentation. Now, we have a few questions from the audience. Of course, if you have a question, you can still write it down while we answer the ones we already have.

The first question is about open-source technology. What is the role of open-source technology in developing a modular toolkit for algorithm regulation? The question raises the issue of whether open-source would be a possible solution to have less costly ways to develop these technologies that will be needed both by the private parties and by the public entities. I would love to hear your comments, both from Professor Coglianese and Professor Gal.

Cary Coglianese

It's worth putting that out on the table. I think it's a great question. One of the things we're seeing with the development of AI tools in a variety of areas, private sector and public sector, is the emergence of a lot of soft law around governing the use of these AI tools. An open-source framework is one way really of developing and managing that governance capacity. It has great transparency. It's a natural question, to ask for addressing some of the concerns that do arise when governments and other entities are using AI tools. But, of course, if this is a tool that is going to be deployed for purposes of enforcement, maybe that's not going to be appropriate, and there will be limitations on transparency. I think Gal's question about how much transparency and how much explainability is needed will vary from context to context. That's true today, even outside of the context of artificial intelligence. I have written this paper called Transparency and Algorithmic Governance, in which I highlight how in the non-algorithmic world we have, for example, exemptions from the Freedom of Information Act disclosures here in the United States that's required of government agencies when they are dealing with information related to law enforcement. Transparency and explainability are largely pragmatic, and it does vary depending upon the use case. Open-source tools might be a really great way of actually developing some industry standards that could be useful for monitoring market activity and standardized market activity. But it may be limited in other circumstances as well.

Alessia Lercari Bueno

Thank you so much, Professor. Professor Gal, do you have any comments on that?

Michal Gal

I completely agree. I think that open-source can sometimes help. But again, it depends on what is out there. With regard to open source, for example, it would surprise me to know that a deep learning algorithm can be written in five code rows. And, so a lot of the algorithms are already out there. The problem sometimes is the data rather than the algorithm, and training the algorithm on data, so it will learn from it in this specific case.

Alessia Lercari Bueno

Thank you so much, Professor Gal. Professor Coglianese, maybe you would like to provide your reactions based on the comments of Professor Gal. Due to time constraints, I think we have a few minutes for this.

Cary Coglianese

Oh, well, let me thank Professor Gal for those terrific comments. I really do appreciate them. There's a lot in what she said that enhances the argument that we're making in our paper. The point that she makes about starting with the easy cases before moving on to the more difficult ones is really very important. In some ways in other contexts, we've missed that a little bit. Here in the US, there's a lot of controversy over the use of algorithms, they're not even machine learning algorithms, but use of algorithms in the criminal law setting to make determinations about sentencing and bail decisions, highly consequential decisions about people's liberty and locking them up in jail. It's not surprising at all, that this has generated an enormous backlash. Part of it is precisely because we didn't really follow Professor Gal's advice. There's a risk I think it's worth putting out on the table here in the antitrust context. There's a risk that any antitrust authority around the world that goes forward and uses some of these algorithmic tools in ways that are, first of all, highly consequential, they're really deeply relying on it, and, second, they're doing so without really competently or responsibly handling the design and validation and execution of the algorithm, that you could generate a lot of controversies that could actually set back the sort of the cause of computational antitrust for everyone around the world. So, I think we need to be aware that this is something to approach with a good degree of care and thoughtfulness, and responsibility. There's a collective good here in using this new approach and these new tools. So I want to really emphasize that point by Professor Gal. Starting with easier cases, building successes with them, being able to validate and develop the institutional capacity to use these tools well before going on, and actually tackling the biggest and the most difficult uses.

Alessia Lercari Bueno

Thank you so much, Professor. We have a few seconds, just for one last question. I would like to hear comments from you both. The question is how do you envision these new capabilities being implemented by antitrust agencies? As new departments within organizations, or by adding data scientists to exist in departments? And maybe I would like to add, who is going to be training this new personnel?

Cary Coglianese

Stanford, right? That's not the answer you're supposed to give here?

Alessia Lercari Bueno

I agree with that.

Cary Coglianese

I will say, just to piggyback on my last point, there was a controversy by folks over at the Medical School at Stanford in using a machine-learning algorithm to determine who should get the COVID vaccines first. It got a fair amount of public attention because it appears, and this is just based upon my reading of the media reports, it appears that the algorithm prioritized vaccines given to those members of the medical school who were most experienced, for example, which makes sense in most contexts, but they were the ones who had the least exposure usually to patients. So, if you were looking at it from a risk standpoint, all of these senior physicians who spent most of their day in their offices were prioritized to get the vaccine and the residents and the doctors who were spending time really with the patients and who had the greater risks of exposure were less priorities. Anyways, this is just to say that any organization, if they are not going to approach this with care, can run around a foul of blinders. We've seen this in other contexts. I have a paper called the Limits of Performance-Based Regulation. Regulators have had these blinders in a lot of other contexts over the years.

On the precise question about where, how should we do this? I think the answer, quite frankly, is all of the above. There are some economies of scale for building some specialized expertise that could cut across different areas within an antitrust authority. But I think you'll also need expertise within individual offices as well. And one thing that the question didn't mention but is also a reality here is that a lot of the expertise for many public agencies is coming from the private sector. So hiring consultants, and there's a good bit of care that needs to go into thinking about your procurement contracts with these private consultancies. You hire, in particular, thinking about the concerns that Professor Gal was mentioning about how ultimately, if you're using this tool, and it winds up in court, and it's being challenged, one of the things that the court is definitely going to insist on is good information about how this algorithm has been designed, validated and tested. Some consultants have as their normal default to claim IP protection over everything that they do. So, in the procurement process, you have to also be thinking ahead and making sure that those contracts are written in a way that gives the antitrust authority the ability to, at least in court, be able to disclose more information about how the algorithm functions.

Alessia Lercari Bueno

Thank you so much, Professor. Professor Gal, maybe a 30-second remark about this, please.

Michal Gal

I think that it depends on how many people you have. Because I think that if you only have a few, they should sit together, and they should create some tools and be able to talk with each other. Another final point I want to make is that what I hear from some agencies that the problem is that is getting data scientists and computer scientists, because right now, they get very high salaries in the world. So, I think

that one way around it is to create a strong motivation, and for what these algorithms are used for, then you're using it for a really important cause. So, Thibault, your project is right on point.

Alessia Lercari Bueno

Well, thank you so much to you both for this panel. I look forward to reading the article. We will be sure to publish it and let everyone know. Thank you so much.

#4 – The Legitimacy of Digital Market Surveillance

Thibault Schrepel

Thank you very much. Now, we only have one more discussion to go through today. And, you know, I want to react because there are quite a few questions in the chat as to how to build this expertise. And what you're going to hear about right now is the actual expertise. And the way to integrate the expertise is the subject for Wednesday discussion when you're going to hear from the agencies. And please feel free to ask them questions as to how they deal with all that. I'm sure they'll be more than happy to answer your questions. But for now, before building the expertise, we should agree on what we want to do. That's the question we're going to try to answer, and I'm pleased that I am not the one in charge of answering it [laugh]. One question is whether or not we want to do market surveillance, how to do it and whether or not we should push for it in the field of antitrust. For that, we have three wonderful speakers. I realized when preparing for the panel this morning that this was a Dutch panel because of your affiliations. So, we are welcoming Catalina Goanta, who is an Assistant Professor in Private Law at the Faculty of Law in at Maastricht University. At the same university, I welcome Jerry who is an assistant professor at the Department of Data Science and Knowledge Engineering. And we have Marietje Schaake, the International Policy Director of the Stanford University Cyber Policy Center and International Policy Fellow at the Stanford Institute for Human-Centered AI. Marietje also served as a member of the European Parliament from 2009 to 2019. So very excited to hear your thoughts about market surveillance and for you to share your experience. Without further ado, Catalina and Jerry, the floor is yours.

Catalina Goanta

Thank you so much to both, first of all, allow me to align myself to the other speakers in thanking you not only for your vision in terms of making this project happen but also for this wonderful event that I can only imagine how much time it has taken to organize. It's really a pleasure and an honor to be in this panel here and to bring a more European dimension to some of the things that we already heard.

And I am very happy that especially I am going to try to build on a lot of the things that Cary specifically has been mentioning. But what I'm going to try to imbue in this presentation is the consumer protection perspective as well. So, you can have

computational antitrust, and you can also have Computational Consumer Protection. At the end of the day. It means the same thing. You're trying to look at digital markets and trying to understand what exact violations can be detected. And how can the automation of these violations be detected and actually pursued?

So, what Jerry and I have prepared for today is a discussion of a paper that we also have upcoming in the series that is so completely unique and current right now at Codex in the computational antitrust Project. And what we're going to try to do today is, first of all, to think a little bit about the idea of market surveillance and the concept of digital enforcement. We are going to try to define that from a terminological perspective with some examples from European public policy. And then what I'm going to do is also take you on a journey through an example of how digital enforcement looks like currently in the European Union, by taking you through the CPC regulation, the consumer protection cooperation regulation, and giving some examples of powers that authorities have, and also making a bridge to what Jerry is going to present, namely, the idea of how exactly to put these powers into practice through technology, by focusing on the pros and also the cons in terms of also what Cary was mentioning earlier. That is the idea of putting these technologies or the deployment and the development of these technologies in private hands. And we're going to conclude with some questions for debate, which hopefully, are going to also fuel the conversation and the discussion and with Marietje.

So, perhaps even before going into this context and terminology, also an anecdote, because I know Thibault that you like anecdotes. So, I was sharing this event on Twitter a few days ago, and basically also with the title, the idea of legitimacy and market surveillance. And I remember there was a comment, and someone said, "Wow, that sounds really creepy". Well, it definitely doesn't sound as creepy as what Cary was mentioning in terms of predictive policing. So, all of the algorithms that are being used in the criminal justice area, or some of the examples that we also had in the Netherlands, with more social policies and the detection of fraud, for instance, in tax or, or social subsidies.

But actually, if you think about market surveillance in the history of market surveillance in the Netherlands, I think that this is one of the least controversial, if we may say so, topics that are going to bring with it the question of how do we detect things, what kind of carbs do we want to focus on. And that is mostly because of the fact that when the internal market was created, the idea was that the actual freedom of movement of goods and services brought with it a responsibility to make sure that the kind of goods and the kind of services that were offered on the internal market were safe, that they were as safe as consumers would expect them to be. And this is something that Monique Goyens, the Director of the European Consumer Organization, also emphasizes in her scholarship. The fact that, actually, consumers in Europe expect that when they go, and they try to purchase a product, or when they go into a store or get a service, they expect that to be safe. And safety is actually the main goal of market surveillance as we have it also defined in a new framework.

So, there's this Market Surveillance Regulation from 2019, that also reformed, perhaps years, if not decades, of prior market surveillance activities at a policy level in the European Union. As you can see the entire idea of market surveillance that we have in the European Union law relies on the idea of product safety, the fact that products should be made available in a safe way, and that we shouldn't be able to engage in transactions that don't do that.

Now, from this concept of market surveillance, which has a long-standing history of legitimacy, which we actually discuss in greater length in the paper, we go to the concept of digital enforcement. And we have seen that in the AI White Paper that European Commission came out with last year and the AI Act, where we see so many references to the idea that digital markets require digital and enforcement.

So, from this perspective, digital enforcement has this dimension of a broader meaning that does not only entail product safety, but it entails several narratives, and we can find some in the work of Margot Kaminski. For instance, this idea that whatever is available in the offline world should also be available on the online world in terms of redressing harms and taking notice of harms that may occur. Of course, we also have a different narrative because there are some harms that might be specifically or intimately native to the digital markets. And this is why even this concept of digital enforcement right now has a little bit of an identity issue. But basically, what we try to do is to clarify it and to say that it is broader. And it is a trend that we see in European public policy of doing enforcement on digital markets.

An example of such enforcement we can see in the consumer protection cooperation regulation, which entered into force last year. It is one of the most important sources of power for authorities in the European Union, that are trying to look for harms that are affecting consumers, specifically on digital markets. And we see that it focuses on three categories of infringement. Now, the tests for these infringements are highly, highly complex. But let me just take you through an example — the widespread infringement. So, you can see that there is a calculation of how many member states need to be to have been affected by specific harm, which can be past harm, ongoing harm, or a likely harm. And then that harm needs to be detrimental to the collective interests of consumers.

My role here today is not to really take you through the nitty-gritty details of this regulation, which I as I mentioned earlier, is highly complex. But to give you some examples of why it is important, and that Jerry is going to build on with the more normative questions that we have in our paper.

This regulation actually really reforms a lot of the powers and gives more grit to the powers of consumer protection authorities, which in some jurisdictions are the same as competition authorities. For instance, it actually formalizes a right to do mystery shopping. Any kind of authority can go and try to purchase goods from an e-commerce website, and then look at what the results are and what kind of violations may result. This is an example of investigation powers.

We also have the so-called enforcement powers, which, in a way, try to compensate for perhaps at least two decades of less effective regulatory frameworks in the realm of consumer protection. This is something quite generally acknowledged that traditionally, for instance, for digital markets, the enforcement of consumer protection dealt with a lot of legal uncertainty, and therefore was often basically brought up as being ineffective. However, now, the doubling down on what it means to have effective means for the situations of this widespread infringement, the three categories of infringement, you can see some examples here, such as the idea that an authority has the power to remove content or to restrict access to an online interface, which is huge. It's really something that gives authorities a lot of leeway in dealing with rogue traders.

And an example of that is what has been happening in France, just a few weeks ago, the way in which French consumer protection authorities have been using these new powers. Basically, they have used them against a company that is based in San Francisco actually wish.com. And that is because they have done this mystery shopping and notice that a lot of the products actually posed a lot of harms for consumers because they did not comply with, for instance, labeling, or they had, perhaps, ingredients that shouldn't have been on the list of ingredients. And this has been a problem. So, this is an example a very recent example of how the French consumer protection authorities have made use of these powers. Of course, the thing is that, if you've read the entire press release, you can see that the investigation that has been done has not been an algorithmic investigation.

And what I'm going to do is pass the floor to Jerry and discuss the potential and how perhaps these kinds of investment investigations could be scaled up as an example of what Computational Consumer Protection can do and how that can also inspire perhaps computational antitrust.

Jerry Spanakis

Thanks, Catalina, for taking the lead on this and the presentation of the paper. Thanks to Thibault and his team for this wonderful event. I'm happy to be here as well, also joining a panel where people have weird names. That's why I'm going with Jerry, which is not my original Greek name, but my international name.

So, in our paper, we were discussing the context of digital enforcement and automation and how it should be done, and by whom. And here, what we are, debating is exactly the different consumer and competition authorities. They are all eager to use technology and in different ways. However, what we also see and mentioned in the paper, based on the work of Deirdre K. Mulligan and Kenneth A. Bamberger is that what we see usually in these cases is a procurement mindset, instead of a policy mindset. What we mean by that is that most, let's say, state agencies' administration just tries to contract technical systems that might employ some kind of technology, could be machine learning or anything else, and the government relies on exactly this technology, which comes from private parties. And of course, nobody can blame them because they're not technology leaders,

they have no resources, and they have to rely on what is out there in order to provide this automation and these solutions to the public.

One of the criticism points is the emphasis goes on tools. Because people use tools thinking that there is this magical one that can fix anything, right? And, of course, one of the other challenges is the resources and priorities of consumer accommodation authorities, which are, of course, very limited, both in terms of financial resources, but also into knowledge and technology resources. The question we have is who should design these computational frameworks that do not yet exist? And should we just rely on the private sector? And all the industry actors that are taking the lead on this? Should the public sector take a role in that even with limited resources? And of course, what is the broader role of academia here? Academia is a great partner for many different reasons. First of all, because we are knowledgeable about what we're studying, but also because we're not profit-driven. So, we're really driven about what and how things should work. And here, we mentioned the example of dark patterns, which is an example of a digital enforcement project, and how that can scale up. It was led by an academic institution at Princeton University.

So, moving on, what we want to stress is that we have this needs not for tools, but for a technological framework that can work today and tomorrow, and then we can discuss how that can be specifically implemented. That goes a bit also to the presentation of Cary about what is the technological framework? And how can we put machine learning and AI into use? So, what we need and we propose here is investing in people and infrastructure. If you want a tool today, that's great. It might do your work for the next month or so. But then, technology moves very fast. So, you might be very well outdated with something that was built today. Instead, investing in people and being able to go past this procurement mindset will give you the room to have technology that can work at any point.

And the focus should be in this framework and what should include, and not just in technological solutions, but also focus on transparency, fairness, and accountability. And usually we try to shape that around three pillars, about what kind of data do we want to utilize, should this data be used, who should decide what data should be used, what kind of organization model we want to visualize as well. Algorithms by themselves are not going to give us a solution because they rely on the data we use.

Finally, the approach is on how we want to scale things. And what we want to know is how to evaluate this approach to how do we measure the outcomes. And here we have just a screenshot of the dark patterns' repository on GitHub, which is open source, and as it was mentioned before, open-source is, I would say, the way to move forward. It's also what we are using to educate our students these days. So, this is the mindset of development, however, it does come with challenges. This is a public repository, it can be used and picked up by anybody including researchers, but also consumer authorities. However, having something which is open source means that it also needs to be maintained. And this also needs additional resources. So, what we've maybe experienced so far is that businesses—and you can imagine anything here like retail, business and airline or anything else—have just been

profiling consumers. We all know that, even if we are not aware of it, we know, somewhere, somebody was trying to monitor our consumer behavior and try to recommend new products or try to pressure us to buy the products or show us relevant advertisements. We've seen all these stories many times. At the same time, the state is at least monitoring and trying to go into this surveillance of citizens, especially in the context of COVID-19. I sure you're aware that different states and different governments try to implement all this Mission Impossible of vaccination in different ways, and different forms of digitization.

There is this triangulation here. On the one hand, businesses monitor and profile citizens for different reasons. Social media is a very good example. And what we want to raise as an ethical point of view is, what if States start implementing this massive surveillance operation on consumers just in the sake of digital enforcement? So, what we want to say here is how surveillance should be set up in a transparent way so that consumers are not harmed? So just to summarize a bit and also trigger a bit of questions for debate. Here are four questions that we have in our minds and would also like to hear your opinion.

First of all, how will this policy mindset be put in practice when it comes to designing, but also applying computational enforcement? At the same time, how do we defend and protect, first of all, government agencies but also consumers from snake oil risks? If you don't know the term snake oil, which was introduced by Arvind Narayanan, you can just Google it or put very simply, it refers to approaches that promise something, but it's not really there. I would say that most of the approaches around that call themselves artificial intelligence tools these days might fall under the snake oil category.

At the same time, I talked about transparency and the role of it in the whole design and implementation of these administrative procedures. How should that be integrated and implemented? And finally, there are concerns. How can we make sure that weaker party—consumers—are protected from increased surveillance and maybe extreme powers just because we're implementing this monitoring? And I think with that, we are done, hopefully on time. And yeah, we're looking forward to the discussion.

Thibault Schrepel

Thank you very much. Marietje, it's going to be very easy for you to answer all of those questions, which are just tiny questions. Right. But I'm very much looking forward to your comments. So, the floor is yours. And I will just say that we will address the questions in the Q&A. So, feel free to enter more questions there. We have quite some time for discussions.

Marietje Schaake

Thank you very much. And congratulations to Catalina and Jerry for their very thought-provoking paper. I have to say I didn't know we were all having a Dutch connection, even though it feels to me that we're a very international team. But it's

very nice to have such a Netherlands anchored crew at Stanford because it connects to my two of my worlds.

Anyway, indeed, a lot of questions. I'm not necessarily an antitrust expert. But of course, antitrust is a very foundational principle for fairness in the economy. And, so, everyone who's ever thought about tech policy as I have, or who has been a member of Parliament's has thought a lot about antitrust, and also how it is applied. So, my comments will probably come more from an applied side and from a research side, but I hope that they can contribute to your thinking as well.

So, on your first slide, you spoke about legitimacy. And I think that that is really a key question. Not only in terms of who has the authority to apply oversights today, but also what might we be missing, and are there questions about whether companies, for example, have the legitimate mandate to assert as much power as they do to monitor behavior and to engage in surveillance. What really helps also to clarify some of the questions that you ended with regard to protecting the citizen from undue surveillance versus legitimately mandated surveillance is to ask yourself—and this is, in the context of law—who can legally know what? And, also, who benefits of that? And this may also be a little bit of a remedy against all the snake oil or tech utopianism.

To put an example on, surveillance is creepy when data is hoovered up by big companies, or when it's repurposed. Or perhaps we actually need more surveillance of big companies in the interest of protecting consumers & citizens. And I should add, and I think you made that clear in your paper too, that monitoring the markets is a normal part of compliance. Think about checking whether goods on a ship are properly cooled, or whether they meet the quality criteria that apply in a certain market, or the safety criteria when protecting children or passengers in cars, for example. So, monitoring goods is very normal. And some of such checks, and some of such monitoring can be done with the naked eye, or information can be deducted in a lab. So, let's say you have a drink, you don't know exactly what's in it, you put all kinds of testers in it, and you can assess whether the company has been compliant and so on. But trade and economic development has gone from goods to services, where there's much more to monitor in the chains of commands in decisions taken by leadership. Think about financial services, for example, or the emails that were also made public when they were seized from Apple or Google. When their actions dating years back, were probed to assess whether their decisions made then were actually intended to negatively impact their competitors, which is also an element of antitrust enforcement. So, after goods and services data is now a growing part of trade and economy, a larger part of the so-called markets and also a context, digitization data processing, a context where wrongdoings in the context of antitrust and competition law can happen.

One key observation here is a to place your paper in a broader context is that while consumer protection is a goal of antitrust enforcement, there are also many other methods and other criteria to assess antitrust violations. It's one slice of assessing harms. But it's not the only slice of assessing harms. Antitrust rules may actually have been violated without a single consumer having been harmed. You don't

necessarily have to prove that one consumer or the other has been harmed. But if there have been price agreements made illegally, or a merger has led to monopoly forming or things like that, that can already be harmful in and of itself.

With digitalization and more and more information and data in the hands of private actors, the big question is, what do we know about the market? And where are the markets? So, a lot of that monitoring, and I think also monitoring that you talk about is external. It can be observed, it can be scanned, it can be scraped. But there's also a whole different side, which is, you know, internal proprietary data processing that happens outside of everyone's purview. And I think that that is exactly where the antitrust authorities want to look. Because that is actually the domain where it's much harder for a person to prove that their rights have been violated. And, to prove that one's rights have been violated is sort of a *sine qua non* on for redress. I mean, if I cannot show that I've paid too much, it's really hard for me to show that I've been wronged and then to claim any kind of compensation or action from the regulator.

The real question is, in a digitized, increasingly proprietary governed context of businesses in the digital economy, how might an aggrieved party—which can be a consumer or a business—discover? And you talked about transparency, fairness and accountability, which are, of course, major objectives of regulation in themselves. But then the question I think, becomes when we look at this frame that you are using in your paper between procurement mindset and policy mindset, will there be standards regulated for that are going to help facilitate oversight? For example, accounting type standards that trace algorithmic decision making, so that even in hindsight, when an outcome is there, it can be traced back to what kind of decision was made, how data was processed, how people's behavior was changed, and so on, and so forth.

That also leads to questions of oversight of software. And also, then, and I think this is interesting, in the context of what you were saying, how will the market surveyors be overseen? Who is checking the one that's checking? If these are companies that are hired by a regulatory body, there will never be blind trust. Either there has to be clear guidelines, or there has to be some kind of standards according to which companies can surveil the markets and be in line with the democratic mandate that antitrust authorities have.

If you had more time to work on this paper would be really interesting to link much more directly to antitrust law. So not just to consumer harms, but to antitrust law. And the reason is that antitrust authorities are remarkably strong. They are actually, broadly speaking, stronger than a lot of consumer protection authorities because of the ability to access and demand proprietary data. It's almost like a legal subpoena to demand for information that otherwise would remain hidden for everyone involved. They have extraordinary mandates. And there is also work being done now to evolve those mandates to specify the application of antitrust and competition law in the digital era. So, think about the Digital Markets Act, for example, in the European context, that is really intended to spell out even more specifically for tech companies, what their responsibilities are as major market

players, what is expected of them in terms of, for example, behavior, vis a vis, smaller size, businesses, consumers, and so it's a really interesting and dynamic field, in and of itself.

So, to conclude, the separation or the two mindsets that you spell out in your paper, the procurement mindset for the past mindsets, I think they actually interact with each other quite a bit. So, on the one hand, regulators, including antitrust regulators, may choose to bring onboard technological solutions to help them monitor and regulate. But at the end of the day, it has to be crystal clear through policy, who is ultimately responsible and who sets the terms of use and who oversees the overseers. And what is important to remember is, in antitrust law, which we often see as a huge hammer to go after companies with, the aggrieved company, the company against whom a case is brought Microsoft, Google, Gazprom, also have strong rights. So, the regulator who claims that antitrust rules have been violated must have a very strong case that stands up in a court of law. Therefore, there should not be and there cannot be any room for ambiguity about how the software works, what the details are exactly, and so on, and so forth because the mandates and the legitimacy of the regulator at the end of the day come from the public. And so, everything has to hold up in a court of law. And I think that is a sort of backstop that can be helpful, and that also will lead to new criteria for what any kind of market surveillance software will have to adhere to. So, I foresee a lot of policy, a lot of criteria, and then possibly on the basis that more procurement from authorities. I'll leave it there, and I've learned a lot. So, thank you for including me, and congratulations on putting together this great conference.

Thibault Schrepel

Well, thank you very much. And what I see is more academic papers, which makes me very happy. Now, what someone is mentioning in the chat is the irony for an antitrust agency to investigate a company using one of the tools provided by the company. And here, who's surveilling those agencies if they make use of such tools? There is another question in the chat, which is related as to how do we measure the harm that could be caused by computational antitrust? And one question that I had, and you see it all comes back to the same idea is, is there an example in which you will ask for the agencies to bear the cost of their analysis? Because it seems to me that it could be a great way for them to be very cautious. So, Catalina, Jerry, if you have any thoughts on the subject as to how do we ensure that we are not creating a monster, that'd be great. Ideally, we will fight fire with fire, but not monsters with monsters, right?

Catalina Goanta

Maybe just to start, thank you so very much, Marietje, for your really important comments, because they were taking away two main things. So, first of all, the linking to more antitrust. What we wanted to do with this paper is to create also an illustration of how the same discussions about computational analysis have been happening in consumer protection. I'm happy that you mentioned this idea that antitrust has been the policy sector where the enforcement has entailed going into

companies, physically removing information, removing servers, removing data and that this has been considered such a scarecrow in terms of enforcement, and it has been considered to be so effective. This is exactly what also inspired me to bring in the paper the CPC regulation. I think that you can see that consumer protection has taken inspiration from antitrust in this respect. You can see that the CPC regulation is very much also in line with the DSA where there are these new kind of revelations about how we are going to have access to the data that we need to in order to audit the activities of these companies. And this is where it becomes quite the answer to the question of how you do that in a consistent and transparent way.

I think it really boils down to the actual tasks that we're speaking about. So, what are the monitoring tasks? I wouldn't go into any examples—because I see Michelle is also asking in the chat—about what kind of computational tools are being used, because I think that these computational tools that are currently being used, are developed with some degree of secrecy. They are part of this adversarial environment where you have a market player trying to do something, and then you have to find a gap, and explore a gap, or maybe a regulatory gap or, or a specific practice that they try to circumvent the law with. Then the regulatory agencies try to catch that or to fill the gap, and so on and so forth. So, there's this ping pong.

But what I could say is that, if you should look at very specific tasks, for instance, I think competition authorities are right now very interested in this, and this is why we gave the example and perhaps Jerry can build on this as well. I am talking about the topic of dark patterns, that basically all consumer protection authorities in the world have gotten to work with some version of. What are we going to do about this problem? For those of you who might not be familiar with dark patterns, they refer to the idea of having interface elements that are supposed to be manipulative, and that would change because your behavior to such an extent that, for instance, under European law, would be considered an unfair commercial practice. This is a very big problem. And I mentioned this example, because it is normally, under EU law, covered by the UCPD directive, which is unfair competition. So, it's a realm where consumer protection and antitrust really will meet each other.

But if we look at how exactly do you make dark patterns detection transparent and accountable, and what how do you measure the harms that arise out of that? I guess my answer would be that it depends on which dark pattern you want to measure. So, I don't know if Jerry has some very specific thoughts on this. We heard those from Daryl earlier, the idea of doing NLP on consumer complaints. And then for instance, looking at testimonials, as fake testimonials, as a dark pattern would be something different than looking at maybe sneaking into basket, even from the measurement and evaluation of that. Jerry, if you want to build on that, or anything else.

Jerry Spanakis

I also want to build a bit on the comment related to investigating Google with services they use, and also something that I think Marietje shared about the

internal process that exists in companies. As computer scientists, we're designing all these kinds of studies that we're going to start with that website and trying to measure everything they do. Is like we are looking at the mirror of the real world and we are hoping that what we're seeing is a good reflection of what was happening underneath. So, in a way, it might feel like "why are you doing this". Because, in the end, we're trying to get data that maybe already exist, right in the company. And of course, they have way more data than we do. So, what they want to say by that is that it's not just about maybe monitoring, but maybe taking a step back and seeing what shall be allowed or not allowed? Should Uber, for example, have access to my battery level? I don't know. And how I'm going to measure that? There was this other example, I think from Amazon where a book was listed for 23 million, plus shipping. Just because there were two pricing bots that were colluding with each other in the end there was no violation from the algorithmic point of view, but there was [a violation] about what kind of data they were using.

And, building also on the examples that Catalina mentioned, you need to know what you are looking for. Otherwise, it will be like "let's scan the whole web and try to find what is wrong". These are important points that we need to take into account on how to design these processes in the context of transparency, fairness and accountability. They're not just broken mirrors or wrong reflections of what we're dealing with.

Thibault Schrepel

Marietje, is there anything you want to say? Then I'm going to ask an even more complicated question.

Marietje Schaake

I'll brace myself. Two things. I think it's really important to distinguish between which actor has which authority so to really not to confuse the type of monitoring that companies are doing in the interest of improving their products and services, selling more, being smart about prices, possibly manipulating prices, knowing what competitors are doing. You know, the whole Amazon model of using its platform for knowing what is sold best, and then putting those products in the market. That kind of surveillance done by a private company is a completely different situation from the regulator, let's say, monitoring the exact same steps. The mandate is different, the role they have is different. And so, compared to other cases, where, if you're a civilian, and you're using an arm and you shoot it at someone, it might be a completely different situation than a police officer shooting. So, it's really important to keep in mind who's doing what and why. When regulators are doing it to keep fairness in the economy, they're doing it on behalf of the entire society, whereas a company is acting probably mostly in self-interest.

The second thing I wanted to say is that we see a lot of companies hiding behind the argument that they simply didn't measure something. So, we see a lot in disinformation. It's a convenient way of denying something. Simply saying, well, we don't know how many of x we take down. But it's a choice not to measure. I think one way to look at is to say companies know that they can't violate antitrust rules,

and they cannot get away with saying, we accidentally monitored when the fuel runs out in the car. And then we figured out x y & z. Companies should be held accountable for violating Antitrust. But what I can imagine, again, is that there will be standards of how they must monitor their own activities so that when they're probed for possible violations of Antitrust, that they have something to hand over, and that they can't say, the machine made all the decisions, we have no logs, we have no idea. It's a black box, and sorry, the black box may or may not have violated the rules. I just don't think that that's sustainable in an open market and open society.

Thibault Schrepel

Okay, let me come up with more complex questions. Something we heard today is the need to integrate expertise. And, of course, it goes both ways. You want to work with computer scientists, potentially, not just at the end of processes and without asking them for the magic tool because it does not exist. It also means that you need to work with lawyers. And it seems to me that if you were just to hire a computer scientist and say, "okay, you design and you do everything yourself", then procedural fairness will go down. And this is not what we want. And so, if you need to increase your expertise in computer science, and to make sure that you have lawyers dedicated to working with those computer scientists, what you need is to hire more and the right persons. And something that Michal mentioned is that if you are trying to hire a computer scientist versus a big tech company that may pay 200k a year, this will be a challenge. And so, it seems to me that a big issue is the one of money, and you have to be able to put some money on the table eventually. You will be able to attract a few experts that may believe that they will increase the common good if they work for the agency, which is true, but for the most part, you need to put the money on the table. And so, my question is how would you present that to a parliament? How would you make the point that we need more budget, and this budget needs to be allocated to those issues? Is it something that you think is possible? How would you introduce that? I guess that's one of my worries. Being in touch with a few agencies, I see they are convinced for the most part that computational tools can help. But they may fear that, politically, it might be complicated for them to go and ask for a budget and to hire the right person. So; how would you, go about that?

Marietje Schaake

I think one thing, that's the good news, is that, if you're the regulator, antitrust enforcement is actually quite a popular thing to do. And it is non-polarizing. So normally speaking, conservatives and progressives, they all agree that you need some enforcement of market rules. So, when I was a member of the European Parliament, I always felt strongly about not politicizing it. Members of Parliament, from time to time, would say, oh, break up this company or break up that company. And I would say, well, you know, the authorities have the power to break up a company if they think that it's necessary. I don't have the access to information, you know, etcetera, etcetera. So that's one thing.

The second thing is of course, that fines in the antitrust context bring in a lot of money. You may have seen the Italian antitrust regulator has just fine, I think, 1.8 billion euros yesterday, and there will be appeals and so on and so forth. But there is a case to be made that some of the financing of antitrust enforcement should simply come from antitrust sanctions. And I'm pretty sure that, I'm not aware of the budgets exactly, but I'm pretty sure that what is spent to enforce antitrust rules is already quite significant. Because there are also lawyers involved, and antitrust authorities have experts. I mean, they work with strong teams, because their case has to stand up against the bank account of big companies that can hire the best lawyers in the world. So, I, if it would be me, as an executive, to make the case via parliament that we need proper resources, I would not be afraid to have that debate, I think Parliament's would actually grant that authority because they see the benefits. And because actually, the fines that come in would probably cover all those expenses and more.

Thibault Schrepel

Alright, so we'll just give the floor to you Jerry, and Catalina. I'm curious as to what is your experience with consumer protection agencies and how they deal with computer science expertise. I just want to mention that something that the FTC is now doing, and I think this is excellent, is to say, okay, for each dollar that you spent on the FTC budget, it actually gives back the American people, I think it was \$18, eventually. This is very convincing. I know that the French competition agencies engaged in the same kind of studies. So, publishing this type of metrics in the annual report will be something that I would love to see more, instead of just focusing on fines, because it may create a perverse incentive. But what is your take on that, Catalina and Jerry. Is it very difficult for consumer protection agencies to hire computer scientists? Are they're doing it? Are you positive about the future? Or are you about to leave us in a negative note right now?

Catalina Goanta

Definitely not going to end the event on a negative note, because I think that there is only one way ahead. And that way is basically for more a more kind of hybridization of enforcement where we have these brick-and-mortar situations and consumer protection and in competition law. But then we also have digital markets, and the only way in which we're going to be able to keep track, because this is also what you started out with. All of the data that is available on digital markets, the only way in which we're going to be able to do this in a scalable way is going to be through some degree of automation.

Now, the question is how to do this automation in a responsible way, and not in a way where we kind of idealize technology as the universal panacea, and then you get a tool as if you're buying a Microsoft license, and, with that, you can just do absolutely everything under the sun in terms of antitrust violations, or consumer protection violations. A lot of the technology that we are talking about here today just doesn't exist. I'm very curious to see the next days of the conference, I'll be following that closely. But there are a lot of pilots, that is for sure. I think Cary also gave a lot of very interesting examples. But these pilots, basically they bring

together a set of interests of agencies, and then a set of possibilities that are technical. We have seen also in our work with Jerry, marrying these two perspectives, it just it takes time. Interdisciplinarity that entails breaking down a legal norm and trying to create a computational framework for that is not something that happens overnight. And it's not something that you can go and solve through maybe more go-to approaches that also computer science has had. For instance, crowdsourcing, as a lot of labeling is being crowdsourced and computer science, that's not the same as legal labeling. That's very expensive. And we have seen in a lot of different projects that it takes a lot of time to get some serious legal labeling going on. But this is the only way, and in my experience, I've seen that there are there's a lot of positive attitudes and consumer authorities in terms of getting data scientists on board. The question is indeed how to do that in a way that doesn't that is not going to create, at least in the European Union, a European Union in two paces, where if there is going to be resources and then you invest these resources then you will have the state of the art at least, in vision and policy. And then if there aren't enough public resources to invest that into recreating the architecture, then we might even end up having this type of like forum shopping where maybe if you do business in some member states, then you know for a fact that you're not going to be necessarily traced online as a company. So those would be some thoughts. Jerry.

Jerry Spanakis

Yeah, I totally agree. Just to add to that, two things. One thing is that interdisciplinarity is difficult, but if I can talk to Catalina as you can understand me and the other way around, and somehow, we can respect each other's limits, then things are possible. The other thing is in relation to the budget spending. I think it's just a matter of also how you sell it. Because the idea of investing in people and long-term solutions might be cheaper in the long term, instead of "I don't know how to do that, or I don't know how to do that. So, I need to invest in some procurement tools to give me the AI tools". So, in the end, it's also a matter of putting down the numbers and persuading the right people. But I think, based on our discussions with Catalina and different consumer authorities, that the mindset has changed. But of course, you know, it takes time to land for good.

Thibault Schrepel

Let me ask you a very final personal question. Jerry, are you receiving offers from the private sector? Are they reaching out to you?

Jerry Spanakis

I wanted to talk about that because that is always the question we have as offshore academics. When we're looking for Ph.D. students, we can offer them a fixed amount of money, but then the companies are offering way more. So yes, that's always happening, but then the fact that we still have a strong academia in things like data science, and machine learning is hopeful that the same can happen in public administration.

Thibault Schrepel

Alright, so that’s the positive note I needed. Now, let me mention that if you are a computer scientist and you are interested in working for competition agencies, just be in touch with them. There are lots of advertisements in the space. So, I think the main message is not to be a dinosaur, because dinosaurs disappear. And the dinosaur, for me, is the one ignoring all our discussion right now. This ties also to the other question, which was unanswered, is there a pricing algorithm being used right now on the markets? The answer is yes and no. What is being used by companies are algorithms to better collude when they’ve decided on their own—without a computer—to collude. Tools can help companies define the optimum collusive price. However, there is zero antitrust case in which algorithms colluded “by themselves”. With that, I hope that I answered all the questions.

Thank you very much, Catalina and Jerry. Thank you also to all of you who stayed online so late in Europe. Tomorrow, same time, we will dive into different projects in which they used computational tools. They will be showing you how to use computational tools for better enforcement or to analyze the case law. There’ll be expanding the results, and showing you how to reproduce their methods, which I’m sure will be helpful to all of us. So, with that, this is the end of day one. I’ll see you tomorrow. Take care. Have a great night. Bye.

Day 2: Computational Antitrust Approach

Transcript by Teodora Groza

#1 – Doctrinal Implications of Computational Antitrust

Thibault Schrepel

Welcome to the second day of the computational antitrust conference. I thought I will be starting with a quick reminder as to what we discussed yesterday. If you were not with us, we discussed the end game of computational antitrust: what can be possibly achieved, the kinds of safeguards that we need, how to structure antitrust agencies, how to change procedural rules, etc.

Today, we discuss practical approaches to computational antitrust. As I said in my closing remarks yesterday, we are scheduled to dive into different projects that will show us how to use computational tools in the field of competition law. So, during the next three hours, you will learn about concrete results. But I think most importantly, you will learn about the methods that are already available in 2021, whether you work in a law firm, in-house, for an agency, or whether you are a researcher. To be very concrete and more specific, the first discussion is dedicated to the use of computational tools and visualization techniques to analyze thousands of antitrust cases. In the second discussion, we will explore the use of natural language processing (NLP) to analyze legal contributions and capture how market players understand the core concept of antitrust, such as gatekeepers, monopoly, etc. And finally, in two hours from now, the third and final discussion of the day will be dedicated to the use of machine learning to detect patterns in the case-law of the FTC which are invisible to the human eye. So, in a nutshell, today is dedicated to antitrust 3.0, in action.

Without further ado, let me introduce our first discussion. We welcome Felix Chang, who is a professor of law and the co-director of the Corporate Law Center and also the Associate Dean for the faculty and research at the University of Cincinnati College of Law. We also welcome Pinar Akman, who is a professor of competition law at Leeds University School of Law, and the director at the Jean Monnet Center of Excellence on Digital Governance. We are joined by Erin McCabe, who is working with Felix on computational antitrust at the Cincinnati College of Law. And finally, we are joined by Gijs van Dijck, who is a professor of private law at Maastricht University and the co-director of the Maastricht Law and Tech Lab. So, without further ado, Felix and Erin, the floor is yours.

Felix Chang

Hi, everyone. Hi, Thibault, thanks so much for bringing us into these conversations. Thanks to Alex Sotropa and Monique Norquist and others at Stanford for organizing this. I think it's really exciting to be at the forefront of this application of computational law techniques to antitrust.

So, Erin and I are going to present more of a methodology piece. This piece was recently published in the *Stanford Journal of Computational Antitrust*. It's a project that grew out of a collaboration that the College of Law at the University of Cincinnati has undertaken with the Digital Scholarship Center at UC, where we're using topic modeling. It's an algorithm that maps the statistical relationships among words in a corpus of documents. We've been funded by the Mellon Foundation by a large grant. And the Digital Scholarship Center (DSC) at the university is partnered with a number of academic units across the university to run topic models on a variety of research questions like pain indicators and medicine. I mean, on citation of authors and biology cases, and even the placement of bone in archaeological excavation sites. So, topic modeling can be done on a visual corpus as well. Our project, though, uses topic modeling algorithms on this platform that I'll show shortly that DSC has built to detect really latent patterns in a large body of antitrust case law.

Let me begin by framing what I think are, and have been up until now, some of the obstacles to big data research on antitrust, the antitrust texts. So first, for a long time, there was a lack of easily accessible legal texts. Until recently, it was really difficult to extract a large amount of case law. When we started this project a couple of years ago, we negotiated with Westlaw and Lexis, for like nine months for licenses to do bulk downloads of cases. This is a really onerous process. And in addition to the license negotiations, it just wasn't really convenient to download a lot of these cases. So that all changed in the last couple of years, because now we have the Caselaw Access Project. And that's this partnership with Harvard Law Library and the Berkman Center. They've digitized almost every published decision in almost every US jurisdiction up until 2018. So, as you can see here, that's about almost 7 million unique cases and over 1.8 million unique federal cases. And they make it really easy for scholars to download cases in bulk.

You can see playing around their website, there's this historical trends function down here. And I've already uploaded the search, where I'm looking for all US cases with the words antitrust and regulation. As you can see, you can run some rudimentary searches on the Caselaw Access Project. So that's been really helpful. What we've done is through the case law Access Project, we extracted 35,000 federal cases with the word 'antitrust' in it. So that's a pretty large corpus—35,000 federal cases with the word 'antitrust'. And then the research question that I had began with was how do federal courts balance antitrust and regulation, because that's what a lot of my prior work was on. So, then we downloaded a second corpus of legal texts from the Caselaw project, every federal case with the word 'regulation' in it. And that's 360,000 cases. This is a huge corpus. And later, we'll talk about how we went ahead and filtered down those corpora to more manageable corpora.

Now, I think the second problem is that a lot of big data work on legal texts is with modeling. It can be really difficult to translate models into visualizations that the audience can grasp intuitively, without a lot of statistical training. So, a lot of this work has been done in the subfield of computational legal analysis. And that's projects that explore questions like how and whether you can model legal reasoning at all. And that's the subject of a lot of papers on these series of online workshops at UVA that Mike Livermore runs. So, what Erin and I have tried to do

is to introduce topic modeling to antitrust audiences. Erin can talk about that a little bit more, but it's been around for a few decades. Again, it illustrates the probable distribution of terms and their co-occurrence within a data set.

So, what we've done is that the DSC has built this platform that uses different variations of topic modeling to show what the occurrence is, or what the spread looks like in these two corpora that we've extracted from the Caselaw Access Project, that is cases with the word 'antitrust' and cases with the word 'regulation.' We went ahead and we filtered down those cases a little bit more, though, so that now we're working with two bodies of text. One is every federal case with the words 'antitrust' and 'market power', and the other is every federal case with the words 'antitrust' and 'regulation'. Let me show you what DSC has devised. It's called a model of models visualization, and this is in the corpus of 'antitrust and regulation'. So, we further filter down to 2720 documents. The antitrust and regulation corpus itself is 7308 documents but you can filter down a little bit more. On the DSC platform, they have the ability to get us cases where the words 'antitrust and regulation' occur many more times in a case, so that they're not just random results. You can see that what the model of models view excels in really big-picture visualizations of how this corpus breaks down. I should talk probably a little bit about some of the variations of topic modeling that we've undertaken and why we're starting with a model of models visualization. So, one of the key criticisms of topic modeling for many years is that you can get these random aberrant results. What you're doing is that you're cutting and splicing cases across this really large corpora of texts. How do you know that? How do you know that these results are useful when they're decontextualized? One thing that the model of models visualizations tries to do is it tries to get us greater context.

The basic point of this kind of model of models or multi-level visualizations is that you're running up to six different iterations of topic models, and you're aggregating them, so that you see them in larger and smaller contexts. Usually, topic modeling gets you topics. But now what we are able to get is clusters of topics. And that's what you see in the right-hand panel. Again, each topic is defined by the most co-occurring words within that topic. And you can see some of those terms here. But what we've done is we've clustered those different topics so that they show up together. In this way, we get a little bit better macroscopic context to see how the topics come together. And what's interesting here is that when you look at the antitrust regulation corpus, you see that the topic clusters tend to cohere along sectors, industrial sectors.

Here, it looks like in this cluster here at the top, the green one, there is some sort of an antitrust immunity topic cluster. Then later in this purple cluster, we see some sort of a class action cluster. If we go on down the line, you see that fairly frequently. There's a healthcare cluster. There's a telecom cluster right here. There's a banking and financial cluster. There's some sort of a commission cluster. And many of these are old ICC cases. There's a patent cluster. And then there's another kind of evidentiary or procedural cluster. That does make some sense. It seems that when you're thinking about the balance between antitrust and regulation, some of the people who've done work in this area have said that it's really dependent upon the

sectors. So, it may very well mean that when federal courts are looking at competition concerns in a regulated industry, the way they wade through those competition concerns is really sector-specific.

All right, now, I'm just going to go through some more of these visualizations to show you some of the added improvements that the DSC has added. So, on the left-hand side, you see this panel here, and the panel gives you the ability to pull up the top relevant decisions within a topic cluster. So, you can highlight this cluster here, and then you might get the top decision in the United States versus visa. So, in this way, what we've also done is in addition to the macroscopic context, we also have a more microscopic context where you can actually go through and pull up each of the decisions, to read through those to see if the vocabulary that's used in the decisions really coheres with what you see as the top terms within each of the clusters, or each of the topics. In this multi-level visualization, there's also a couple of other views that may or may not be as useful. But this circle of views shows you certain terms and how they might cohere across the different clusters. The other thing that I think is a little bit more useful is what is you might have seen in a word-to-vector type visualization where you are situating the clusters against this vector space visualization. This is the network view, which can show you the distribution of these topic clusters as well. Now, I'm just going to show you a couple more different types of visualizations and then I'm going to backtrack and talk about some of the doctrinal generalizations we might have. And maybe some of the methodological criticisms and also the checks that Erin and I have undertaken to ensure that the results are robust. So, another variation of topic modeling that the platform has created is the topic browser view. And this is less of a macroscopic view, it doesn't aggregate the topics into different topic clusters. But what you can see, which I think is particularly helpful, is this little graph of the cases over time. So, you get more of a historical view of what types of topics might have declined as importance, at least as a proportion of the overall corpus of the text itself over time. So, what's really notable is that you have the rise of these class action topics over time, you also will tend to see the decline over time of a lot of topics where you have the top cases with terms like 'Commission', in particular, the Interstate Commerce Commission, and this does seem to cohere with some of the earlier work that scholars in antitrust have done. More about deregulation: it's more of an imprecise label. But the concept of deregulation is moving from rate regulation and more into a regulatory regime where industry regulators are just trying to set baseline rules for competition. So, this is the antitrust regulation corpus very quickly. Then this is what the antitrust market power corpus looks like. This is the antitrust market power corpus. What I think is interesting over time is that you see the decline in the importance of the tying cluster. And that might be for doctrinal as well as theoretical reasons. Doctrinal reason might be that over time it has been much more difficult to bring tying cases. This is a Stack View, which shows much more easily the relative importance of each topic in proportion to the cluster over time, where you see this real increase over time in the importance of class action topics, and also topics that pertain to antitrust litigation in evidentiary matters.

Okay, so let me talk for maybe a couple of more minutes about some of the big picture, some of the doctrinal or results, and then I'll give it over to Erin to talk about some of the robustness checks that we've undertaken. Some of the less

surprising results we saw over time are the fact that Interstate Commerce Commission cases, the real rate regulation cases, have declined over time. It's also not entirely surprising that we find clusters of topics grouped by theme. For instance, you've got patent clusters, healthcare clusters, banking and securities, labor class action clusters. You know, here one of the surprises was that any way we ran the models, this little cluster of milk cases appeared and that seemed really odd to me. I wasn't really aware of the milk powder cases. But apparently, in past decades, there was a lot of milk powder antitrust litigation that went on.

What about relevance? The way topic modeling defines relevance is really different than how we might think about the term relevance doctrinally or traditionally the way we do Lexis and Westlaw research. Remember that if a topic is comprised of terms that are most statistically likely to appear, then the top decisions are going to be those that are statistically most likely to reflect these terms. So, it's a statistical gauge of relevance that's different from the legal understanding. And the top cases that we find are less ones that have precedential value but probably more statistical value. Here, I think I'm going to pause, stop, share, and let Erin talk about some of the methodology and robustness checks.

Erin McCabe

Sounds good. Thank you, Felix. Felix covered in his section most of the conversation discussing the grants. I want to talk about what the algorithm is actually doing rather than getting into the formulaic composition of the algorithm. Essentially, we're measuring or assuming that words more frequently co-occurring across a higher number of docs in your dataset are a part of the same topic. But you must keep in mind that topics are lists of words and not labels and that each document has a distribution of those topics. So, it's not one topic in one document. Back to the original coherence metrics that Felix had talked about, or led into this, we test models at the DSC with kind of a dual approach. So, like Felix had mentioned, there's a lot of human judgment in these models. And we incorporate that subject matter expertise at two points in our analysis of these models. There's the actual analysis of the models itself, but there's also the evaluations of the model's sense-making ability, before we really get into the meat of what the model is telling us. We must assess whether it's doing its job in a sense-making fashion. So, this is kind of our first step. It's just a human eye.

Felix and myself come from very different realms. I work in NLP research and law, and Felix is in law, and that combined subject matter expertise allows for two different eyeballed perspectives on our models. So, we look for things like, do the topics have an inherent sense-making capacity? Look at the example I gave here, cable-licensed television programming. There's a certain inherent sense to those words belonging together and to a topic, so we can kind of go through the topics and check for that. We'll also look at top terms and topics. We use kind of a mantra of 'affirm and complicate': you want a certain number of your expectations of the data set to be met. And then maybe, later on, add a nuance to that, but you want to meet a certain number of expectations. And then we also look for things like meta language, things like plaintiffs, and defendants that occur kind of everywhere. And

then we kind of consider that language that’s endemic to your specific data set. So, we look for that and decide, does that prove our model is working? Should we remove those first step words, things of that nature? So that’s the human judgment element, that can be a little more subjective.

To combat any subjectivity bias, we also use coherence testing in a more algorithmic approach. And coherence, in this sense, is a metric that scores each individual topic by its degree of semantic similarity between the highest probability words in that topic. These are tests that we run across an entire data set. And we adjust the different hyperparameters of our algorithm across all of these tests. So, the LDA algorithm that I had mentioned earlier that we use for this has two different hyperparameters that control how many topics can be in a document and how many words of that dataset can be in a topic. So, we can run this multiple times across the same data set and across models at a different topic count. So, 10 topics in a model or 15, or 20, and then we decide what are the best hyper parameters in to get the highest scoring coherence metric there. For Felix’s data set, we found that the range of coherence metrics was actually pretty tight. And it seems that case law as a data set, compared to other datasets we work with at the DSC, like news media text, or even academic journals, is a lot higher coherence scoring across the board. And I think that really speaks to the linguistic hyperstructure of case law.

So yeah, that helps us pick the best hyperparameters for a data set. So we have a stronger, more coherent model in an algorithmic way, and also based on our subject matter expertise. We’ll leave it at that.

Thibault Schrepel

Thank you very much right. I just want to make it clear to everyone that should you have questions, please enter them in the Q&A. In the meantime, Pinar, the floor is yours.

Pinar Akman

Thank you, Thibault, and many thanks to Professor Chang and Miss McCabe for their presentation. Well, where do I start? I should probably start by saying that I’m a mere antitrust lawyer. And so, there is not a lot I can say on the modeling. And I think that’s been explained extremely well today to us, having read the papers. So, the thoughts I’m going to offer to revolve mostly around what we, as antitrust lawyers, can learn from the findings. I’m probably going to have more questions than comments myself. If I may, I’d like to start by summarizing in my own words and in the author’s words from the paper, what the paper achieves. And the authors should absolutely feel free to correct me if I’m wrong when I say anything about the paper.

I’m doing this in the hope that this might help others like me, who are not familiar with these models, to understand, almost in layman’s terms, what the paper is doing. So, the paper uses so-called topic modeling, which is a form of natural language processing that depicts the probability distribution of certain terms over a body of texts. And here, the body of texts is case law. And topic modeling also

shows essentially how words cluster around one another. And therefore, it's used to unearth patterns in the body of texts that it's being applied to. In that context, clusters show us the relationship between different topics. The topics show us the relationship between terms that are being sought for in this body of texts. The paper here aims to contribute to building the empirical foundations of possible reforms in antitrust by essentially offering us tools with which we can assess how federal courts approach certain, in this case, doctrinal questions like market power and the relationship between regulation and antitrust. To me as an antitrust lawyer, this is the most exciting bit because I myself have tried on some several occasions to do empirical research myself. And I see this as the main missing piece of the puzzle in almost every aspect of antitrust, that we don't have enough empirics to work with. And therefore, I welcome and applaud the authors for actually providing us with this venue of research, which I'm sure others will be able to build on as well.

So as has been mentioned during the presentation, there's a very large database of over 300,000 federal cases which deal with regulation and 36,000 which deal with antitrust. The authors have used elimination to reduce this to a shorter group of cases. And what the models that are being used do is to provide us with visualizations of cases, grouped by certain recurring terms, and then that allows us to see the relationship between terms and relationships amongst those cases, and there are three different types of visualizations used, one with multi-level visualization, where we have seen the tree and the circle and network format, the topic browser visualizations, and the Python based LDA visualizations that show the distance or the proximity between different topics. And these visualizations enable us to more easily interpret the machine analysis of essentially what the algorithm finds. What the authors seem to find is, as has been mentioned, that regarding market power, certain topics such as telecoms, health, health cases, tying patents, technology, and so on are coming up. And in the antitrust regulation corpus, we have terms and topics, such as immunity, telecoms, energy banking, health cases, and Interstate Commerce Commission featuring more highly than others.

Thinking of what we have heard and having read the paper, I have two, types of questions to the authors - and those are really my comments. The first one revolves around what was mentioned at the beginning, namely that of decontextualization, which was said to be one of the criticisms of topic modeling that comes from digital humanities. And again, I'm speaking here as a lawyer and not as an expert in the method. And as has also been mentioned by Professor Chang in the beginning, this is really a methodology paper rather than a paper that presents the results in terms of the substance. And I note that there's a separate paper which I'd be very interested to read as well, in the context of decontextualization. The question I have is whether contexts merely mean which cases are most likely to feature the terms that constitute a topic for the purposes of the research. And I guess that by being a lawyer, to me context means a lot more than what it specifically means in this modeling context. And I wonder what these models, what this strand of research, or what machine learning can tell us about the broader context, as we lawyers will understand the context to a given case. So, how much of the context that's external

to the subject matter at hand could we detect by using algorithms or machine learning, if at all? And I don't know this. That's why this is the genuine question in that sense. The research is absolutely fantastic in terms of providing us with a venue that would allow us to observe trends over time and case law that will allow us to categorize terms and see which categories of cases appear more often. We can also see when they do appear and when they do not appear. It gives us, I think, a very potentially coherent bird's eye view of a body of case law. My question, again, as an antitrust lawyer is, well, what else can this research do? And particularly in terms of substance? For example, can this type of research be used to make predictions in terms of possible outcomes? Can it be used to make predictions in terms of the likelihood of success?

I have some more general questions as well. So, for instance, who will be the most likely beneficiaries, in your opinion, of this type of research? And how do you think your model could be used in other jurisdictions? What would those other jurisdictions need to have to be able to use, for example, your topic modeling, and to be able to, I suppose, transpose your study into their own jurisdiction? And one final question, I suppose, what do we learn or what have you learned in terms of the substance? So, we have seen really the methodology in this paper. And as I said, I know that you have another paper in which you explain, or you present the results on the substance. But as a lawyer, to me, it's also very interesting to hear from you. If you don't mind doing that, tell us about what you have found in terms of the substance once you have applied your methodology to the large databases of cases that you have. And I'm sorry, I have more questions than comments, but I hope they might be helpful. Thank you.

Thibault Schrepel

Thank you very much, Pinar. Gijs, the floor is yours to ask even more questions, to be sure.

Gijs van Dijck

I'll do my best. So thanks a lot for this invitation. And also, thanks a lot for the presentation. This is really wonderful work. So, when asked to comment on this, and I tried to pull up the paper turned out to the PDF file was corrupted. So instead of asking what the content was, I started looking online, and I encountered, I think, a previous draft, which is much longer, a fact which allows it to provide more context. So, I'm basically placing my comments based on that, but it doesn't matter; the results will be the same.

In the bigger picture that you presented in that paper that I read, based on the large pool of these antitrust cases, you try to perform an algorithmic analysis aimed to basically answer to doctrinal questions related to the measure of market power and the balance between antitrust and regulation. I think that's very interesting. I'll tell you why. It's because a lot of these methods, LDA methods, but also the computational analysis of law, kind of float in the air sometimes. So, the idea of bridging the gap between computational analysis and legal questions or legal analysis is a great achievement. Overall, I think this is really a nice approach. We

should definitely have more of these and especially because I think you include a very honest assessment of what is possible, but also what is not possible, and what are the challenges and I think that is what we really need in this.

Let's look at the findings. Basically, information retrieval is probably the best use of topic modeling at this point, not so much for prediction, but to find the relevant—I very much agree with that. And also to suggest areas of overlooked scholarship or doctrine. I have a very bad cough, so I sometimes have a cough break; I hope you don't mind. I think you have more added value to the present. The added value of automation should not be underestimated, even if your topic modeling results do not provide results that the doctrinal scholars would find optimal. You have to keep in mind that if you do this search, and you have a sort of a pre-trained model, or at least you have set parameters, you can automate this process quickly. And even though you don't have optimal results, you still have relevant results that you get much quicker than if you have to read the whole picture. And I think the visualization of relationships and trends is extremely powerful. And if you read the paper, and I would recommend everybody to do that, then you actually see how powerful these findings are.

I have some questions. I'll start with two minor ones concerning the selection effects, and we're going to talk about which cases are included or not included. Those are common, the 'usual suspects' questions that I'm fascinated about and to which I don't have an answer myself. In our lab, we conduct these similar types of analyses. What are we actually studying? We're studying cases, and cases are anomalies by default, depending on the field of law that you're working. So, in a lot of areas of law, the ordinary cases will become cases because they're not special enough, and they can easily be decided based on existing case law or existing rules. So, where on the spectrum does antitrust law and these cases find themselves? Are all antitrust cases being brought to court individually? Or not always? And what does that mean for the interpretation of the results?

And maybe, to make it more concrete, when we talk about precedents, for instance, you see in these various niche topics, rivers that show to what extent topics become more prevalent, or less prevalent over time, overall, are compared to the other topics. There are also different ways of presenting data, of course. But the question here is, what are we actually seeing? Are we seeing a trend over time of indeed the topics that are becoming more interesting or more important and more relevant? Or are we also seeing an effect of the cases that are just decided, and then they're gone? There's no way that computational analysis can decide that, and I don't think there's any way legal analysis can decide that. But I think it opens up the door of asking questions, posing hypotheses. To kind of put these marks on this timeline, like in 1984, this decision occurred. And you see, after that the questions actually diminished. So, what could be going on and what effects you couldn't observe and then test it against the data? That's also the question of independent observations. Basically, if there's a precedent, you're going to use the previous case to decide a subsequent case. What does that mean for a selection of cases that are in your data set? And just out of curiosity, I'm curious to what extent these results are fair. And I'm not talking about fair in a legal sense, but fair in a computational sense. To what

extent are these results, the software and the methods Findable, Accessible, Interoperable and Reusable? In the paper, you refer to some of these software packages, and also even platforms that you can use. It would be great to learn as a community more how to reuse these elements because what you're doing here in the field of antitrust can also be put into practice in tort law, contract law, or whatever field you're talking about. So, learning that and sharing that not only as an individual project, but also beyond the thing could be very valuable.

My main point is, I guess, again, that you're selling yourself a bit short. And I think that has to do with doctrinal research, which you hardly define in your paper. You seem to connect it somehow to precedents and semantic similarity. But I think you can flesh that out more. So, if you look at how the final scholarship is defined in the current research, it's defined in very different ways - systematic exposition, analysis of consistency between rules, explanations, future developments. If you narrow that down, you can actually look at, for instance, case law analysis as already a field where we can pose different types of questions. You can look at precedents; you can see what the relevant precedents are, to see if the laws apply coherently. But you can also look at organizing and reorganizing case law in terms of factual similarity, so given a certain precedent or given a certain rule. To what extent are similar cases decided similarly? Then you have to basically look at to the facts and the outcome.

I think you can really play with the facts and with the legal reasoning and arguments and then with outcomes as well. I mean, you can do that for other parts as well. I already said about the policy analysis proposition testing that you can formulate a certain hypothesis that has been forwarded by the literature, or maybe something that you think is going on, or experts thinks is going on, and you can test it against the data. And I think it becomes more robust and more systematic if you do that. And you can actually improve academia with that, by showing that the possibilities are much bigger than that. I think one element to look for is to retrieve legal precedents, in which you could leverage the citations. You mentioned in the paper that I read that you can leverage that, also compared to and by using semantic similarity. But in order to do that, you have to look at the facts, the arguments, the outcomes, and you have to peruse that. What is problematic often is that these legal decisions are often badly structured for machine learning. So, you have to do that yourself. We are trying that in our group for European cases. It's a lot of work. It's not very rewarding. But on the other hand, once somebody has done that, it's going to become extremely valuable. And then you can actually start doing what I was just mentioning - you can basically say, if I have a given decision, can I look for a similar decision based on the facts? Can I look for a similar decision that use similar legal concepts, and then can I compare it against the outcomes? And if you can, then combine it with legal citations, and then maybe we're able to leverage that information altogether. Then you could even have a slider where you go to the citation weights or to the semantic similarity weights. But at least that allows us to play around with the data and fine tune the results and gear them towards the questions that we're interested in. And then we can basically apply topic modeling, zoom in, etc. with these different layers, as you presented. I think it will open a world of new answers to the same final questions. That was it for me.

Thibault Schrepel

Thank you very much Gijs, I've never seen someone mastering the ability to mute oneself right before coughing—you really do that very well [laugh]. I guess it's fair to give you a few minutes, Felix and Erin, if you have anything to answer as I suppose this might be the case. I've done my best to cluster all the questions in the chat and some of the questions that Pinar has been asking into three big questions. So hopefully, we can all do all that in 15 minutes.

Felix Chang

Okay, so maybe I can start by responding to some of the questions that were raised by Pinar and Gijs. Thank you so much for reading the draft. Gijs in particular, thank you for looking it up on SSRN. I know that the PDF was corrupted; we realized that after publication. And there are a lot of questions in the chat and Q&A. Some of those are methodological, I can defer them to Erin, I know she's responded to some of those.

So maybe some of the bigger picture questions that Pinar had raised to begin with. What is the broader understanding of context? Can it be used to make predictions? What are the beneficiaries, and what have we learned in terms of substance? You're right that in terms of context, we're thinking about it, restructuring context, giving more context algorithmically. The criticism of topic modeling was, and it applies to a lot of these computational methods, is that you're slicing and dicing words, you're rearranging them. And that's devoid of context. And that has been problematic in the field where we are used traditionally to close readings of text. That's what legal analysis is. What we've tried to do, for instance, with a multi-level visualization, is to give us both larger and smaller contexts. You can actually look at the cases and begin to read those to do some of the backend checks that she suggested regarding reasoning. Are they coherent across a set of cases within this topic? Are the facts consistent? We're noticing at the top terms there is some sort of consistency. You see the three different ways in which the multi-level visualization shows you these topics in clusters to give you a larger, more macroscopic context. But again, you're right, that is largely more algorithmic, although sometimes that does cohere with a doctrinal understanding when you see these topics that seem to align with our understanding of the case law. Maybe the case law does splinter along industries. Maybe there is a decline over time in rate regulation cases or a decline over time in tying cases. Can it be used to make predictions? Probably yes, we really steered clear from there. I mean, a lot of this stuff is unstructured. And we really try to open up the black box through coherence testing.

Other things that can be done is what they do in data science: they sometimes take random samplings just to run topics across these random samplings to see if the results hold. Until we can do that, with a lot of confidence, we're really hesitant to make any predictions, regarding for instance who are the most likely beneficiaries. I think that it definitely benefits researchers. I assume industry can probably work and invest in the robustness of these methods much more quickly than we can, and they can probably take it to a level where they can make predictions, and I'm sure

they are, making predictions already. But what we're interested in is offering a service to academic researchers. What we think is particularly useful at the end of the day, something we haven't talked about yet is that in being as transparent with the methods as possible, we show that searches don't always return the same results, even across commercial databases that are using whatever their proprietary algorithms are. Lexis and Westlaw are notoriously opaque about what their algorithms are. And our hope is that we're trying to push them towards transparency as well. And all the research questions that you had suggested, Gijs, are exactly what we hope to do at the next step. It is a lot of work. A lot of the computational legal analysis, a lot of this work is kind of tedious. You've got to read the results. But it does have to be done. I think maybe I'll pause here again, just say thank you, again, to both of you for such close readings and such helpful comments. I'll connect with each of you individually.

Erin, I think there were a set of methodological questions that I will defer to you on. I'm not sure if the participants can actually see your answers. I think they are just for the panelists. So... shoot.

Erin McCabe

Jack, you had a question about clustering algorithms. And it's a good one, because we do run six LDA topic models, and then cluster them to create those topic clusters that Felix was talking about. K-means clustering is the algorithm. So, we run six LDA topic models, that's one algorithm. And then k means clustering to create those topic clusters. That's kind of part two. We had previously used the affinity propagation clustering algorithm. And the results were not as visually cohesive, and also just had a long statistical tail of one topic clusters of hyper specialized language that really didn't give us a lot of actual meaning. And K-means clustering proved much more effective to all our goals, and we didn't lose anything important in that transition, either. Patricia Villa had a question about labels. So, topic modeling falls under a term category. Unsupervised machine learning, as opposed to supervised machine learning, which would require labeled data. The label would say, this document is antitrust, and it would do that 100 times. And so, the next time it comes across a case that has a similar look, it could say this is an antitrust thing. But the LDA topic models' goal is really about that latent relationship. Topics are lists of words that frequently co-occur, that is in the algorithm itself, built from the data that we're giving it. So, there is no training data set that the machine is learning on. Its learning while it's going through the documents themselves and building those statistical relationships of word co-occurrences. So, there is no pre-labeling data. We're really looking for that latent relationship. Let me know if I'm missing any other ones.

Thibault Schrepel

Those questions were related to the type of algorithms. So, I guess you answered them all. Yeah, indeed, another one, which disappeared, was very technical. I couldn't understand the question. And so maybe you will retrieve it. But let me move on to the substantial parts. I'm wondering, and that's one of the questions that Pinar asked and Julianos as well, if you think that's what you've done is solid

enough to be used for policy guidance? Do we need more cases or can we already play with the results?

Felix Chang

At this point, I'm not sure that it's solid enough. And one of the things that I was thinking about that recently I attended this conference where Doug Melamed had asked the question of whether, since antitrust law is just all over the place, maybe what we must do is just streamline it so that we get back to the basics. And he had had two propositions for what we really need to do. When we were looking at the cases, I took another look, again, with a view toward asking, are there trends that we can observe over time that can cohere, that can show us whether or not we can get to that endpoint, which is to clarify antitrust doctrine or not. In the more concrete example of market power, I had thought that the decisions would be all over the place. And I think to some extent they are, but we do see, at least in the way of the structuring of the topic, some sort of coherence where we do see topics that are breaking out, where you have industry-specific topics, where you have rises and declines over time with these topics. A point well taken earlier from some of the questions is whether that just meant that there were more decisions, and they were just decided without very deep analysis. I think all that is to say that there are still a lot of questions that we have to sort through right now. And I don't think we're at a point where we can offer that much policy guidance right now. Maybe eventually, we can get there. But there's just so much here - this is an area that's replete with questions. I think we just have to explore a lot more to get more comfortable before we can get to the policy guidance.

Thibault Schrepel

I was surprised by the trend for tying cases: they are declining, when in fact, we thought that this was the number one practice, and it still might be. But if you're working at the FTC, it might be that it's declining, because you decided 10 years ago to focus on other practices or because there is a decline in such practices being implemented. So, you need also to know the kind of policy you've been implementing on the inside. And that relates to another question that Faith was asking: do you think that this type of methodology could be used throughout the world? To be very specific, what about an agency with just a few cases? We know there are some antitrust agencies which were just created a few years ago. What would be the number of cases and data you need for your analysis?

Felix Chang

Yeah, so let me take the first part first. Now, when DSC built this platform, it was on some other packages that other computer scientists and digital humanists had created, like maybe Andrew Goldman. So, some of those visualizations are taken from those platforms. It might seem that the barriers to entry are rather low. But you still have to have money to be able to create a platform, you still have to have the technical know-how to be able to apply those packages, even if they're free and open source. So, when we had started, I think we were much more, at least I was

much more optimistic about the democratic nature of this technology in particular, now that you can get cases pretty easily through the case law access project. But then, as I started thinking about it some more, my personal view is that it's not - there are some barriers to entry, it does take some money, it does take some technical expertise. So, it's not like in any advocacy group or that consumer advocates group can easily replicate this stuff. And also, I'm particularly conscious of what this means for university research, because this is funded by a grant. So initially, the Mellon Foundation had given DSC a million dollars, and then they realized that's a lot of money. And so, I've been also concerned about what this means for the future of research. In particular, if law gets away from sort of his humanistic grounding and more toward computational, but these are all questions for another time. And then the second part of your question to what the sample of cases is, I can defer to Erin on this as well. And you can chime in. But I think that this tends to work with larger bodies of cases, I don't find it particularly helpful with a small sample of cases. So, I'll give you an example. I had run some, some visualizations early on for cases with the word Trinko. On the balance between antitrust and regulation, we got maybe a couple of 100 cases. I just didn't find that particularly helpful.

Thibault Schrepel

Very interesting. Erin, I will give you the final words, and then we go to Gijs. One question we had in the chat relates to the ability to use your technique to enhance search engines such as they exist. So, I'm curious as to whether you work with the private sector, if Westlaw and LexisNexis got in touch with you. What will be the commercial application of your work?

Erin McCabe

Previously, we've worked with patent law offices to explore commercial applications to patent searching with topic modeling. But I don't believe anything came of that. I think there's still something of a skillset gap because there is this analysis component that not everyone's prepared for. And then to the document count kind of aspect of it, I think Felix is spot on; when you're in the realm of a couple of 100s documents, it's better to pick a different algorithm, really, instead of topic modeling. We go down to a word level. And oftentimes, we'll use things like word-to-vector, which is more at the word level relationship rather than document level. So, when I get into smaller realms, I kind of tend to go towards that space to help us a little better with our analysis. But, Felix, I don't know; if you've had other interactions with commercial applications or anything, I assume LexisNexis wouldn't be keen on it.

Thibault Schrepel

Alright, Gijs. You have a minute, like in a presidential debate, to convince people to vote for you or to conclude.

Gijs van Dijck

Nobody will ever vote for me. But I'll give my two cents on this. So, I think the implicit remark of pooling resources is very important. These projects are project-based when the funding runs out, and it's difficult to maintain the software, whereas it's actually valuable for other projects as well. So, I think there goes a long way to include the community and to build a community for this topic, but for a computational analysis of law in general. And to kind of see like how we can contribute to each other's software, each other's packages, so that most of it stays alive beyond the death of or at the end of the individual project.

Thibault Schrepel

All right. Thank you very much. Pinar, the last word is yours.

Pinar Akman

Thank you, just to pick out what you said and bring that together with Gijs said, I think context seems to be really important in the sense that we are looking at a sample of reality, in essence, that is the outcome of certain parties having made the decision to take your particular case to court. And even there, there will be differences between jurisdictions where you have to go to court in an antitrust enforcement system, as opposed to where you don't have to go to court to prove an infringement and so on. So, contexts and whether, for example, the authority decided that, okay, we haven't done any tying cases recently, let's do some tying cases or vice versa. I think those things are potentially really important when it comes to being able to use this model to make predictions and so on. And the other thing is, I'm really intrigued by this statement around Westlaw and LexisNexis. And if they had been regular platforms, we might have thought maybe they're engaged in self-favoring because they'd rather see these other results. But I'm sort of almost a bit disturbed by why they're top result when you search tying, and market power is so different to what you're finding. I think this is probably the topic for a different conversation. But I find that really interesting. And I'd like to perhaps know more about that on another occasion, and others just if this is the final word, I'd like to congratulate the authors again, this is really as cutting edge as it gets, I think, and I'm sure this will lead to many more other fruitful building upon this particular research and different papers and so on. So, congratulations. And thank you.

Thibault Schrepel

Thanks Pinar, Gijs, Erin, and Felix, thank you very much. Just a quick note on WestLaw and Lexis, if they were indeed sanctioned for doing self some self preferencing, they will never publish the decision, and so... we'll never know. I suppose [laugh]. In any case, thank you very much for the fruitful conversation.

#2 – A Computational Analysis of the DSA and the DMA

Thibault Schrepel

It's time for us to move on to the second panel. I'm very pleased to leave the floor to Aleksandra. Aleksandra is one of our editors in the computational antitrust project. She's a Sciences Po and Columbia Law graduate and she will be moderating the second panel dedicated to the computational analysis of the DSA and the DMA, Digital Services Act and Digital Markets Act. So just give us a minute to bring all the participants. In the meantime, thank you all so very much for all your questions. Should you have some more questions feel free to send our panelist an email, or use the Q&A. Alright, Aleksandra, the floor is yours.

Aleksandra Wierzbicka

Thank you, Thibault. Thank you for the brief introduction. And thank you also for the opportunity to moderate today's panel. Our discussion indeed will be on the computational analysis of the Digital Services Act and the Digital Markets Act, the twin proposed package regulating digital platforms. Our panel will focus on leveraging the power of computational tools, and more specifically, the algorithmic technique of natural language processing, in order to explore how different stakeholders understand terms such as 'gatekeepers' or 'self preferencing'. And I'm very happy to welcome today's speakers Fabiana Di Porto, as well as Daniel Crane. Fabiana Di Porto is Professor of Law and Technology at the University of Southern Italy. Professor Di Porto presides the funded research project on algorithmic disclosure regulation. She works with Tatjana Grote. And on the other hand, Daniel Crane is Frederick Paul Furth senior Professor of Law at the University of Michigan. Professor Crane served as the Associate Dean for faculty and research from 2013 to 2016. For the next 20 minutes, we will listen to the presentation of Fabiana Di Porto. And in the next 15 minutes we'll hear from Daniel Crane. Please don't hesitate to ask questions in the Q&A section. And without further ado, I'm giving the floor to Professor Fabiana Di Porto for her presentation.

Fabiana Di Porto

Good evening, everyone. I wish to express my gratitude to Stanford Law School and the CodeX Center, and the Stanford Computational Antitrust project. I want to express my gratitude also to my co-author who could not be here but who also thanks everyone.

So, let's start. Back in 2020, Executive Vice President Margarethe Vestager presented the DSA and DMA. She highlighted that it was complex but that they were looking forward to it. Of course, the breakthrough reform Commissioner Vestager was referring to was really a major breakthrough reform because it was shifting all the antitrust intervention from an ex-post enforcement towards an ex-ante regulation, and it was meant to be really a blueprint for regulation of digital tech technology around the globe, in the US and China as well. Such a shift is profound and pertains both to the goals, the tools and the terminology of competition law. According to the Commission, this shift is powered by a growing

consensus that traditional antitrust tools are inadequate to tackle the competition problems of the years to come, and therefore, an ex-ante regulatory tool is much better and more appropriate. So, when looking at the DSA and DMA proposals and the long debates that surrounded their drafting, one will quickly notice one thing. Many key questions, for instance, how to define the scope of application of the new rules, involved some relatively new terms like 'gatekeepers'. There are also some anti-competitive practices like self-preferencing, or remedies like blacklists, that are really not yet part of the traditional competition law vocabulary. Hence, we ask specifically, are all stakeholders sharing the same understanding and, therefore the same use of the relevant terms and concepts of the DSA and DMA? Or can we identify different attitudes towards these issues? And to do so and to answer these questions, we decided to employ computational tools.

So how can computational tools contribute to this? Actually, there are several advantages arising from the use of computational tools in the legal analysis and antitrust law specifically, a lot of them have already been outlined by the prestigious speakers who presented before me. I will only delve into two of them in this presentation. The first one is velocity. Computers can of course process documents much faster than humans, and sometimes within a fraction of seconds. For instance, if we must read 1000s of feedback documents to a consultation, we could have it done by algorithms. Hence, algorithms allowed to scale up our analysis. The algorithm can have it done not only faster, but much deeper. Of course, this reading would be different from one made by a human being. And for that reason, that would require human intervention. That human intervention would be needed for selecting the input for manually instructing the machine. The second reason is the ability to discern hidden patterns. And that humans would not see with close reading. Because of the novelty of the many terms used in the DSA and DMA and the issues surrounding digital competition, there is no established common understanding of a large part of the debate on the DSA and the DMA. There was actually a lot of debate that was centered around how to define the relevant terms like gatekeepers, self-preferencing, but also established terms that were new not in their understanding, per se, but in the way that they were to be applied to digital markets. The Commission concluded that all stakeholders were demanding new rules for the gatekeepers, which explains its prohibited anti-competitive practices. This sounded like a consensus at first, however, how could we know that all participants are having the same platforms and practices in mind? Especially when the key terms they are using, like, for instance, gate keepers are themselves still up to debate.

As I was mentioning before, even terms that were more consolidated, like, for instance, dominant position, or even dominant firms, were still not so clear in the sense of how they could be applied to digital markets. So therefore, as we shall see this was also reflected in the feedback documents to the European Commission's public consultation. So, what I am telling you is basically that the feedback documents are a source of an extremely valuable source of knowledge. In order to understand how those affected by those new rules, from individuals to big platforms, perceive the problems addressed, and the solutions proposed by the European Commission, simply reading these feedback documents might, however,

not help to actually develop a precise understanding of the stakeholders' take on certain critical questions. And therefore, we might not be sure whether the competitive issue at stake is adequately tackled. Consider that many tools depend on the stakeholders' behavior to be enforced. This is where algorithms enter the scene. Algorithms can help find some underlining differences in understanding that are not visible to the human eye.

So, all we can see are our words, but not necessarily the way they are understood and used by the drafter. We used NLP, which has already been mentioned, and especially word embedding models. Word embedding allowed us to express generally speaking what are the semantic relations between words in numbers, and hence to quantify them. It allows us to compare the semantic relationships between words for two different collections of documents, and therefore we used the unsupervised word embedding alignment. Only through this technique we can meaningfully compare embeddings that were trained on different documents. To see whether the differences we found were not only random noise, but we also then employed classical statistical testing. Lastly, to better understand the potential reasons for the differences we found we use those for sentiment analysis. As I will explain better in a minute, each computational tool has a precise purpose.

Let's have a closer look. We first scraped all the relevant documents, and we cleaned them all not to have noise and then clustered them into bigger groups. We combine the commission categorization scheme, and I continue with the analysis of which documents should be clustered together to do so. We then train the neural network on the documents for our cluster. Based on the number of times words appear next to each other, the network is allowed to calculate a vector for each word in each corpus. This is the so-called 'word embedding model.' Just to give an example, these models are really remarkable in the sense that they can capture the semantic meaning of words in a set of numbers. For instance, if we have a model, a well-trained model, the distance between the vector of the words Paris and France will be more or less the same as between Rome and Italy. Hence, the relative positions of these two vectors will roughly represent the meaning of certain terms. This means that while a simple algorithm would require that researchers formulate explicit rules to approximate the semantic meanings of words, machine learning or this neural network would just train the neural network on data and learn this relationship automatically. And therefore, we will not need to make this relationship-specific and explicit, but just to train the neural network. To make the model, we trained on different corpuses comparable with the use of vector space alignment. The vector space alignment, just to make it very simple, allows us to bring the vectors that we train on two different corpuses together on just one model space. And this allowed us to make them comparable. And it's like if you project it in a space in a 3D space. But in order to make it visible for you, in this presentation, we just projected on a 2D space model. And I will show you in a minute. And yet, we still needed to ascertain that these differences that we noticed were not just merely incidental but were actually of certain significance. And therefore, to do so we just employed the statistical test. So, we had this semantic distance, this semantic difference, and we had to establish that this was a difference in meaning and in the employment of the terms. We used a controlled vocabulary to get an understanding of what a non-semantic or random difference would look like. Since

we found significant differences for more than 1000 words, most of which were not really interesting for our analysis, because we wanted to focus just on some very specific terms, we had to manually code what were the results of our analysis.

Therefore, we have to select what were the terms. We focused on the documents that all the stakeholders, individuals, businesses, NGOs, governmental agencies, lobby groups, submitted to the open consultations of the three impact assessment documents that the European Commission published, and that later became the DSA and the DMA. We collected and analyzed almost 3000 feedback documents and replies to the questionnaires. We then collected them into three corpuses. So, the three corpuses you can see on the slide became our three groups. They are made up of individuals and micro-organizations, the small organizations, and then the medium and big organizations. The corpus is made of what is commonly known as the big text.

Let's get to the results. We had we collected the results, dividing them into three groups, then we had the first group of results, the actors, and then the terms relating to actors. Then we have the terms related to the conduct side, and then finally, terms relating to what to do. The majors the first result is on the question of how the gatekeepers should be defined, that was hotly debated. And our results actually confirm the results of the European Commission consultation. In fact, in all the three corpuses as you can see, there is a difference in the use of this term by all the three groups, and moving on to the terms relative to the conduct, something which is quite interesting is that there is a statistically significant distance between the vector for self-preferencing for the corpuses, A and B as well as for B and C. While the corpus is A-C, which are the individuals and big tech, the term self-favoring is used differently, hinting that it is the general concept behind those terms that seems to be still elusive. This is very interesting to me, because if one knows that the DMA in its latest version that has been released on November 26, 2021, it only literally prohibits self-favoring and not self-preferencing. Now, in corpus A self-preferencing is most closely related to monopolizing, suggesting that the anti-competitive outcomes of self-preferencing are a key concern for individuals and micro-organizations, while small companies on the other hand, associate this practice with over-regulation and reports. This is interesting not only with the view to the prohibition of self-preferencing in Art. 6 of the DMA, but also in light of the comprehensive reporting duties are on ranking in the DSA. Our results could be interpreted, but this is of course, very tentative as a hint that the small companies and organizations might fear comprehensive transparency and reporting duties, and therefore highlighting the need for a difference in approach. Let's get to approach the conclusions after having discussed the actors and their conduct.

So, the next step would be what shall we do? What to do concerning the ex-ante measures? An interesting finding is the blacklist. We find that there is a significant difference in the way stakeholders use the word blacklist and data sharing. Concerning blacklist, the micro-entities and medium-big organizations use the term very differently. And in this disparity, there is a strange association with the terms on the side of larger entities, there is an association with the term problem. While this interpretation is, of course, highly tentative, and would need to be

confirmed by a qualitative analysis, the second closely related associated term is dominance, which is a little bit more telling, and highly interesting.

This is my conclusion, as I was telling, we actually can tell a lot about these differences. We found a lot of significant differences in the understanding for many central terms. However, we cannot explain why those differences exist. So those hidden misunderstandings can of course, be exposed thanks to our algorithms, and that means that we could help the Commission as well as those competition agencies, or legislators and regulators that are approaching the process of modifying the rules towards digital markets, to help clarifying the rules that they are adopting, especially when those rules depend, as far as the enforcement is concerned, on the action of the addressee.

Of course, our analysis has some limitations. Our corpus is rather small, and we could adopt some techniques like bootstrapping in order to enlarge our corpus. But we could also let our analysis evolve. And also, in order to have enhanced replicability, we should try to limit as far as possible manual coding, which was exposed, as we already said, but in any way, we do manual coding. And finally, this is my last word, and I conclude. We think that despite those limitations, there is really a plus for competition authorities and for rule makers more broadly, in order to have to use these tools to combine the usual job with the computational tools. And eventually, in order to strengthen our analysis, we should have a sort of a computational toolbox, made also of other techniques like topic modeling, which was already mentioned, but also text and document similarity measures, which could be used in order, for instance, to compare the two versions of the DSA and the DMA, especially the last version that was released the two weeks ago by the European Council, which is slightly different by the one that was adopted one year. And then sentiment analysis, which also might help elucidating on the position of the stakeholders. And, of course, transnational is also an issue which is extremely useful, because clarifying those terms would also have to consider the intrinsically transnational nature of digital markets. So, what this analysis would necessarily need to be is transnational. So therefore, including actors and stakeholders that are in other parts of the globe, but not just in Europe. Thank you very much for your attention.

Aleksandra Wierzbicka

Thank you Fabiana for your insightful presentation. Just a quick reminder for the audience that you can ask your questions in the Q&A section. And I'm leaving the floor to Daniel, for your comments on the fabulous presentation. Thank you.

Daniel Crane

Great. Well, thanks very much, Alexandra, and thank you to Thibault for putting on this wonderful conference and for inviting me to participate in it and to Fabiana and her co-authors, congratulations on a really nice and interesting paper, which I think is really the starting place for much future analysis.

Like Pinar in the last segment, I am someone who does not have any expertise at all in the techniques of computational law, or computational antitrust, or NLP. I'm really coming at these as a consumer of the work product you create as opposed to a producer of it. And so, what I want to do is really ask some questions, I'll ask for particular questions that strike me as a potential user of these kinds of analyses, about what I should make of them. The first question is, what have I learned here? Or what should I be learning that I wouldn't have predicted with a high level of confidence without this kind of a study being done? So let me give two examples. The fact that larger and medium-sized companies have a more favorable disposition towards the term self-regulatory and that smaller entities and maybe individuals have a less favorable disposition towards the term self-regulatory strikes me as being not surprising at all. It's exactly what I would have predicted. Regulation by the government, of course, is in place to protect the weaker from the stronger and so self-regulation is really non-regulation, if you will, and so it strikes me as being exactly what I've predicted, that smaller entities would have a less favorable view of self-regulation, whereas larger entities would of course, think that they can manage these processes by themselves.

On the other hand, I would not necessarily expect this is the key observation of the paper, that different stakeholder groups would actually use the terms differently, right. And this is an important observation because it suggests not only different perspectives among different classes of stakeholders, but that the language itself is manipulable, that it can mean different things in different people's mouths, or that they can construct it differently. And that's really important because this suggests that these ideas do not necessarily have a fixed objective meaning. As we know from theory, language words do not, and they can lead to the same language deployed in different occasions in different contexts leading to different outcomes, particularly if different regulatory processes are captured by one stakeholder group or another. The choice of language may not be that significant if the different stakeholders have the ability to manipulate it to mean different things in different occasions.

The second question goes to explanation. So, what actually explains this dispersion and attitudes among key terms among different stakeholder groups? And so, I can throw out a few different possible theses or explanations and see if any of these or they can all be true on some level. The first one is attitude note, which is that these stakeholder groups simply have different underlying preferences that translate into different understandings of the regulatory ecosystems in which they operate. Now, those could be normative preferences in the sense that they have different actual preferences as to what's best for them. Or they could be just revealed preferences simply, this is how they kind of vote with their feet or with their language. But the observation would simply be that by observing the differing deployment of language among different stakeholder groups, we can actually observe how they think differently, how they have different preferences about the regulatory landscape.

The second one is more strategic. And that is to say that different stakeholder groups are trying to shape the meaning of regulatory language as they interact with

regulators. Now, of course, language is not merely reflective language, is not simply taking what's already created, it's every time a speaker touches or appropriates language, they can be constituent to it by trying to change the meaning of language. So, part of these dispersions in the use of key terms by stakeholders may simply be a strategic effort. Not that they actually understand these terms differently in terms of their baseline preferences, but they're actually trying to shape what it means to be a gatekeeper or what dominance means to sort of create a different outcome that satisfies or maximizes their preferences. And then the third kind of bucket of possibilities here for me is institutional, and that is that perhaps different stakeholder groups are coalescing around the same political intermediaries. That could be law firms, that could be trade associations, that could be lobbyists, that could be other sources of information. But their deployment of language simply reflects the ways in which they're accessing the language or the or the intermediaries to which they're channeling their comments to the regulatory agencies. And that'll be an interesting story because it would suggest that part of what's going on here is not so much the underlying stakeholders themselves but that they are intermediaries who are shaping the regulatory landscape through their deployment of language.

The third question I wanted to pose is how are stakeholder uses of language shaped by the background, political, legal, regulatory, or constitutional conditions in which those stakeholders operate? In particular, do stakeholders adapt their deployment of language differently in different regulatory environments? So, for example, in American laws, we know the attitudes towards dominance have been much more permissive towards dominant firms, even dominant firms that exploit their market advantages than EU law has been for a long period of time. Now, that doesn't have to be forever. As we know, the United States right now, that may be changing. There may be legislation, there may be case law developments, certainly the antitrust agencies now are trying to change some of that perspective. But as a background condition, the stakeholders would simply have a different view of the connotation of language-based upon the regulatory baselines in which they operate.

For example, self-preferencing - the paper mentions the *Google Shopping* case in which you found Google liable for abuse of dominance under Article 102 for self-preferencing. For now, at least the leading preference in the United States, or the leading precedent in the United States, is the 2013 decision of the Federal Trade Commission, holding that Google's self-preferencing, its vertical search bias, was not a problem, that has benefited consumers. And so even though, of course, that is being reconsidered, now by the Justice Department, and perhaps for the federal courts, perhaps by Congress, perhaps by FTC rulemaking, it still sets a kind of a different baseline of expectation on whether self-preferencing has a negative connotation, or a neutral connotation, or maybe a positive connotation, depending upon whether the entity, the stakeholder is coming from one regulatory tradition, or another regulatory tradition. So, I think it'd be really interesting to know if we could code stakeholders geographically and determine the effect of background regulatory socialization on their deployment of regulatory rhetoric.

And then my fourth and final question is: what use can be made of this sort of analysis by different constituencies? Or really, who's the audience of this information? And how can different audiences use or abuse this kind of information? I think the paper, as I read it, seems to be targeted primarily at regulators or trying to give regulatory authorities more information about how stakeholders think about the deployment of language in key things like the DMA. But it could also be the case that the stakeholder groups themselves will make use of this kind of analysis. If regulators were to begin using this kind of computational techniques to inform regulatory design, what strategic responses should we envision from competing stakeholder groups. And just follow some public choice theory down the pike. And you'd predict that stakeholder groups would get savvy, that is if these kinds of computational methods are being used to shape regulatory outcomes, that we can sort of paper are our record, but in ways that are strategically designed to achieve the rogatory outcomes that we prefer. So how should I think about the repeat player game here, which is if this kind of analysis comes to inform regulatory decision making, are there ways it can be exploited, it can be manipulated by key stakeholder groups? And would that mean, and then are there counter strategies by regulators? Right? Can you outsmart the regulated entities who may want to use these kinds of interventions in ways that would actually defeat the regulators' purpose in also using the same kind of regulatory technique? So again, I know that that's a question really beyond the scope of the paper, but as a consumer of these kinds of analyses, they're the kinds of things that I'm thinking about. But again, really interesting project, I really learned a lot from it and enjoyed it. I look forward to hearing more. Thank you.

Aleksandra Wierzbicka

Thank you, Daniel. Your questions were really thought-provoking. Fabiana, do you want to respond?

Fabiana Di Porto

Sure. Thank you very much for your careful reading. And thank you for the comments, and the questions, which I think are around the idea of manipulation. I think you used the word manipulation several times. I'm not using my algorithm here. But I think that you use it in almost all your questions, which means that you're very worried about the idea that the use of algorithms by competition authorities, rule makers, legislators, regulators, and whomever is really worrisome somehow for you. Eventually, this of course would require some more qualitative analysis. Now, coming to your questions, of course, people tell different things in different sessions, depending on the environment, and depending on the language they are using, and on the institutional environment. Of course, we analyzed documents that were drafted by several different stakeholders, most of which were non-native English speakers, we excluded them, especially documents that were not drafted in English, of course, this is because our algorithm needed to run a web analysis only in one language, at least for now. And, of course, all the big texts were not European. They were American, and they contributed to this consultation process. Did they respond the same way an NGO would respond? Of course not.

Were the questions the same? Some of them were the same, some of them were not. This is the reason why the Commission allows all respondents to both respond through the questionnaires and to also attach a document whereby everyone is allowed to just answer freely, completely free and use the jargon they prefer. And this is where our algorithm could identify the largest differences and the widest differences, but also where our algorithms had the widest vocabulary and therefore could perform with the best data in a way. But of course, the US companies might have referred to concepts that are not European or are not in line with the European competition law jargon like dominant position; they might have used concepts like monopolization, which belong to the US antitrust jargon. This is where we could measure this distance, and this is why it was useful.

Another thing which is also interesting and relevant is that there we have the problem of how do we build the clusters, because one thing is considering small and medium firms from an economic point of view, and another thing is to consider the same entities and compare them from an NLP perspective. A language is completely different, and the way a bigger firm expresses itself is totally different. Because the document that is produced by a huge firm is not the same as an academic produces its own document of course, but from an NLP perspective and from our perspective, it is completely the same—we just analyze documents and words and association between the expressions. when we had to clusterize, we end up having medium and large entities in the in the same group or something that in economic terms does not make any sense. But from a linguistic point of view, this made sense. So, would that be manipulated? I don't know; I cannot tell you straight away. Because first of all, imagine that the European Commission decides to use algorithms and NLP techniques in order to streamline the job, the heart job of close reading all the documents, the 1000s of documents. That will not exclude the close reading. And second, that will not exclude the lobby, that will not exclude all the remaining processes, that will not exclude cost benefit analysis. This is just something that will be on top of the rest.

I hope I somehow answered some of your questions. I responded the question about the background, and also about the context and institution. The other question was, who is the audience of this paper? Well, I think the audience of this paper is the legislator, the competition authority, and more, more broadly, the rule-maker, which can, of course, include regulators. Do I think that algorithms could be used and computational tools more broadly, in the future, in order to produce rules or to implement rules, to enforce rules? Well, with a high degree of visionary attitude, my personal opinion, based on my research, which I actually showed in my last slide, the thank you slide, I would say partially yes. And also, I would add that I think that in the future, we would have rules that will be eventually self-implementable, whereby rule makers would provide the specifications to the industry that would eventually implement them.

Aleksandra Wierzbicka

Thank you Fabiana. We have a question from Thibault, so I will give you the floor.

Thibault Schrepel

Thank you very much. And we have a few questions in the Q&A. But it seems that indeed, as a regulator, you may want to avoid the use of certain keywords if you know that they are not understood the same way by all market players. So, this could be something which is relevant for them when they draft legislation. My question is actually for Dan. You are working on the history of anti-monopoly. I'm wondering if the results that Fabiana is showing are consistent with what you found in your research? Or do you see that there is a change in the way big players approach antitrust laws?

Daniel Crane

My research on anti-monopoly goes back hundreds or 1000s of years. And so, I think you get different attitudes in different times. I do think some of these techniques and also the techniques presented in the previous panel would actually be very useful, though, in doing historical research on attitudes towards monopolies. For example, I'm doing a read right now of every case, US case and state or federal court in the 19th century that use of the word monopoly. It's several 1000 cases, and whether some of these techniques can be useful to me just analytically to swordfight cluster words together, I think that that could be really useful. So, I'm interested in learning more about that. I'm more of a consumer than a producer of these technologies, but I can certainly see how they can be very useful.

Aleksandra Wierzbicka

Thank you, Dan. As Thibault said, we have some questions in the Q&A section. And there is one about the linguistic considerations because the EU is very diverse when it comes to languages. So, the question from here is, whether there's some kind of correction taken into account because the usage of English words might not be the native language of the party using them. So, the question is whether you take that into account?

Fabiana Di Porto

Yes, we did. We did take it into account. However, we could not translate them directly into an instruction to be given to the algorithm. We knew that we accounted for that, but we could not translate it into any instruction.

Aleksandra Wierzbicka

Thank you. So, I have another question about from the Q&A section as well. So, can an AI algorithm trained with machine learning create new words? Words that still do not exist in the legal field.

Fabiana Di Porto

Okay, there we have natural language processing, we have also, let's say, a branch of the natural language processing, which is natural language producing, which is something that is developing of course, and we have actually several services that are based on this technology. More and more chatbots in the US use this technology and Alexa as well. But to make an example of this use specifically for the legal domain is extremely difficult, just because rules, I mean legal rules, entail so much more than just prohibitions. As we know, from the times of Professor Lessig's code, rules are much more and imply so much more than just duties and prohibition. So, there is so much more content that cannot be coded. And therefore, let's say that at this time and this level of the level of technology thus far available, I don't think that rules are completely codable. And therefore, we cannot thus far produce rules completely automatically. I am working hard on these types of projects. I also hosted a seminar last year convening a lot of experts in this area. It was called the law and RegTech seminar. But I can tell you that it is not yet the time. I'm not saying that it will not be possible. I can also add that Regulatory Sandboxes stemming from the financial, the FinTech world, are something that we should look at, because they are going very fast - but we are not there yet.

Aleksandra Wierzbicka

Okay, thank you so much. We have three minutes left. So, I have the last question for Daniel, and then I'll give you both one minute for the concluding for the closing remarks. I was wondering, what is your take from the American perspective? So, for example, if there was a similar study run in the US, do you expect the results would be the same? Different, completely different?

Daniel Crane

Yeah, that's a great question. And I think my instinct is that the results would be similar. I think what we're picking up here is that different stakeholder groups use language in different ways. And whether that's sort of innate or strategic. I'm not sure this goes to the question about where the causal stories are, what the explanations are, but my suspicion is that there are inherent differences among stakeholder groups in how they even deploy regulatory language. Now, that's a very interesting observation because it goes, as I said before, to almost regardless of what language is adopted in regulation, it's going to continue to be contestable in regulatory practice. Of course, we know that that language is not self-defining. But this really just shows very viscerally that there is a power in the deployment of language in different voices, that begins even well, before regulation is adopted.

Fabiana Di Porto

If I can just add one word, do we have one minute?

Aleksandra Wierzbicka

We have two minutes left. So, it would be as a part of the closing remarks.

Fabiana Di Porto

We have two different legal system, the European and the US, so we should absolutely take into consideration the role of the case law, which can be definitely factored into the idea of rule coding. And as far as the EU is concerned, definitely the decisions by the European Court of Justice, these should be taken into consideration because they clarify the way terms should be interpreted. While in the US, I think that definitely we should take into account the role of the precedents. This explains also the relevance of databases like Westlaw, and all the rest, which are less relevant from a European perspective.

Aleksandra Wierzbicka

Thank you. And Daniel, do you have a final word?

Daniel Crane

Thank you Fabiana for a really interesting paper. And I'm very curious to see the extensions. One suggestion, if it's possible, this goes to previous questions as well about nationality of different stakeholders, whether it's possible to code for nationality, and see how that affects the results. And if it's possible or not, I'd be very interested to see that.

Aleksandra Wierzbicka

Thank you very much. Our time is up. Thank you for your presentation. Thank you for your comments. Thank you for the discussion. And thank you for the questions. And I'm giving the floor back to Thibault for the next panel.

Thibault Schrepel

Thank you. Thank you very much Dan and Fabiana, it's been a pleasure. Now, Fabiana and her team excluded the non-English documents from the analysis. What you will hear in the next panel is how, even though you may try to train a machine learning using the case law coming from the same language, but also from the same country, it might be really hard to make it work.

#3 – Gleaning Insight from Antitrust Cases Using Machine Learning**Thibault Schrepel**

We move on to the next panel, and I can see that pretty much all the speakers are here—Ashwin is here, Giovanna, Daniel, and David need to be promoted back. So... let me introduce our speakers. The subject for this panel is “Gleaning insight from the antitrust cases using machine learning.” You will hear about how you could actually teach antitrust to an algorithm. We are joined by Giovanna Massarotto, who is an academic fellow at the Center for Technology Innovation

and Competition at the University of Penn, and also an affiliate of the UCL Center for Blockchain Technologies. We are also joined by Ashwin Ittoo, who is a professor of computer science at the University of Liège, more specifically HEC Liège. We also have Daniel Linna who has a joint appointment at the Northwestern School of Law and at the School of Engineering, as a director of law and technology initiative. I think it's fair to say that Dan is one of the leading scholars in the field of computational law. And we are also joined by another of my favorite scholars in the space. He has a website which is absolutely fascinating. I hope David that you can talk about it. David Colarusso is the director of the Suffolk University Law School, Legal Innovation and Technology Lab, where he is a professor of law. So, what we have decided is to first give the floor to Giovanna and Ashwin for their presentation, and then we'll move on to the discussions. And again, feel free to just drop your questions in the Q&A. Enjoy the talk.

Giovanna Massarotto

Okay, so thank you so much, Thibault. I'm really delighted to offer you a few thoughts on a subject that's of direct and immense interest to all of us in the antitrust community. And the takeaways of from our presentation and our paper are, first, that AI is everywhere. Today, AI is from speech recognition to AI techniques that suggests what movies, you might like on Netflix, and AI can be used also in the field of antitrust because we have more than a century of experience in enforcing antitrust law that can be exploited by using an AI algorithm. And the third takeaway is that our paper experiments, aims to pave the way for future AI systems in the field of antitrust. So, the reason why I believe that it is fundamental to reason on how to exploit AI techniques is because, as Richard Posner observed back in 2001, antitrust enforcers do not have technical resources, and do not move faster enough. So... the idea of using, experimenting with AI stems from my book on antitrust settlements, in which I collect data related to both the US and Europe, and then perform a multi-regression analysis to predict the possible effects of the Google breakup using AT&T coefficients. And differently from most of scholars that were exploring how AI could be used to engage in anti-competitive practices, I was more interested in exploring the question of why don't we use AI to fight antitrust practices in today's data driven economy.

I discussed the idea with Ashwin, and we translated this into a reality. And we developed a proof of concept that was built by using data from FTC antitrust cases under Section Five of the FTC act and we adopted supervised learning techniques. Then, the project was developed in four stages. So first of all, we collect data and define the most significant variables that could be interesting to detect useful patterns in our data set, code the data, and clean the data. And afterwards, we ran a machine learning algorithm to see whenever we were able to find that useful patterns. And we assess the results from both a technical and an antitrust perspective. Now, I will leave Ashwin to proceed.

Ashwin Ittoo

Thank you, Giovanna for this introduction. And good evening, or good afternoon, or good morning to all of you. So as my colleague says, we are trying to build an AI

algorithm that could understand antitrust cases. And the first step was to create a data set of these cases. So, we tried to formalize each one of those antitrust cases in terms of a number of variables, which are on the screen here - the year, the industry, the conduct being investigated, the remedies adopted, and so on. And in all, we had 72 cases that were ruled in the second file, the FTC act that constituted our data set. In the next slide, you can see a list of all our variables.

These are the variables that we use to describe each one of those cases. As you can see, we had a maximum of four types of conduct and a maximum of five remedies. Because the cases we examined had a maximum of five remedies and four types of conduct, we had the year, the industry, and of course, the case name. I will describe how we formalize each one of those cases. Here you can see examples of each case. Of course, you can only see a partial view of all the variables that I described earlier, which were used to describe these different cases. Right, so now we have our variables to describe each one of those cases. We had 72 cases in all, and as you will see, in the next slide, we will start to first explore our data. The main aim is in doing this kind of exploratory data analysis used to detect outliers in inconsistencies or incoherencies Irrelevant variables and highly correlated variables, you don't want your learning procedure, the bias, for instance, by variables, which are highly correlated to each other, or which have extreme values, which are outliers. One way to address the specific issue of trying to find out which variables tend to be correlated with each other is by a heatmap. I just showed you an example. From this heat map, we saw that, indeed, our data was correct; there were no outliers, there were no variables that were specifically not correlated with each other. Then in we said, now it's time to start applying our machine learning algorithm.

So, our aim here is not to make any predictions. As you will see next on the next slide, we are mostly concerned with trying to find patterns in our data. And since we are not trying to make predictions - we could, but it was not our aim—our aim was to try to find patterns that characterize those cases. And in machine learning parlance, we're doing unsupervised learning, and one popular type of unsupervised learning is clustering. As you can see on the next slide, there are several types of clustering algorithms. K mode and the bisecting K means, they all adopt more or less the same operating premises, which is as follows. They take each case, which is described by the 16 variables, each of these cases is created as a vector, which is plotted as a point on a multi-dimensional space.

Each case is now a point in space, and then we compute the distance between those different cases because it's quite easy to compute the distance between two vectors. Once we have the distance between the different cases or the different vectors, we can group them based on the similarity of these cases; those which are more distant are going to be less similar, and vice versa. On the next slide, you can see the groupings of the different cases. So on this particular graphic, all dots represent one case. And similar cases have a similar color, that is, in the vector space and distance, they were close to each other. And they have been grouped into a single cluster. Here we have defined to have four clusters of similar cases, identified on this graphic by different colors. And know that each dot here is a vector of 16 variables, which corresponds to each case, which I have showed you earlier.

Now we have applied our clustering algorithm on our data set of cases, we have found out which cases were most similar to each other, which means that for each case, we could identify a cluster to which it belonged. And then, once we have this information, we could determine, for example, that the first case on this on the slide here, as well as the fourth one, and the fifth one belong to the same cluster, that is, cluster number two. The algorithm did mean that these cases were similar to each other; they were not distant in the vector space. And therefore, we grouped those cases into a similar cluster, which was now cluster two. Obviously, these results are quite dry. They need some kind of antitrust interpretation, which I will leave my colleague Giovanna to continue with.

Giovanna Massarotto

Thank you so much, Ashwin. So as mentioned, we assess, evaluate the results, also from an antitrust perspective, in addition to a technical perspective. Cluster one was interesting because it detected data and computer industry together with the healthcare and pharmaceutical industry cases. The most frequent practices investigated in this cluster were price-fixing exclusionary conduct, and the remedy suggests work contract limitation and compliance obligations. Cluster two, instead ss interesting because it concerns cases that investigated a conspiracy. It's interesting because as remedy was suggested limitation in the exchange of information as the most frequent. In cluster three, the majority of cases concern an investigated conduct for which there was no remedy, besides reframing from the anti-competitive conduct, and most frequent practice was an agreement to non-compete and an invitation to collude. The last cluster, four, was dominated by cases in the healthcare and pharmaceutical industry and detected as the main antitrust practices price fixing, suggesting limitation in exchange of information as remedies, which makes perfect sense from an antitrust perspective.

The idea was, okay, let's see whether the pattern also makes signs from an antitrust perspective. And it seems that it makes. I believe, and we believe in our paper, as mentioned, aims to pave the way for future AI systems. Because I think there is no alternative in the enforcement of antitrust in today's fast-moving technological markets because, as mentioned since the beginning, antitrust agencies need to move at the same speed as markets to be effective. At least, these are my thoughts. AI has a lot of limits, our experiment has limits and limitations, but this should not prevent the exploration of these systems also on the field of antitrust.

I believe that my message, our message today was clear. It was clear and straightforward, in a sense that we don't think that AI can replace antitrust agencies such as the FTC. But we believe that the AI systems can turn out to be an invaluable antitrust tool to enforce antitrust principles in today's fast-moving technological markets. Thank you. Alright, thanks a lot. And we, and we look forward to receiving your comments and your thoughts because it was a nice experiment. And what I can tell you is that since the beginning it was great to work together to combine our expertise. I hope to hear that there are other projects on their way now.

Thibault Schrepel

I remember that one of the first questions that I asked you in the podcast is, how long did it take? And how costly was it? I'm sure Dan and David will have more comments and questions for you. Dan, the floor is yours.

Daniel W. Linna Jr.

Sure. Thank you, Thibault. And thank you, Giovanna and Ashwin, for this work. Let me see if I can share my screen as well. And I got just a couple of slides here. So yeah, I thank you for your presentation, I really appreciate the combination of law and computer science expertise that it takes to work on projects like that. It is great to see the continuing work that's happening in this space.

I just wanted to kind of use my 10 minutes to walk through some of the contributions and a few questions that I had. And maybe also I'll highlight some of the things that, from the audience perspective, I might be repeating a couple of things that you said, but just want to kind of try to reinforce what some of the contributions are, and what exactly you did. And so, you know, the contributions, I think you're illustrating here, the way that machine learning tools have the potential to aid antitrust agencies and preliminary screening analysis of cases and potentially ultimate decision making. As you said, you weren't trying to make predictions here, and this is the first step, you're trying to lay a foundation, demonstrate the application of some methods, demonstrate the validity of those two so that additional work can be built on top of this. I think it's interesting that you pointed out as well that it wouldn't just be necessarily antitrust agencies that would use these tools, but regulated companies too.

As a fan of computational law, generally, this idea of creating tools that cannot help just the regulators and enforcement agencies, but also individuals and companies who are subject to law and regulation, better understand what's expected of them and how they should be acting in the environment is important. And then also your use of unsupervised algorithms for legal analysis, your contributions here, your research questions, you laid them out very clearly in the paper.

I want to talk about each of these research questions just a little bit, asking the question about whether AI can assist antitrust enforcers, and then also this idea of accelerating and harmonizing global enforcement of antitrust principles, many of these sub-questions about the extent to which characteristics can be extracted from antitrust cases, or weather patterns can be identified. And then it did raise an interesting question of bias that I'm hoping we can talk about a little bit in this discussion as well.

I'm at the first couple of questions. I think that you demonstrated, you showed some work here to suggest that we can extract characteristics from past cases and identify patterns. One of the things that I did want to talk about a little bit is that I'd like to learn a little bit more about how you selected these cases. And just for the audience, I would want just to highlight again that when you first started, you're looking more

broadly at European cases and cases from the Department of Justice antitrust division and the Federal Trade Commission but had to end up settling on just Federal Trade Commission cases because the heterogeneity across those different types prevented you from finding any patterns. Now you said you train the algorithm mainly consent decrees; I'd like to understand a little bit more about that, about which cases were used. When I went to the fed the FTC website, I came up with, I think it was 234 cases in that time period. And so... you can tell us exactly how you selected these cases; maybe that was the full population. But if it wasn't the full population, how did you select them?

You already walked through the variables that you use and talked a little bit about the different methods. I thought it was interesting. It was really interesting when I was first reading through that remedy, one was not relevant. And that really surprised me. But then I saw that later on you talked about it that the way you coded remedy one, it had to do with the FTC requiring that the entity refrain from the investigated conduct. There was no variance across that variable. You ended up discarding it. And so... I did have a related question because I know that the remedies were random, you said. I guess there's some arbitrariness and kind of deciding whether a particular remedy be reclassified as remedy two or remedy three. And I wanted to know if that was the same case for the type of conduct as well, because I think that could potentially affect the analysis. And again, really, I mean, I think you summarized so much work that you did in this paper and 20 something pages in the appendices, which I thought were really helpful.

I'd just like to understand more about kind of how these variables are constructed for the analysis. You talked about a couple of these charts already. But I'll just stop here on this one about I think this question is important, too, because, about the type of conduct, it's really interesting to see that the second category of type of conduct, it appears, had almost four times the impact as the first type of conduct. And I'm kind of curious, like, is that an artifact of the way that the cases were being coded? Potentially, what's going on there that might help us better understand what's happening.

I appreciate your analysis of the cluster. I think there's some interesting trends that you're observing. I think this is one of the challenges when you're doing this kind of work. Well, how do we really know how well the model is performing? Right? We've got some anecdotes about the way about some of the groupings and things like that. How can we have a better sense on whether the model is picking up on some sort of signal versus just some noise that may have been introduced in the way we put together the data? On your initial questions about accelerating and harmonizing antitrust principles, I think the accelerating part; there's work still to be done to show theoretically. I think these methods can help. And then it's important to consider the processes - who's gathering these facts, who's applying the labels to the facts, what's not being enforced, what's in the data, what's missing from the data. And as you discussed in your paper, there are challenges with harmonizing. And I think asking us questions about harmonizing antitrust principles globally, we know what kind of progress we made.

This bias question, you raised it briefly in your paper. And I think it's a really interesting question to ask. There's been a lot of research, on-going research, I should say, being done at Stanford, David Angstrom, and Dan Hoe have done research on algorithms in the administrative state, also by Cary Coglianese at Pennsylvania. When we start using these algorithms, it is important asking questions about potential bias. I think compass is an example of bias in the criminal setting. I think it's really interesting to consider not just bias and technical questions, but the fairness questions raised and compass because some have defended the methods, actually even used the compass. But there are some really challenging fairness questions. You've laid out some reasons why antitrust can be a safer testing ground. For it, we're looking at efficiencies and competitiveness and things like that. There's some reasons to believe that antitrust could be a good place for testing these. I think we have to ask, though, then how much of that would rely on the assumption that protecting human rights isn't a part of it? As you know, there are some normative questions. And in fact, there's a lot of discussion now about thinking about the extent to which antitrust law should be thinking about racial equality and thinking about the ways in which there might be bias, or there might be disproportionate effects of enforcement actions or failure to enforce, for example. So I think there are some interesting questions around bias. How might we test for bias? What sort of features away do we need? What sort of data would we need? How might we get a sense of some of these antitrust decisions? Who's affected by them? Are there underrepresented groups that are disproportionately bearing the costs of enforcement action or failure to undertake enforcement action? How might we assess impacts and potential harms? So those are a few of the things that I just had a couple of wrap-up questions on, on slides. And I'll paste these in the chat after I finish here. So we can have them on top of our minds after David concludes. And we can hopefully discuss some of these.

Again, to understand a little bit more about how the cases were selected, I would like to know if that was the whole population during that period of time on these part two consents. When it comes to the coding of the variables, understanding if the type of conduct also had this randomness in it, I think it'd be really interesting to talk about the limitations and about using consent decrees to train algorithms to make predictions. I am wondering the extent to which we know that in the judicial space, if we look at the facts section of a case and use that to train a model to make predictions on outcomes, well, one of the problems is that the facts are written when we already know the outcome, so that there can be some challenges there using that type of data for prediction. And do we have any sorts of challenges in data like this?

I think we've got to Google one other thing to kind of keep closing the gap here. And moving forward is asking questions about what data is missing. And of course, there's investigations that don't proceed to enforcement action. There maybe are activities that should have been investigated, but were not, what are the impacts on people that we have data about. There are plenty of things that are difficult to measure, some would say impossible. I think we can come up with a lot of ways to measure a lot of things. But first, we also have to identify those things that ought to be measured, that we have to collect data about. I just noted that you noted that

during that period of time, you think enforcement proceedings probably did not change much. I think it'd be interesting to ask if that's a reasonable assumption. I mean, that'd be something else we want to control over time. And of course, there were different presidents. During that time, I've seen other researchers suggest that maybe there were some changes in the way the FTC was enforcing Section Five. So that's just something to think about. And then just finally, I hope we can have some discussion about where are we? What's the state of the art? And how do we keep working to close the gap between where we are and where we want to be with these tools? So those are some of the things I look forward to us discussing.

Thibault Schrepel

Thank you very much. The floor is yours, David, for more questions to be discussed.

David Colarusso

First, I just want to say you've done a great work here, creating structured data from the FTC proceedings. You had to go in and do all that coding to get out that structured value. So great work there. I also liked seeing some of the toolings you were using.

In fact, I really wanted to get into the data and dig in there. And so, one of the things I really want to know is if your data set and your code is published anywhere for people to work with, because I think that would be very, very useful. And then I must admit, my comments and all my questions were sidetracked by a single slide in your presentation, and I just need you to talk me down from this. So, I'm going to share here. And we can look at this together. So... this is where I really hope you can answer these questions. I look at this as the sort of the features and the targets. This is what you were working with your after you've gotten your clusters. This was the data you looked at, presumably awesome. Everything minus the target. So, minus the cluster, this was the data you fit into your clustering algorithms.

Two things really jumped out at me. One is that it looks like maybe there's a confusion of categorical and numeric values. For example, under industry, I see their numeric values between at least one and eight. And if I'm interpreting right, I suspect it means that you had one industry you call it number one, and the second industry you called number two, but you know, the biomedical industry is not two times the manufacturing industry. So I hope that's not what happened, because if it is, then I don't think you can get reliable information out of the industry column. I think the best practice would be to take that and create dummy values for each of the categorical values.

Assuming you had an industry using to have eight dimensions, each would be a binary value, and you would see that present. I'm worried about the fact that maybe those got completed. I'm also worried about the types of conduct one, through four, doing the same thing, but in piecemeal, and the remedies two through five, doing the same thing. I noticed a remedy two; I see a remedy one. And I also see a remedy one and remedy five. And the thing that also really worries me as I see remedy zeros and types of contact zeros always on the trailing ends, which makes me think that

those are no values, which is to say that you're coding as zero a no remedy or, no conduct.

Now I look at those four-classification bends. Again, that would be giving a value to something that didn't have a value. I think the best practice would have been to take that and remember, what was 14 or 18 types of conduct, split those into 18 dimensions, as binary classifiers. And so, I just hope you can talk me down from this because I'm worried that that basically means you need to throw out your industry, your types of conduct, and your remedy features. I think that also would help explain Daniel's concern here, it might be that the reason why code of conduct is showing most predictive power, because most predictive power, in this case, would be for classification in your grouping simply because that's the one with the most entries in it, as you have no values and the subsequent codes of conduct. So, I really love the tooling. I love the work you did and the coding and stuff. I love the idea of the research agenda. I just had some big questions about that, about the underlying data. I think I'll just stop there. And then look for your answers to that. So yes, thank you.

Thibault Schrepel

All right. Thank you very much, David. I've also put the link to your website in the chat, codingthelaw.com. I'm not sure if you do have a new version of it.

David Colarusso

It changes every year. And I also should make very clear... the no value error... I have made that myself, and I had to retract it. And it was horrible. But it happens. And so... yeah, there you go. That's that.

Thibault Schrepel

All right. Thank you very much. I think I'll give the floor back to Ashwin first.

Ashwin Ittoo

Thank you very much for this. I will stop on your question, which was in fact a very interesting question, and a very pertinent question, I understand the concern. In fact, you have to be assured that the values we use were categorical values. So, two for tech industry, and one for pharma. It does not mean that tech is twice as important as pharma. So, they're categorical values. And we did use dummy variables that didn't show in this slide here as it was going to explode.

David Colarusso

Makes my heart feel a lot better!

Ashwin Ittoo

So, I wrote PowerShell on top for that reason, specifically. And also, what we did to address these issues or categorical variables was to use a different algorithm known as K modes, which is more suited than K means for this kind of data. And we've tried to combine demos as well, and then we stopped because it was getting too complex. But be assured that everything has been done according to the best practices, we did use category dummy variables, sorry, but I didn't show them on the screen. And that's a very important point because many people tend to forget about those dummy variables and confuse numerical values with categorical values.

Now for the second question about the type of contract zero, I have to check the other one on that, and I think zero had a specific meaning in our case. I don't recall exactly which one had to look at the data again. But it had a specific meaning, which I don't recall. Zero did not mean no value, or null, or empty. It had a specific meaning, which I don't recall from the top of my head now. I can get back to you on that later. And then you have one more question technical question.

I can jump back to what was asked by Daniel earlier about the type of conduct, something which David also asked. So, as you know, all these algorithms they have some randomness inherent into them. And whatever the credit we should not take it at face value, because randomness can be quite problematic sometimes, especially in that case, because we had only 72 proceedings to work with. So, what we did when we tried to assess the variables' importance was to try to run two different algorithms and compare whether their results were aligned with each one another. And what we saw was that the same variables were evaluated to be important by both methods, random forest and support vector machines. But the order of the magnitude were different. I think what I showed you on the slide, or what you showed in the slide was one of the results for random forest, which showed conduct two, three most important, but the other algorithm, support vector machine, showed conduct to be the second most important variable, I think. So, they kind of ally in telling you, these are the top five most important variables that can be used to assess those cases. And it might be, as you said, there was some randomness in the types of conducts, in how the corpuses were rolled. And this is why the results show that conduct one was not as important as two.

So, just to wrap up David's question first, yes, we did use dummy variables to convert those values into specific categories and so on. And then the zero values didn't mean no values. And we did try K modes, which is very commonly used for categorical variables clustering. And for Daniel's question, conduct was most important, but it is difficult to take it at face value, you have to compare the results with some other predictions and see if they align or not, especially, because you have randomness, which is inherent, and even more so our data set was extremely small. And this could be one limitation of our research, such a small data set. And I was in a conference last week in South Korea, and it's been more of the same question, which was whether we had considered cases that will have been dropped. So that would be an investigation, and then that will be dropped from further

investigation. And so, this, again, tends to start to point out that there was a lot of bias because our data selection was biased, you can say, because we focused on the cases that were with consent decree. But I will leave it to my colleague to elaborate more on these legal aspects. And one last thing before I forget, yes, the codes are available on Google Colab. And I think the data set also is available somewhere, by link in the article, or otherwise, we can publish it online on GitHub also.

Thibault Schrepel

So yeah, I remember. So, this was one of the first article that we published. But we've had a link to all that in the paper somewhere.

David Colarusso

I can't recall where it said in the appendix. So perhaps it's hiding?

Thibault Schrepel

No, it wasn't in the appendix, it was a link because it was too big. And we thought, well, you know, it has to be short. So, this was our trick to put it outside and pretend we didn't have more than 20 pages. I will just ask a question before I give you the floor, Giovanna, and the question is actually for you. And indeed, here Diana is asking the question about the fact that, by definition, you've been labeling data using the anticompetitive conduct that we know already. Right. And so how would you envision using a similar technique to capture novel practices, whatever that means. You've been working with supervised and unsupervised machine learning so maybe it's time for you to discuss all that and to answer Dan and David's questions.

Giovanna Massarotto

Thank you so much, to both of you. Such comments and questions help not only us, but also other researchers that are working in the field. I also thank you for the advice in understanding where we should focus our attention to improve, and to keep exploring these new techniques in different fields, but also in the field of antitrust.

With respect to your first question as to why we have chosen the FTC, we initially thought to include other agencies. But it was the first experiment, and we thought it was more useful to focus on one agency. Because if you look at the other agencies, the rules of law that they apply, they have differences. And they have different powers. So, the chances that we were able to find useful results from combining data from agencies from all over the world was really low, I guess. So... we trained our machine learning focusing on only one agent. We might explore the opportunity to extend our data set, and include other agencies. Honestly, I don't know what the concrete possibilities are to find interesting patterns, because as mentioned, we focus on FTC antitrust cases under Section Five of the FTC. But if you need to include the other agencies, you need to include the different rules of

law and create different variables. It's something that is interesting if someone is willing to explore the idea to combine and build an AI system that includes all the agencies, I would be very interested in knowing the results. But for us, it wasn't worth the effort. I will say, as a first experiment, as a first proof of concept, we decided to narrow our invitation to the FTC, mainly consent decrees; in a sense, we include a few cases that weren't settled, including Google. But 90% of the FTC cases except for mergers are settled by means of consent decrees. This is why the algorithm was trained to use mainly cases related to consent decrees. I emphasize this point just because it was something that I thought was interesting to note in essence.

Daniel W. Linna Jr.

Giovanna, can I jump in and ask: were 72 cases the full population of a certain type of case from September 1, 2005 to November 30, 2019? Or was that just a sample of some of the cases? And if you sampled, then how did you sample? I'm just curious.

Giovanna Massarotto

Yeah, sure. We use the FTC database. And in that FTC Database, there are some filters. So, no mergers, no consumer protection. We focused only on antitrust and acknowledged that they include mainly FTC consent decrees because this is the largest practice of the agency. But I'm happy to double-check and assess. Thank you so much for raising this point. We started this project years ago. But to my knowledge, we include the cases using the FTC website and looking at Section Five, no mergers in that timeframe. Thank you very much, Daniel.

Thibault Schrepel

I have a question for you, Giovanna and Ashwin. And then a last question, which appears in the Q&A. First, there is a question as to whether there is an industry standard for clustering techniques, and whether they could be used to structured databases that exist already. I believe that yours may be working on a structured database with EU antitrust cases. And so... how easy would it be, in a sense, to translate your methods to EU case law? And if I may just take 20 seconds for a public call, something that would be very helpful would be for the European Commission or another institution to create a database with all the cases coming from EU member states, which is something that we don't have. Of course, the language barrier will still exist, but it would be nice to have. And I see no reason why this doesn't exist. I just wanted to put that on record. But going back to my question, would it be possible to use some of your techniques and translate that to EU competition law?

Giovanna Massarotto

From my point of view, yes. Ashwin is much better than me in responding to technical questions, but what I envision is, as you mentioned, the construction of a data set that enables antitrust agencies to collect data, and if they want in the future,

to run a machine learning algorithm with the data collected. They should start with the new cases that they are working on. So... this is really something that I hope the agencies will start. No matter what techniques you use, I think just the data collection will be extremely helpful. The FTC is very good, because individuals and also other antitrust agencies can find many public data. And also, you know, they will create a dataset that is more reliable and also will help to avoid inconsistency.

Ashwin Ittoo

I will try to answer your question and at the same time address a few questions on the chat. I would like to read one of them: whether we should use data analysis based on theory.

Thibault Schrepel

That's the question I wanted to ask next.

Ashwin Ittoo

So how does this apply to the question on the EU jurisdictions? First thing that we need to do is to understand how can we formalize a case? How does antitrust, competition law in the EU represent the case? And how can we operationalize variables to represent that case? In our study we selected 16 variables that we deemed best described a case. For the E on jurisdiction, I think the most important thing to do is to understand how is a case characterized? How do antitrust enforcers characterize a case? So... we try to get into their head, try to see how they formalize a case in terms of variables, how do they structure it, then once we have that structure in mind, I think that the same algorithms that we use can be applied to any data set that is at hand. Except, of course, if we need to tell it that we need to deal with textual data, because those data demand different kinds of approaches. But initially, I think we could apply the same algorithm to numerical or to categorical data, regardless of the jurisdiction which they come from, as long as we can formalize a case correctly. And there was one more question on the chat, which said that we applied our date our algorithm to known cases and asked what happens to new cases. This is also a very good question because machine learning, especially unsupervised learning, is trained on past data. In unsupervised learning, we learn patterns from past data which have been created by human beings; algorithms cannot learn anything on their own. By themselves, I mean from scratch, they have to be fed new data. That being said, if we had to detect novel practices, one approach, using what we have built so far, would be that given a new case, to take that new case and see to which cluster it belongs. If it doesn't belong to any cluster, then it might correspond to a new type of practice, which has not been seen before. So... this is one way in which we could do that. I can think of other ways now, but this seems to be a promising way to address this issue.

Thibault Schrepel

I will give you two minutes so you can wrap up, but first, there is a last question in the chat whether we can work with existing economic models. I think this relates to the question of interdisciplinarity of such projects. I can tell you that we have invited computer scientists to write articles for us and you see that the way they approach antitrust and especially merger control is totally different from the way an economist trained in the field of antitrust approach the issue, and even more from antitrust lawyers. So... I think the two might be complimentary. But my question to you all is whether, when working on computational antitrust projects, you have ever worked with non-computer scientists, for example, a philosopher, or someone specialized in sociology?

Giovanna Massarotto

I'm particularly lucky, because I had the opportunity to work in an economic consulting firm, in addition to law firms, and to work also with people with different backgrounds. This experience with Ashwin was particularly useful because we built a proof of concept. Honestly, I love the idea of having philosophers. I always try to study more physics; my passion is linguistics. And in my spare time, I study linguistics. So... it's really something that I will encourage other people with different backgrounds, because they can provide different perspectives. Now, the last question about a new kind of conduct that might raise. Honestly, if you think of a new kind of conduct, in a sense, the antitrust conducts remain the same. It's just that they find different ways to perform, to engage in anti-competitive practices. But yeah, so far, the rules are the same. We don't really deal with new antitrust practices in the legal sense.

And now with respect to the bias. That's a very interesting question, as the comment of Daniel, related to the aspect of why antitrust will benefit from AI systems. I think that antitrust has a great responsibility because it's the first arm of government regulation. So... this is why I think this is particularly compelling. It is also compelling to see if these techniques raise bias because that means that there is a bias in the enforcement of antitrust law.

Thibault Schrepel

You have 10 seconds.

Giovanna Massarotto

Okay, so thank you. Thank you very much to all of you for your amazing questions, and for keeping the community full of thoughts. I truly appreciate that. We will work on your comments and questions, and we hope to see other proofs of concepts, other projects in this way. So... thank you.

Thibault Schrepel

I'm going to try to nudge them to write for us. Ashwin, the floor is yours for a short conclusion.

Ashwin Ittoo

I'll try to be very brief. We see where the field is evolving. And we see that AI has the potential to create much welfare, but also much, much harm. So... it needs to be very well regulated. But as we said, too much regulation kills innovation. We need to work on the fine line between too much innovation and too much regulation. AI is a wonderful tool, but it needs to be properly regulated.

Thibault Schrepel

Daniel...

Daniel W. Linna Jr.

Thorne's point is a great one: how do we find approaches where we can combine, use what we already know to improve these tools? Of course, what Giovanna and Ashwin are doing as a proof of concept is showing how we can use these tools. Again, I really appreciated the work, I love to see computer scientists and legal experts working together. And yeah, we've been working on projects of court data projects, for example, here at Northwestern, where it's so great to have people not just in computer science in law, but from other fields, sociology, journalism. I think we need to see more and more of that work. It's difficult to do it in the academy, and most of us are not incentivized to do that work. And then these conversations, we need to have more of them, so that we learn to have better conversations and develop the vocabulary and the body of research, so we can produce more and more of the work that we need. Thank you for organizing this, Thibault. And thank you for your research, Giovanna and Ashwin. Glad to be a part of it.

Thibault Schrepel

Well, thank you very much. David?

David Colarusso

I just want to say thank you, the coding part of getting the structured data out of the things is great. And I'm just going to put a plug for the code and data; I feel like it could solve so much confusion.

Thibault Schrepel

All right, thank you very much. I will just conclude with a few words. First of all, thank you very much for staying up late with us, for those of you in Europe as I am. We have a packed schedule tomorrow. In fact, we'll be together for four hours. You'll hear from MIT, Chicago, and Toronto academics on how to improve merger control by studying the tech which is being acquired, instead of just focusing on the turnover and the market shares. You will also hear from a computer scientist regarding the inside of the Microsoft case and how to automate remedies. And finally, we will have presentations from three antitrust agencies, which, I believe, is a world first. They will explain which use they make of computational tools in 2021. See you all tomorrow and take care. Thank you very much. Bye.

Day 3: Computational Antitrust in Action

Transcript by Maura Margaret Carey

#1 – Merger control — Time for a New Antitrust Era

Thibault Schrepel

Welcome to the third and final day of the Computational Antitrust Conference. If you were not with us yesterday, we discussed concrete projects using computational tools to improve antitrust law. We dived into natural language processing, data visualization and other techniques of machine learning. To make an analogy with the world of cinema we ventured into movies like Inception. Today, we are moving into an action movie. I haven't watched the latest James Bond movie, yet I prefer OSS 117. If you haven't watched it, you should. But I'm convinced that Daniel Craig will like our script very much. First of all, we will explore how computational tools can help better predict and detect killer acquisitions. Second, we will explore how to automate merger procedures to address under and over enforcement in the space. Third, we will be joined by a computer scientist to discuss the unknown insights of the Microsoft case US version and how using computational tools could have made a big difference in that case. And fourth, three top officials from Brazil, the UK and France antitrust agencies will join us to present their work in the space. As on Monday and yesterday, I will remind you to please ask your questions using the Q&A function. We will do our best to address all the questions. If you have something nice to say, but only if that is nice, you can use the hashtag computational antitrust on Twitter. I will put that in the chat in just a minute. Last, make sure to subscribe to our newsletter. The link will appear also in the chat in about 20 seconds. And I will put later on today, a link to a Google form so we can get your feedback on the conference. Now it is my pleasure to give the word to Jan, who is our Academic Outreach chair, and who will be moderating the first discussion.

Jan Jakub Przerwa

Thanks very much, Thibault. And thanks so much for the introduction. Very, very grateful to Thibault and Stanford for the opportunity to moderate this panel. I'm a student at Stanford currently and prior to coming here I worked a bit in private practice on merger control. So, I'm very, very excited to discuss the future of merger control in this panel, and we have a great lineup. We're joined today by Professor Alex 'Sandy' Pentland, who is a Toshiba professor at the Massachusetts Institute of Technology, where he directs the MIT Connection Science Initiative. We also have Sandro Lera, who is Assistant Professor at the Risks-X Institute at the Southern University of Science and Technology in Shenzhen in China and a visiting researcher at MIT. We have Robert Mahari, who is a Ph.D. student in the Human Dynamics group and a JD candidate at Harvard Law School. And we also have Professor Bill Kovacic, who is a Global Competition Professor of Law and Policy at the George Washington University Law School and Director of the Competition Law Center, and also a visiting professor at King's College, London. Without further ado, Alex, Sandro, and Robert, I'll give you the floor.

Robert Zev Mahari

Perfect. I will share the slides. And hopefully everybody can see those.

Thibault Schrepel

Yes, it works.

Alex Pentland

I'll kick this off with a little bit of background and give it over to Robert and Sandro. I'm at MIT, and so... why is MIT doing computational law? Of course, at one level, it's obvious: code is law. But law is also code, and in fact, law is a special type of expert system for the most part, and so it has network properties. It has all of the sorts of properties that you see in the scientific literature where you have citations and things, and this is the stuff that Robert is working on. But perhaps more importantly, in this case, the sort of general framework for law is one where new rules happen very slowly. They often have implicit context rather than explicit contexts. We naturally think about things like markets more than we think about things like networks, and yet today, most of the fastest growing and most influential new companies are networks. And those behave very, very differently than traditional markets.

How? Well, one is that they use preferential attachment. In other words, the ones that are most visible get the most new customers without doing anything. It's just the top of mind. So a basic, mathematical function that derives from the fact that we have limited memories and limited attention. We can only name a few companies or resources at any given point, which drives preferential attachment. And what that means is the rich get richer, but not in the normal way. This is in a way that is exponential and explosive. And, that dominance is what we're seeing today. And it behaves very differently than market dominance. In fact, market dominance, you know, all the players are somewhat even in terms of the sort of mathematical equilibria, the Nash equilibria, things like that. But the paper that we've put out in the National Academy of Sciences proves that actually in network pings, you get a different dynamic, and that is that the dominance is more a function of the competition over the small guys, the general health of the participants, than it is of a small number of largest participants.

And in fact, if you do something that is conventional, antitrust policy, where you cut the head off of one of these big companies and you break it up. It's actually worse potentially than doing nothing. Because each of them being so large already, so far up the network dynamics will grow to be another monster. So, you're really hurting competition. Particularly the small guys by doing traditional sorts of antitrust things with network organizations. What can you do about this? Well, there's some simple things that you can do. One is that you want to discourage explosive growth. We want to make them basically pay their fair share, which is one way to think about it. But more importantly, you want to have nascent companies grow. That's how you actually reach an equilibrium which has the greatest innovation.

Let me give you a simple example that I like that is far from the only way. Imagine that I bought a certain number of network customers and if the network reaches more than a certain number, you begin to pay an infrastructure tax. And that infrastructure tax goes up not as a number of customers but as the square of the number of customers because the value of a network is generally proportional to the square of the number of customers because they can connect with each other as long as you're the center. What that does is that slows the growth of these monster network and it's called dragging's in physics. But you could use that same sort of tax to be able to promote local people, for instance, to be able to make the smaller customers with smaller companies, much more fit. And I like the idea of adding this tax on the big guys, that you turn around and use to give tax breaks to small local companies. And that way you're promoting the growth of small local companies, which has lots of other advantages. But also making the whole ecosystem a little more fit. And that's actually what you're after. The classic version of antitrust where you get rid of disproportionate ones by breaking them up is the wrong thing to do in a network economy. Instead, what you have to do is you have to raise the average fitness, and that's what you're going to hear about from Robert and from Sandro. Robert, will you take it away?

Robert Zev Mahari

Absolutely. Thank you so much, Sandy, and thank you all for letting us be here today. You know ,if we take a legal perspective, maybe a bit of a historical perspective, time and time again, US antitrust law has responded to the changing needs of the economy. In the 1800s, we had trusts, and the Sherman Act was passed to address those that was followed by a historic wave of mergers, and the Clayton Act was passed to address those. And so, we kind of asked ourselves, well, in a networked economy, individual economy, what are these new antitrust needs? How do we respond to new ways of competing?

We really came up with two concrete proposals. The first relates to monopoly power under the Sherman Act. So right now, what judges look at is price control and market share. And those are used to assess whether or not a company has reached a state monopoly power. But what we realized is that those are primarily symptoms of anti-competitive behavior. And you can exert anti-competitive influence long before you have a majority of market share, and long before you charge monopoly prices, if you ever do, and the way you can do that is by controlling data. And so, what we advocate for is an expanded definition of monopoly power that takes into account data control. And in the paper, and we're happy to go into this during the Q&A, we go into different types of data ownership models, and we note that some ways of owning data raise more concern than others and in particular, we advocate for this idea of data trusts as a way of owning data that can be done in a way that doesn't rise quite as much of an antitrust concern.

The second recommendation we made is related to merger review under the Clayton Act. As Sandy was explaining, there's a lot of empirical data and a lot of network science that makes us challenge this intuitive idea that we should be going after the very biggest M&A transactions and stopping the concentration of market share at that level. What we really need to worry about is today's dominant players

buying up tomorrow's competitors and this can happen completely under the radar in this term called killer acquisitions, because the transactions are small, too small for regulators to take note. And so, we focus a lot on the idea of fitness and Sandro will explain what that means in quantitative terms in just a moment, but at a high level what we want to do is break out of anti-competitive feedback loops by thinking about the market as a whole and not just the very largest companies. And so, if I leave you with one thing for today, is that until now, antitrust regulations have been very backward-looking. It's been a reactionary discipline and in the increasingly fast paced digital economy, what we're hoping to do is to move towards a real time approach to antitrust. With that, let me hand it over to Sandro.

Sandro Claudio Lera

Thank you, Robert, and hello, everyone. It is my pleasure to explain to you how we measure these monopolistic tendencies in the data. In the context of firm sizes, US law states that the growth of a firm is proportional to its size, and the proportionality constant, which is essentially just the normalized percentage growth rate is what we call the fitness. Qualitatively speaking, fitness is just a measure of how fast the firm is able to grow, how innovative it is, or how well it can compete. And the key insight that we derive from network science is that the growth rates distributions rather than the size distributions are relevant to predict monopolies. To sustain an economy where there is no firm that completely dominates, it is important to have a lot of high fitness firms which are able to compete. It becomes problematic when there is only a handful or even just one company that exhibits high growth. But if you have a lot of firms, all of which exhibit high growth, competition will predictably emerge. It's really about growth, rather than size, and to show this, we have analyzed the acquisitions of startups with data from CrunchBase.

Here is what you can see along the x-axis which show firm fitness, which we have proxied by the growth rate of follow-up investments, but there are many different proxies for growth and for fitness, and we have tried several, all of which lead to qualitatively similar conclusions. And along the y-axis, we just show the relative frequency. So now, if you first look at the gray lines in the back, these show the base rate distribution across all firms. But then the dashed lines on top indicate the distributions of the acquired firms. And well, maybe it's not so surprising, yet certainly concerning to see that the high fitness companies are acquired much more frequently. And from an antitrust perspective, this is problematic it's these acquisitions that involve startups that could compete later on, think Instagram, for instance. These acquisitions gradually deplete the pool of high fitness firms, and unfortunately, they often go unnoticed.

The main message here is that it's really about growth of competition rather than size of individual firms that are the concern, and regulators should focus on sustaining a high fitness, high growth environment, rather than just focusing on the few largest firms. To reiterate what Sandy and Robert already emphasized, punishing the largest firms is just addressing the symptoms, it is not addressing the root cause. And with that, I would like to hand it back to Sandy to wrap up. Thank you.

Alex Pentland

Our argument is, as a function of this different structure of the network economy, is to be able to have different metrics that are involved rather than just market share and to promote greater innovation. And, a lot of it comes down to controlling data. The data is the elixir by which network economy companies grow. A number of customers, things to know about the customer, the relationship you have with the customers. And what you need to be able to do is maintain and encourage decentralization of the data, rather than allowing the concentration of the data. And that's part and parcel of big shots acquiring startups is that by acquiring them you get a centralization of the data, which gives greater power to the corporation to grow. And what we want to be able to do is examine the acquisition of highly fit small firms and encourage merger structures that tend to preserve fitness. If the data is the thing that allows this explosive growth to happen, and if you want to buy that firm, you ought to have ways that put that data into a trust, that allow many corporations to use it under a controlled sort of governance and not be something where what they're doing is they're locking away this valuable resource. Happy to say more about it. And I see we're sort of out of time. So why don't we stop there and get to the next piece here, the Question and Answer and comment?

Jan Jakub Przerwa

All right. Great. Thank you so much. You're right on time. Professor Kovacic, would you like to comment? I think you're on mute.

William E. Kovacic

Thanks, Jan. It's a delightful paper. I think it fulfills all the hopes of the creation of the project, and congratulations to Thibault and the computational project for providing a forum to do this. And I think the basic observation of the paper is that you need a system that's fit for purpose and where you have extraordinary dynamism in commerce that has to be matched with policymaking dynamism that is responsive to the concerns that commercial developments raise. It would be a significant departure from existing practice. And I guess one way to think of it is that in the coming 12 months, we'll probably see a rewrite of the government's guidelines involving horizontal mergers and vertical mergers.

And I guess a question one context for the authors to think about is, are you confident enough about the approach that you would build that into the new guidelines? That is, that you would have sections in the guidelines that say, "We are ready to put this into practice. And we're ready to go with it." That is, is it robust enough, rugged enough in your thinking, to be worthy of being applied right now? Because that window has been opened. Alongside of that, just one more thing to think about is, an impetus for the current review of the guidelines is that those large dominant firms you mentioned are by themselves an appropriate subject of concern that the purpose of the law is in some way to take existing concentrations of power and then to subject them to stronger control. But the opportunity to think about new approaches is coming up right away.

A couple of questions and comments about the methodology. In shifting the focus on how we measure what's important, and insight in the paper is that you have to measure what matters in order to come to an appropriate conclusion. How are we going to define when firms have reached a level of market significance that we call them dominant? What level of control over data becomes significant? Well, how do we know when we've got there? Are we going to use traditional market share measures as one screen to identify who a dominant enterprise is? There are references throughout the paper to dominance and data share. What data shares are significant, and what level of individual firm significance within a network is important? I think that it would be very significant to define operational techniques for applying the approach.

Another question is, how we're going to measure fitness and how for purposes of drafting guidelines providing an operational methodology, will fitness be defined? What criteria will we use? Are we going to use highly qualitative criteria? Will we resort to the kinds of quantitative measures that have been seen important in drafting guidelines as a way of giving guidance to the agency staff and to the courts themselves? A couple of thoughts about some of the implementing mechanisms. There's discussion about being able to exempt transactions through the refinement of the Hart-Scott-Rodino merger notification process. The exemptions you're describing only apply to the notification obligation. The FTC and the DOJ do not have the capacity to exempt transactions from review all together. They would not bind a decision by state government to use a judicial approach. It wouldn't bind a private party. That is, the agencies don't have the capacity under the existing framework to literally exempt transactions from review. They have the capacity to determine who must file as part of the notification process. Do you have in mind a broader framework where the agencies could, contingent on the adoption of certain alternative methods for data control and sharing, would be exempt from review? That would be a very significant change for the framework. Right now the agencies do not have that capacity.

Another is how confident are we about the operation of data trusts? Do we have a body of experience that gives them confidence that these will be seen as reliable mechanisms that satisfy expectations about data protection and privacy? Would we see that in some way, these have to be certified or approved by the public authorities as mechanisms for, as repositories for the information? Would there be a further review about the terms on which they provide access? Who's able to gain access to the data set? If we had enough experience with data trusts that, suppose in the coming year we're going to adopt a policy that we are confident in going ahead and relying on them in the way that paper suggests we might? Or is there a way to think about doing this by establishing a prototype that perhaps takes some subset of transactions, takes a single sector, and says we're going to test the different parts and see how they work?

A final thought about the tax break and the allocation of subsidies. Are you confident that the administrative machinery to do that will do it in a way that sends the money to the companies that need it the most? Will there be in the structuring of that a perceived political imperative to distribute funds in a way that satisfies a

broad set of political expectations? Who would be the administrative body that would do that? What criteria would they use? Tax is not my field. I simply wonder how the allocation mechanism would work in that instance. I think it's a wonderful way of thinking about the problem that is not featured in common discussions. Current discussions focus almost entirely on setting absolute caps and prohibitions by the acquisition of dominant enterprises, firms above a certain level of market significance. Simply an absolute prohibition. It deals with the hydra problem by saying you will not, in any sense, be able to make certain acquisitions. This is a significantly different approach, and I think a very fresh addition to the overall debate.

Those are my thoughts about implementation. And I guess, if you were thinking about the introduction of it, would you be confident in putting it in place next year in writing new merger guidelines? Are there areas that you would want to test more before you put it into play? Because the government's analytical framework is going to be retooled over the next 12 months. But I'm most grateful for the chance to read the paper. And again, I think this fulfills the promise of the exciting approach that is bringing very fresh new thinking into the field. Thanks.

Alex Pentland

Well, why don't I kick-off? You asked a lot of questions which are great questions. I think in terms of the description of the dynamics and what the problem is, that businesses don't just cut off the head of one Hydra, I think we're very confident about that. We see that in all sorts of domains. The mathematics of that is very clear. This is not just about these particular examples; you'll see it in all sorts of situations. And the question then becomes as well, what do you do about it in this sort of situation where you've got particularly focused on companies that are network companies that thrive on data? I guess that the thing that strikes me is there's a mental shift that needs to happen. People talk about market share. They talk about sales volume, but nowadays, data is a primary means of production, like money, like labor. This is antitrust, to me, antitrust about data, data monopolies.

And the question then is, does an acquisition create a monopoly within an important data segment, population segment like, young people under 25 or something? And to the extent that they do that gets to be problematic. I think that the thing that we're less confident about is the remedy to that—so all of that, I think, is pretty clear—it's interesting that it intersects with data privacy, data ownership, and all those sorts of questions that are also busy being rewritten. And I think that your suggestion about having a demonstrator within a segment or a prototype is probably a good one for how you actually carry it out. Because this is in some ways, new territory.

One thing about data trusts is I imagine them to be things that are not owned by the companies involved but rather much more public facilities with government oversight, otherwise, I think it misses the whole thing. The other thing I think that's really important to realize is that data trusts like this, as a general principle, should not ever share data. What they can share is statistics about data, statistics about segments of data, and you can do this in a way that doesn't threaten individual

privacy. And what that does is that makes for a much more liquid market in data insights, but without exposing the data to general distribution. So that's sort of what I want to put in. Robert, Sandro, please jump in. I know you have thought about different aspects of this.

Robert Zev Mahari

I mean, I think the key question about our confidence in you know, implementing this in the next 12 months. I think that, in our group and the way we think about these regulations, it's imperative to experiment, right? It's imperative to actually collect data and figure out like whether the things we think matter, matter. The idea of some sort of regulatory sandbox and finding an area where we can put some of these interventions into place and actually measure the outcomes, which requires measuring outcomes and knowing what we're looking for, is key. And I think we would be reluctant to say, you know, we've figured it out, and here's the solution, but rather, what we're hoping to do, it sounds like you were able to do, is to kind of offer a new perspective and also bring in our expertise in network science and some other disciplines to look at antitrust through a new lens. So yeah, I hope that that's helpful.

Jan Jakub Przerwa

Great, thanks. Thanks a lot Robert, Professor Pentland, Professor Kovacic. We have a couple of questions from the Q&A. I think we might have one about data trust that we just discussed. There's a question from Caroline Thomas, I think to the authors of the paper but Professor Kovacic, feel free to jump in with your thoughts as well. The question is, do you think we need to address data problem using Silicon way, and what do you think about the solution that the data should be owned by the big subject, is stored on a blockchain, with the owner able to give and retract consent to use it?

Alex Pentland

This is probably my key area. Let me answer about that. Ownership is a funny concept when it comes to data because most data is co-created. Clearly, there is the notion of personal data, which is sensitive and both the California law and the new law give people control over that, but not sole ownership over for instance, aggregates of that. Aggregates are often owned by the company. Putting things directly on a blockchain is silly. Blockchains aren't meant for that. You can have data storage which is pointed to by a blockchain, that's fine. But that doesn't get to the point of who controls it and how and the point of data trusts which are attracting a lot of attention is that data is complicated. The situation changes.

Individuals like myself who are presumably experts can't handle it. It's too complicated. We can't anticipate it. Like a bank, like a credit union that's owned by its members, that does not take money from its members but holds money for its members. You can imagine a data trust that is owned by the members. It holds data for the people and carries out the sensitive, continuous task of establishing policy

and use and remuneration for data. I think that idea that data is as important the money and needs the sort of infrastructure that money has is a key insight, you can't just say, “Oh, everybody, take care of your data.” It's just not going to work.

Jan Jakub Przerwa

Great, thanks so much. I think we have another question about nascent competitors. I think we can allow one more question, and then we can also have one or two minutes to get final thoughts before the end of the panel. The first question is, I think we can combine the two, it's from Fabiana Di Porto and from an anonymous attendee. We agree that nurturing nascent competitors is a great idea, and small businesses are often fed by the data, and their growth might depend on constant data flows that would come from the hydras, as you call them in your paper. I think question two, to the authors of the paper is, would you also advocate for rules that foster data flows from big to small businesses?

Alex Pentland

Well, I'm sorry to keep answering first, the suggestion of the data trust is that we would have the ability to share insights, stratifications, statistics about things but not the data itself. And if you imagine a world where sharing of those sorts of insights is creating a market in the traditional sense, then yes, small businesses would have the ability to know what Facebook knows about your neighborhood which we currently don't. I would like to know about what the hospitals know about my neighborhood so that we can get better public health or what is known about our neighborhood so we can hold the government accountable. I think neighborhoods, local institutions need at least insights about their neighborhoods that they're currently not able to have, because it's siloed away, not only in companies, but in governments and other institutions.

Jan Jakub Przerwa

I also think in this area, a question to Professor Kovacic, would be that from one of the attendees. Many startups actually rely on being acquired by a big player like Google or Facebook. How would you think that the proposed approach in the paper could affect this trend and potentially have an impact on innovation?

William E. Kovacic

You know, the authors can address this more directly. My interpretation of your paper is that you're trying to address that specifically, you're trying to maintain an environment in which you would have a continuing market or smaller enterprises in which larger enterprises would have the opportunity to buy them subject to some of the reforms you have in place. And that, in that sense, I see the paper as being more sympathetic to concerns that have been raised about what you do, if you simply slam the door shut on acquisitions by larger enterprises of smaller enterprises. And my interpretation of the paper is that you're trying to create an enabling environment in which those transactions can continue but are subject to

alternative policy approaches that ensure that it maintains a porous, open system in which other enterprises are going to be able to come in and thrive.

Sandro Claudio Lera

Yeah, that's certainly one point, if I may add one or two more remarks here. The question is for the motivations behind the startups. First of all, let me add that certainly, there are some other motivations like obviously IPO which is probably the Holy Grail, but also there are some other firms which just can remain private for long term and pay off for the VC investors. There are certainly other exit strategies rather than being acquired, but we really would be also interested to hear from others about certain ideas, how large companies and small companies could cooperate in other ways than just immediately acquiring a company as a whole.

We know we work with a Swiss company, a large telecommunications provider where they often work with local Swiss startups, but rather than just acquiring them, they have a special legal structure that allows them to become minority shareholders in these companies and to interact with them and have them grow alongside each other rather than just immediately acquire them early on. There are certainly other structures, and maybe there is some more innovation to be done in this area, also from a legal perspective, to make sure that there are other ways too for small firms to interact with mergers beyond the acquisition.

Alex Pentland

Some years ago, I was on Google's high-tech advisory board, and I kept telling them they ought to break up the corporation and become something like they are today, Alphabet. But the additional thing I said is they're going to have to control data flows between these startups. Because it doesn't make any sense to have a distributed structure like that where the data is all pooled. But if the data were actually things where it stayed in the startup, and only certain sorts of value and insights and statistics were back and forth, it could be perhaps publicly audited, then I think that makes sense. It's something more than just a holding company, because it's project of, say, a cooperative, but of independent players. As someone who's had small companies, I wouldn't mind entering into a special relationship with one of these huge companies. But you want to maintain autonomy and you want to maintain privateness of the data that you have.

Jan Jakub Przerwa

Thanks, thanks so much for these answers. We have four minutes left, so maybe, Professor Kovacic, would you like to take two minutes for closing remarks and then also two minutes to the authors of the paper, and we should be right on time to conclude the panel.

William E. Kovacic

If this were an opera, this would be the point in which we give flowers to the performers. I think it's a wonderful contribution. I think this is a perfect time to do it because the field has opened for a rethink about how it functions. This is exactly the time to do this kind of work. And I guess in the paper, I mean, just a couple of thoughts again: at what point is the ownership of data, is there a critical mass that it becomes important? You can imagine a smaller amount of data, a larger amount of data, for purposes of identifying when it's significant: where is it important? At what point? And I think the other thing you might reflect upon in a further revision of the paper is, how much of this can be rolled out? Maybe without awaiting the creation of the other institutions. Could you start thinking about this network environment and the way you do as another way of appreciating the market significance of individual firms and what's taking place? Is that something that could be used right away to think in a different way about what's taking place in these sectors?

I mean, I see in a way the again, the data trust as being a technology that hasn't been built yet. Perhaps in the way that it would have to. But I think what you've done if you think about that the Apollo program, there came a point where an engineer said you don't want to take one big heavy thing, land it on the moon and send the big heavy thing home. It was the idea that you could throw stuff away along the way and send back a much smaller compartment. In order to do that you had to create a Lunar Excursion Module; it seems to me like the Data Trust is that technology that hasn't been built yet, hasn't been tested. Maybe that's further down the road as that is another thing to be done. But you could decide that you're using the general approach to go to the moon and also start working on these other institutional reforms. Maybe some notion about that, again, the sandbox prototype, I think, a further reflection on your part about what you can roll out now, what you have to start working on if you want the larger vision to be accomplished. But again, it's the right time to be thinking about a new approach to measuring what's important and addressing it. So, it's a wonderful paper.

Alex Pentland

If I had to add that one thing that you can do right now, think about data, not just market share. That gets you maybe 50% of the way there. Robert, Sandro, final word?

Robert Zev Mahari

I think I want to reiterate the idea about real-time antitrust and how as we move more and more into this network economy, I think it's going to be harder and harder to react retroactively. And the price we pay for waiting for dominance to be achieved is going to be higher and higher. I think this is a central motivation, can we spot antitrust early and not bear the costs of anticompetitive behavior? So, thank you so much. This was great.

William E. Kovacic

If I could come in with a final thought on that. I see one, and it's an agency I'm affiliated with now. I'm on their board as an outside director. But I know of one agency on earth that is creating the internal capability to do the real-time analysis. That's the Competition and Markets Authority in the UK, which has assembled a data and analytics team of 40 people now, computer scientists, quants, people who can do this kind of work and work with the data and talk about what's going on. Many other agencies are only creating the capacity to do this kind of real time analysis. If there's one place on earth that could do it well, it's the CMA. Do it well right now, in real time.

Jan Jakub Przerwa

Great, thank you. Thank you so much to all the panelists for this discussion. It's been incredibly interesting, really important and really timely. As Professor Kovacic said, it's time now to think about these things. Thank you so much. It's been a pleasure. Thibault, I hand over to you.

Thibault Schrepel

Yes, thank you very much. A few thoughts. I think you do have the title for a book, Robert here, "Real Time Antitrust." I promise you I won't be using it, although it's very tempting. And you know, it's something that lawyers complain a lot about that we lag behind the tech, and I sometimes wonder if we really want lawyers to be ahead of the tech. I'm not sure if that's the solution, but to be right behind, I think might be something we may want to experiment. Bill, you mentioned the moon. The question is, who's Elon Musk on the panel? I'll let you decide that one. For the CMA, luckily, we do have Stefan Hunt, who will come today actually to explore what the CMA has been doing. And so, maybe the occasion to learn more about it.

#2 – Micro-directives and Computational Merger Review**Thibault Schrepel**

I'm delighted to hear the next panel. We will bring the speakers to the forefront in just a minute. The person moderating the panel is Teodora, who is one of our editors, and I can tell you that she has been a central force behind this conference. Everything you like about the conference is pretty much her initiative. Teodora, thank you very much, and I see that all the speakers are here right on time. Teodora, the floor is yours.

Teodora Groza

Thank you very much, Thibault. Thank you for this very kind introduction. It's a great pleasure to moderate this rather unprecedented event, and particularly this panel, which will assess how micro directives can be used to address both the

under- and over-inclusive nature of the existing merger control regime. To keep the cinema theme that you kept bringing up throughout the conference, we can easily say that we have an all-star cast for this panel.

Our three speakers are Anthony Casey and Anthony Niblett, who authored the paper that will be discussed, and Daniel Greenwood, who will share with us his comments, questions and concerns about the paper. Now let me tell you a bit more about each of them. Anthony Casey is Donald M. Ephraim Professor of Law and Economics and the Faculty Director of the Center on Law and Finance at the University of Chicago Law School. Antony Niblett is an Associate Professor at the University of Toronto Faculty of Law and also the Canada Research Chair in Law, Economics and Innovation. For the lawyers among us, the Casey-Niblett duo is a household name. So, it's indeed a great honor to have them here with us. The same holds true for Daniel Greenwood, who just joined us, and who is the founder of civics.com, a boutique provider of consultancy services for legal technologies, as well as a researcher at MIT Media Lab, where he works at the forefront of computational law.

We're very lucky to have him here, indeed. For the next 20 minutes, the Casey-Niblett duo will present their paper. Please feel free to share your questions in the Q&A section at all times. Professor Casey, Professor Niblett, thank you for agreeing to present your paper. The floor is yours. You have 20 minutes.

Anthony Niblett

Wonderful. Thank you very much, Teodora. Would I be able to be made a host so that I can share my screen please?

Teodora Groza

Yes, of course.

Anthony Niblett

Thank you.

Thibault Schrepel

In the meantime, I promised myself not to intervene but Dazza, you promised me an MIT hat. I'm waiting for it.

Daniel J. Greenwood

Do a good job finishing the conference and a hat you will have.

Thibault Schrepel

Thank you very much [laugh].

Anthony Niblett

Okay, thank you very much Thibault for inviting us. My name is Anthony Niblett. I'm at the University of Toronto Faculty of Law. I'm also a faculty affiliate with the Vector Institute of Artificial Intelligence here at the UT and this is a co-authored paper with Tony Casey who's at the University of Chicago. Our paper is called: "Micro Directives and Computational Merger Review". And this term micro directives comes from a previous paper that Tony and I wrote a number of years ago, a number years ago called: "The Death of Rules and Standards". And this paper was a thought experiment about what's going to happen when predictive technologies get better and better and how predictive technologies will change the structure of law.

We investigate what happens when you have unprecedented an exponential growth in data, what happens when you have supervised machine learning algorithms get better and better at predicting and with this better predictive technology that will enable legislators, regulators, judges to make better law. Armed with better information, you can make more tailored rules. These court rules, these bright line rules that we currently depend on can be improved upon. Bright-line rules tend to be overinclusive and underinclusive, and you can improve upon that with better prediction. Standards tend to be vague and create uncertainty and are rife with bias and inconsistency. You can take those standards, and you can predict better. You can give those standards more ex-ante content.

And so, we said, well, you're going to move away from rules because rules have costs and you're going to improve upon rules. You're going to move away from standards; you can improve upon standards. And this gives way to a new form of law that we called the micro-directive. We were invited to submit a paper for the journal and this conference, and we thought why not apply these ideas about a micro directive to something in antitrust and those merger notification and review. Merger notification and review is something that has rules and it has standards and how can we improve upon merger notification and review? So that's what our paper looks at, it's how can we use these predictive technologies, supervised machine learning, etc., to improve merger review? And we start off by taking the extreme version. And we imagined: what would it look like if you could have micro directives for merger review, essentially fully automating the merger review process, that parties to a transaction could enter all their relevant information and receive an almost instantaneous response from the agencies.

What we conclude, though, is that there are limits here. There are a number of limits and you can think of the limits of being limits about data. You can think about the limits of technological feasibility; you can think about the limits of trust and legitimacy in such a system. But for us, the biggest limit here is a very human problem. And that is the difficulty in specifying what the objective actually is, what values do we want to promote? What is it that this artificially intelligent algorithm should be maximizing? And that for us is the biggest hurdle it's not to say that there aren't hurdles about data, legitimacy, trust, etc. Perhaps the biggest hurdle is the hurdle of specifying what the objective actually is. This is one way we'd think about

the current notification regime. This is the notification regime in the United States. And one thing you’ll notice is there are some very bright lines. There are bright line rules in a way you notify or not. And when you’ve got bright-line rules, this typically results in over-inclusivity and under-inclusivity.

And just to do a very, very simple way of thinking about this, you can imagine a mapping of competitive concerns from low competitive concern about a merger all the way to high and there are notification thresholds that are in place to screen out certain mergers. If you have a bright line rule, that could be fine for certain types of mergers where you’ve got high competitive concerns, and they’re above the threshold. That’s fine. Where you’ve got low competitive concerns and it’s below the threshold, again, that’s great. But these bright-line rules are unlikely to be predictive of competitive concerns and you’re going to end up with some situations where there’s a low competitive concern, but still required to be notified and this leads you to the world of over-inclusivity, that these resource-constrained agencies are being required to review mergers that are not likely to generate competition concerns, they’re “inconsequential,” and yet we still have to devote resources to looking at this. It also has costs for the merging parties. It might disrupt beneficial mergers. It might even chill beneficial mergers. So over inclusivity is a problem.

On the other hand, you could have mergers that are high competitive concerns, and yet they fall below the notification threshold, somewhat of which we were talking about at the previous paper that is under inclusivity allows potentially anti-competitive merges to fly under the radar. The current system is both overinclusive and underinclusive. These problems can’t be fixed by just adjusting the threshold. If you lower the threshold to try and capture some of those high competitive concerns, then you’re just going to increase the number of false positives. You’re just going to increase the number, the over-inclusivity and vice versa. The way to do this is to get better ex-ante predictions. To get better ex-ante predictions about what types of problems there are and what types of mergers require further review.

And so, we look at some types of predictive tools and we’d say, well, we can improve upon these notification rules we can improve upon the standards in merger review, trying to work out what substantially lessening competition means. Agencies could move away from bright-line rules in the notification process. You could also provide greater ex ante content about what these standards are. We’ve got substantially lessened competition here in Canada, but you could provide more content to those rate standards. In a fully automated system, what you would imagine is that parties could be told instantaneously, they get a red light or a green light as to whether their merger is deemed likely to substantially lessen competition. So, transactions of high risk are flagged immediately. Proposed transactions that are a very low risk are given the instantaneous green light.

Now, this system is not necessarily feasible with the current toolkit. A lot of the discussion in the previous paper was about how do we improve that toolkit. We discussed a couple of ways in which the agency’s toolkit could be improved to help push towards the direction of having more micro directives in merger review. The two systems we talk about use different types of data and they have very different prediction problems. The first type of data are internal agency data on which

mergers have been reviewed, which ones have been approved, which ones have been challenged. And then, a second type of data is external data on the consequences of mergers, and I'll talk about each of these in a bit more detail. But I think it's extremely important to separate the two, it's extremely important to separate the two because they have different prediction problems.

They solve different types of prediction problems. The first one if you're going to use internal agency data on merger approval, then what you're doing is you're predicting whether the type of merger that's been proposed would have created concerns in the past. That's really what you're doing. It's a replication in some sense. Then you've got external data, you're looking at the consequences. It's more like doing retrospectives but more fully fleshed out retrospectives. Where you're trying to predict the risk levels of various harms; increases in prices, reduced quality, reduced innovation, etc.

Let's talk about this first one first, internal agency data. How could we use internal agency data to improve merger notification and review? Well, we looked back at the numbers and we saw that there's, over the last 20 years, there's been about 30,000 notified transactions in the United States. And one way to think about this is you could turn this into structured data, you could turn this into structured data to provide predictions of how future transactions could be dealt with. And you might be thinking, well, that's, you know, a lot of work to put that together, you've got the data unstructured in that way. There are all sorts of problems with doing this. This is something that we have done in Canadian tax law.

And we have taken unstructured data from cases, thousands and thousands of cases, and we have turned them into structured data. These are data that describe every single case that's ever gone to court in Canadian tax law, on particular issues. I'll just give you an example right now that one particular issue is, well, if it is a worker, are they classified as an employee? Or are they an independent contractor? And that's a very vague standard, but we've turned thousands of cases on that issue into data. And so, you can come along with a new case, you fill out your information about what your case is about, information about what kind of work is it? How much control is being exerted by the firm, who has the freedom to turn down work, what tools are being used, etc. Who has the risk here? Are they being reimbursed? And then you can take all those facts and compare it to every case that's ever gone to court in the past.

And you can make a prediction, you can make a prediction that if this case goes to court, this particular worker will be classified as an employee. And this is just a prediction problem that is trying to replicate what's been done in the past. It uses a structured data set to predict what would happen with a new case, and you can imagine similar tools being used in merger review. You take the unstructured data in agency decisions. It's a bit of human effort to turn that into structured data that describe those decisions. But then when the next one comes along, you can make a prediction as to, well is this the kind of thing that has internally generated concerns in the past? Is this a benign merger or is this something that we should be challenging?

Now this has benefits, it has benefits in terms of the speed with which you can provide an answer. It has benefits in terms of consistency, it's not dependent on which human reviewer is given the task of reviewing the merger; the algorithm treats like algorithms alike. Essentially, you are replicating what the agencies have done. And that's for some people a problem because perhaps you think that the agencies have not done things correctly in the past and perhaps you think that the standards should be changed. Perhaps you think that with this growth in data by companies that that's opening up a new problem that we've got to address and so maybe we don't want to do this.

If your objective, though, is to give fast responses. If your objective is to get consistent responses, consistent with what's been done in the past, then this is a good system. But if you think that the law should be more dynamic than that, then this obviously is not a good system for you. In that case, then you want to use the external data that describe some of the consequences of mergers. And so, you're looking at predicting the risk levels of various forms, the increase in prices, reduced quality, etc. And in a number of areas of law, we've seen that researchers have been able to use big data and machine learning algorithms to show that machine learning algorithms actually improve human decision-making. There's a number of papers, for example, on Vale showing how machine learning algorithms are better than judges at predicting risk. And you can imagine in the context of mergers, getting data on market structure, conduct of firms, etc., will enable you to provide much richer and more accurate future predictions. In some sense, this is basically like taking the current system of retrospectives, and just fleshing it out to better understand what these errors are and better calibrate the law.

Now the current system could be thought of as a system of triage, that we're basically just trying to triage you know, if it's a small merger it's unlikely to generate any concerns. Predictive algorithms improve upon this. Improve upon this as a pre-screening tool, and you can imagine, rather than going to the fully automated system, that what we do is simply use this as a screening mechanism. For us, the biggest hurdle to a fully automated system is not technological feasibility. It's not the volume of data. It's not the type of data, it's not legitimacy, the biggest hurdle is spelling out what the objectives are. And these supervised machine learning algorithms make predictions based on a particular objective. And that requires us to be extraordinarily clear about what we want our merger law to achieve. What is the objective of merger law? Is that purpose translatable into an objective that an algorithm could maximize? Ultimately, for us, these algorithms can't solve the human problem of selecting the objective. It's like having a self-driving car. But if these humans don't know what destination to enter, the self-driving car is not that helpful for them.

And so, in our paper, we discussed how competition authorities have a variety of different objectives. It's not like we have a single objective for any antitrust authority. In Canada, for example, we've got all sorts of different objectives. We want to promote efficiency, expand opportunities for Canadian participation, ensure that medium and small enterprises have an equitable opportunity to participate, provide consumers with competitive prices and choices. Some of these

objectives are found in other jurisdictions too. In Japan, protect small and medium-sized enterprises; in China, protect and expand domestic enterprises. In South Africa, they've got a whole slew of different objectives based on the need to have greater ownership by a greater number of South Africans, to promote fair distribution of ownership among different racial groups, to balance the interests of workers, owners and consumers, and in Europe, enable the proper functioning of the EU's internal market for the benefit of citizens, businesses and society as a whole.

It's extraordinarily difficult to try and put that into one translatable objective. And so, the Americans in the audience might be thinking, well, our objective is much clearer, it's maximized consumer welfare. Now, that might be true. It might also – that objective might be changing even if the objective was to maximize consumer welfare. That's still very difficult to translate into an objective for an algorithm. Are we maximizing consumer surplus or are we maximizing total surplus? Or is it like in Canada where we're actually maximizing something in between?

Even if we are maximizing something in between, how much weight should we place on consumers as opposed to producers? When you've got multi-sided markets, there are important questions about which consumers are benefiting and how much weight should we put on each of those consumers? How much weight should we put on non-price variables like quality, service, availability, all those things that affect consumer welfare but are less easy to translate into dollar figures? How do we balance today versus tomorrow? How do we balance static efficiency versus dynamic efficiency?

Now, that's not to say that this can't be done on a case-by-case basis. In fact, that's exactly what is done on a case-by-case basis. The problem emerges when you try and put an ex-ante system in place. It would require agencies to be very, very clear about how they view these various considerations, how they'd make this a balancing force. At the moment, it's done on a case-by-case basis. The agencies have a great deal of discretion how to balance those objectives. But if we moved to computational merger review, that would require agencies to be extraordinarily explicit about how these different objectives are being balanced, and it would require them to commit to those balances and those objectives upfront.

My conclusion, Tony and I have been accused of being techno-utopians. We're not techno-utopians, but we do see that there's enormous benefit here. There's enormous benefit that can come from more data than from better innovation of predictive technologies to improve the merger notification and review system. There's no doubt that there are enormous benefits but the technology also exposes some of the limits of antitrust, some of the limits of computational antitrust, and some of the limitations of law altogether, that the biggest roadblock is not data. It's not technology. It's not legitimacy or trust. The biggest roadblock for us is what do we want the law to do? What is it that we want merger law to achieve? And Margaret Levenstein, in her talk on Monday said, "AI can help us make better decisions, but it can't make those decisions for us." And that's kind of where we end up with our paper. With that, I will hand it back to you, Teodora.

Teodora Groza

Thank you very much for the insightful presentation and for being right on time. I'm sure that the thoughts you shared with us will generate a very lively discussion both on how computational tools can optimize merger control and on the normative side of what objectives antitrust should pursue. To get the discussion started, I will hand it over to Daniel, or should I say, Dazza?

Daniel J. Greenwood

Both, but you could call me Dazza, just between us.

Teodora Groza

Okay, just between us, Dazza, I am very happy to give you the floor. And I must confess that I'm particularly excited to hear your comments since you shared with me privately that you've been a fan of the Casey-Niblett duo well before this conference. You have 10 minutes; the floor is yours.

Daniel J. Greenwood

Thank you so much, Teodora and Thibault and CodeX for putting us on. And yes, it's true. I am a fan. And we've actually published one of their prior works in the MIT Computational Law Report. And, I think right in tune with that, this paper is truly excellent. I believe that they're looking at the right things. Following on from Sandy Pentland and the team's conversation just moments ago, here again, we see a focus on data. And we see a focus on a real-time application that was a theme from the last panel as well. And we see a focus on more to the point now taking it one step further, which is why I'm such a fan of this and it's sequenced so well in the day. It's particularly applied to a very specific aspect of merger review. And so, the suggestion for a computational merger review scheme, I think, is nicely positioned such that it has academic merit, but it really begs to be taken to the next level in terms of action, and of course, action is the name of this track of the conference. But the reason specifically, is that by narrowing the context and the scope to this aspect of merger review, now we can have a tractable engineerable, testable, even capable of deployment kind of idea. This is particularly delightful, I think.

Let me just go down one level and make a couple of observations and then I actually do have some questions. So, I'm glad that the authors are with us. One of the observations is that, as specific as the paper is on the scope of merger review, and the possibility of a very practical benefit, among other things, of just shortening the time for review for easy yeses that just seems like low hanging fruits. It also reveals a major ambiguity that ought to be resolved—almost has to be resolved in order to architect and test a computational merger review scheme such as this, namely, as the authors just mentioned, the goals and the objectives. And I didn't really appreciate until I read the paper, how varied the goals of antitrust and competition law are internationally like just very, very different things idiosyncratic to the kind of social and political and economic priorities of the jurisdictions.

Even in the US, the goals are somewhat squishy, and they seem to be somewhat subjective or change in context, that is not a starting criterion for a system to be engineered algorithmically. It's something that Sandy says a lot, Sandy Pentland, but that we definitely take as a mantra at law.mit.edu is that we need to do law differently in order to successfully transition to computational systems in many respects but perhaps the most challenging one is way upstream with legislators and with regulators. For them to start to make laws differently such that there are defined sort of objectively measurable goals attached to the rules that they are promulgating, and I think that the authors do a very fine job of highlighting that as an open question. I think the challenge is to answer the question, as part of embracing and trying the approach of this paper.

I would like to, I guess, first highlight that but I guess the question would be, how might one postulate and test examples of high-level policy goals and lower level regulatory objectives, and the different priorities in ways that are capable of expression in objective testable systems? And ideally, how might one do so? I would think if it were part of law, not like in a very specific language like Python or something, but actually in a way that's more like a math syllogism that could be reduced to different code bases and different models and different algorithms and still be capable of independent verification that we're achieving the goal.

This is not a normal way to do it, sorry, I shouldn't say normal, this is not a customary way to do law. I think it's very reasonable and almost expected and I'm just going to stick my neck out to say necessary capability that we'll need to resolve not just in antitrust but in general, and yet it's an open question and the authors really put their finger on right on the pulse of it in their paper. And I hope that by doing so it can spark conversation at CodeX and beyond and in industry and government as well to start putting potential solutions to that on together. When I say solutions I just mean, how do you do it? Like, I'm not even talking about what should the policy be at this point.

One other thing I wanted to touch on was to go one level deeper on this wonderful insight about the need for data so there are a lot of assertions in the paper that we have all the data. I guess, you know, arguably, it's true that the data exists, the concept that we have it in some way that something could be done with it I think deserves a little bit more attention. And so, again, in the theme of this action track of the conference. I think it would be interesting to talk a little bit about what are examples of data sets that would be typical in business and in your standard enterprise systems for example, to maybe in the broad categories of pricing, market share, market size of the transaction HHI size of the parties. Like I'm thinking of accounting systems, I'm thinking you know, Salesforce, I'm thinking NetSuite, there are all these different systems that exist that you work on all day, typically as a consultant.

But what systems and what data could you derive, like through ETL and other wrangling methods, the kind of input data necessary for the algorithms that we're talking about? And to the extent that that's a bespoke idiosyncratic data science job every time that's okay. But you know, is that going to actually exceed the 30-day

benefit on this application area? And more to the point, how much value does it leave on the table if there aren't standard repeatable, widely deployed ways to collect and transform and make accessible this kind of data? Because, this juncture of merger review, certainly is a great use case, but there's a lot of law and regulation out there that would benefit from this type of data.

And by the way, I think the biggest benefit of having exactly the kind of data needed for merger review, is just doing business, like every time I'm on a team in companies that are trying to figure out a marketing thing or you know, different product questions and different strategy questions we're asking questions that could be largely answered by the same data that would be needed for these reviews. And it's always like scrambling the fighter jets and doing a data science team to get this together and wrangle it. So eventually, we're going to, I just want to just throw out there, they're going to need to be a way to have more standard approaches. And I'd love to hear as a start from this regulatory antitrust review perspective. What example data sets that are common in business would you think would be inputs, so we could just start thinking about that.

And in closing, I guess I would just ask this, which is in line with the last session's question about how soon this could be ready. My question would be, how could you envision deploying this is? Let's say an antitrust agency was interested in taking you up on this and giving it a try. One question I would ask is, how would you structure a pilot or a test such that you could determine whether it succeeded? What would be the metrics that you would want to test to know that applying this method was better than what we have now? And then how could you structure a pilot I guess, let's call it or a test of the system, such that it could provide that answer?

Those are some of my initial reviews and questions but overall, I just want to say well done. This is such a great paper and you're really pointing the direction to how to apply sometimes highfalutin seeming concepts of computational law in a way that can provide real benefit in the economy and in society. Thank you for this. And I provide the mic back to the conference.

Teodora Groza

Thank you very much Dazza for your questions and for your comments. Does any of the Anthonys or perhaps both want to jump in on any of that?

Anthony J. Casey

Sure, I just want to come in on the big picture point. I think it's super important. The idea that, and Dazza and I have talked about this before, that here and elsewhere requires law to be done differently. And I think this is a good thing and a bad thing. The bad thing is it's super hard, right? Like, we will see that defining the objective of law is a lot harder than people could've recognized before when you try to put it into a portable format or into you have the data tell you what to do. The good thing is, I think it's going to require a little more honesty about what law is doing. I think that's a good thing. The flip side of that is it will reveal a little bit of

the dishonesty that's been out there and be like, "Oh, this is what we're trying to accomplish". No, we're not. When we wanted you to put that into code, we said no, let's change it. On the good side, it does require that honesty about like this is why we have the law, but it's going to reveal that a lot of times we don't know or a lot of times, we're punting because we don't agree.

In the Q&A, Caroline had a good point of who makes that decision? It doesn't have to be the legislators. The legislators might punt to the regulators, and regulators might be the one who has to turn the fuzzy objective into something real, but someone does have to do at some point. And again, I think that's true in the antitrust field and elsewhere. And we're going to be seeing this throughout, you know, whenever micro directive becomes possible. People are going to say, wait, what was the law doing in the last two hundred years, and as we point out in this article, Antony and I wrote about it, in antitrust, there's so many possibilities and there's so much right now where people are talking past each other because they have such vastly different polls. I liked that Dazza pointed out that challenge, I think it's very important. Antony, I don't know if you wanted to add anything to that?

Anthony Niblett

I'll just add something to the calls for action. Just in regard to the different types of algorithms and different types of data and the different types of prediction problems that are coming out of this. And so, you know, the danger exists as it pertains to internal agency data; it exists in an unstructured form. There's effort that's going to have to be undertaken if you wanted to go down this route, to turn that into structured data. And what's the payoff, the payoff is you essentially get an agency decision quicker, but it's replicating what's been done in the past. And so, you know, you think about the 30-day waiting period, if you can get the benefit from that straight away, that's an enormous benefit. How much does it cost to put that interface? Well, we've done this with a small team in Canadian tax law, some of these cases have fewer variables or fewer features than some of these mergers but it's not exorbitant costs and it's possible to do this, to turn unstructured data into structured data to come up with at least a triaging system, for working out the probability of concerns with regard to the internal data.

With regards to the broader question about you know, what other data do you need about pricing, etc. And we talked about HHI and market shares. For us, the biggest concern here is actually not the volume of data and putting something into practice. It's about somebody tagging these things, tagging these things as this was wrong, we shouldn't have done this, we shouldn't have let this merger go through. Or actually, you know, we should have let this one goes through. It turns out that that's going to be the big problem, is identifying what the error was in the past. And if you can't identify what the error was in the past, then it's very difficult to actually have this mechanism for predicting the consequences in some sense. And so, you know, with regards to the bail papers that I talked about, there's a very clear objective that you're truly looking at flight risk here. What risks are we looking at here, and which one should we be concerned about? And that's the human problem of we don't

know where the self-driving car should go. It's very difficult. It's very difficult. If we don't know what the destination is, the self-driving car is of limited value.

Teodora Groza

Dazza, would you like to have a quick go at that?

Daniel J. Greenwood

Yes, I guess. First of all, thank you for those reflections. Anthony, I don't think I'm familiar with the examples you're describing in Canada. And in case I'm not the only one, those sounds very relevant and perhaps very helpful. You know, more broadly, would you just take a moment to describe those and maybe with a little extra emphasis on how they might be applied in the United States and other jurisdictions.

Anthony Niblett

Sure. I and a couple of academics at the University of Toronto have been putting together a startup that puts together turning on traffic cases in tax law infrastructure data. That allows lawyers and accountants to predict what the law is. You've come with your client's case, and you enter it into the system just like I showed. And then it compares your client's case to every case that's gone to court in the past, and it enables you to get a prediction and so, is my clients an independent contractor or are they an employee? Well, here's all the facts of my case. And here's how a court would likely, if the court acted the way that they've done in the past, here's how they would likely rule in your client's case.

This has been brought to use, the regulators use this tool as well to help with their process. You can imagine a regulator in an agency an antitrust doing something similar here. Turning all that unstructured data about their past decisions into structured data. You make a comparison, essentially, you're just making a comparison. What are we...? Here are our new facts from this particular merger, comparing it to the 30,000 mergers that we've had before. This is not likely to raise any concerns because that's what we've done in the past or this is likely to raise concerns.

We've got software, that's in tax law and in employment law, we do estimations on damages, for example. All these things use very simple machine learning algorithms, but the key added value is really turning that unstructured data into something structured that enables you to make those predictions.

Teodora Groza

Dazza, are you satisfied with the answer?

Daniel J. Greenwood

Very satisfied, and I'll just close by singing out the mantra that I hope we'll all derive from this: data-driven, measurable, model-based, algorithmic and adaptive. Let's make it so.

Teodora Groza

Thank you very much, Professor Casey, I saw you wanted to say something, so we can give you the floor.

Anthony J. Casey

I was just going to comment on what Anthony had said, and that gives you this like readily available sources of data in the past decisions, and it goes to the theme of, that's great if we've been doing it right. What whether, if we don't know the purpose is like, and you can't quite define it, but like, you know, let's just keep doing what we're doing. That's great. Of course, it also could identify problems with what we've been doing; you know, I'm stealing this from a colleague of mine at Chicago, someone named Nathan. He talked about the Amazon issue a few years ago when their hiring data was supposed to be doing what they've been doing in the past, and they realized that it was kind of having gender biased decisions and basically the data, the algorithm was using bad old data and perpetuating this problem, but the interesting thing was, what was their solution? They shut the data down.

But it just pointed out, and you've been doing this thing bad for all this time. And the solution shouldn't be, let's go back to what we were doing, because we just found out, because we automate and do what we were doing. It's gender bias. Right? Like, in some sense, that data is great if you like what you've been doing, but it can also help identify that you don't, but that's a new challenge that goes back to the now let's be honest, what do we want to do? And if we've identified that what we have been doing doesn't match that we've got to fix that one of two ways. You know, train humans or train the data, one way or the other, you've got to make a solution. And that's a step forward, whichever path you take because you're fixing a bad, the non-automated past wasn't working either.

Teodora Groza

Thank you very much for those insights, now we'll open the floor to the questions from the audience. And the first question we got was from Professor Bill Kovacic, who was on the panel just before. It's a half a question, half comment. He points out that one of the main asset of your paper is that it shows out how competition laws across the world seek to achieve different objectives. Some of them economic, some of them non-economic. Now, the question is, to what extent do you think that AI could cope with objectives which seek for instance, to protect the integrity of the political process? Does any of the Anthony's want to have a go at that? Or perhaps Dazza?

Anthony Niblett

So that's an interesting question, could AI cope with a system that said identify any mergers that harm the integrity of the political process? And I do not know of any way to predict that particular end. It is an extremely interesting question. Again, it would require people to know how to extensively attack something of, well, this is actually something that is concerning for the political process. And it would also require us to agree upon, what is concerning for the integrity of the political process. I have not considered how to put together that sort of algorithm. I don't know if Tony, if you want to jump in.

Anthony J. Casey

I'm not a voting rights expert, but I would imagine if you wanted to do that, you'd have data on voting and you'd have the data on how mergers between companies affected voting outcomes or voting participation. I'm just using that as an example of trying to identify the thing that is the purpose or the objective that the integrity of the political process, is it you know that these types of mergers changed voting behavior in the past. I don't imagine we have. We have anecdotal examples and people who talk about that, I don't imagine that we have massive data on how to do that on a merger by merger basis. Again, you'd have to really identify what you mean. Is it voting, is it the dialogue in public, what is it that you identify as the failure of integrity?

Anthony Niblett

The question actually raises something that I probably should have emphasized a bit more. You could have algorithms that highlight a particular problem: disruption of political process. You could have one that predicts prices, one that predicts equality etc. The question is putting them all together, as once you've got these additional predictions, it's how do we balance those things? And that's really the main concern. You could have algorithms that make all these partial predictions, predictions that are only partial, they're only part of a problem.

Teodora Groza

Dazza, how do you feel about that?

Daniel J. Greenwood

I tend to agree. Basically, I guess that the one thing just rattling the back of my head, I hope it's helpful, is just that whenever I see the phrase AI and then attached to these big questions, I always just feel it's good to take a moment and take a closer look at or like maybe not even use the word, the AI or the acronym, and maybe just talk about technologies and systems that exist, and then a lot of the presumed capabilities, narrow considerably and become a lot more simple to talk about.

Teodora Groza

Thank you very much for that. Now we have a question which goes into a slightly different direction from Caroline Thomas. She points out that if we define the objectives of the law too clearly, we take out the flexibility inherent in the common law system. Now the question is, to what extent do we want to do that, especially if legislative change would take decades and decades to implement?

Anthony J. Casey

Yeah, I think that's exactly right. Like if you say here's the data set, here's the algorithm, that's the law period, move on. You're going to lock-in things in a way that would be problematic that you've never done before. And so, you have to decide, as she suggests like, perhaps it's at the agency level, perhaps it's, and this is the paper that Dazza mentioned earlier, you can imagine courts are the ones... you have to think about the institution in an optimal kind of updating of old, you want to be getting the data about the world regularly, but also reviewing your policies regularly, reviewing the purposes and you have to decide what institution that is. And I do agree. If you've all of that at a legislative level, you're going to lock things in, in a problematic way.

Anthony Niblett

Yeah, I mean, this comes back to tradeoffs. There are costs from doing that. You're locked in, you lose the dynamism of the law, but you gain speed, and you gain consistency. And so how do you trade those two things off? And that is another human problem we have to deal with.

Anthony J. Casey

Just to give you one example, Anthony, you mentioned stoplights. If you said here's the data on traffic today, here's where we're putting all the stoplights, pass that in law and move on, five years from now, your stoplights would be a disaster, right? And so, you'd want to say here's the agency that collects data and regulates algorithmic stoplights. That's the way you want to think about all of that. That's the stupidest example, but I think it demonstrates the problem. We would have traffic jams endlessly if you just locked it in through legislation.

Teodora Groza

Dazza, do you have any thoughts on how too clear objectives would ruin the flexibility of the common law system?

Daniel J. Greenwood

Yeah, I do at a very high level. Two things. Number one, I believe we're very fortunate to have common law because it's so supple and such a great fit for innovative new technologies, especially of this type that can be tuned in ways that

are adaptive and can take lots of little data points that we could express as precedent, for example.

But the other thing, just in general, is I feel like...part of what is the context of this paper, that's the anchor for this conversation. It seems to me that the best thing to do is give it a try. And then to see. And the last, I mean, arguably the last session when we had this like stunningly big idea about like a new economy that almost feels like almost like science fiction, but it's important, you know, out of academia to set a vision over the horizon of how things could be when the economy and society is in a mature next state of a digital economy and information society. And now in this paper, we have something very practical like we assume no other changes in anything using stuff we already have laying around like data and like we get to build these algorithms out of the box. It wouldn't be that hard.

I feel like my overall take on this is I hope that we could end up somewhere in the middle. And we can look at applying capabilities like smart contracts and other classes of technology capabilities in ways that can solve multiple problems by applying them once and not necessarily in a very bespoke way. Because ultimately, we're going to need to find ways to transform systemically, not everything at once, but hopefully in clusters of types of problems. I know that wasn't exactly on point to the question, but that's at least what I was thinking when I was considering the topic.

Teodora Groza

Thank you very much for this. And now unfortunately, we have to move on to the last question of this panel, which comes from Brendan Maher. The question is, can you speak more deeply about the role of simulation, especially since simulation is different from prediction? I hand on the floor to two Anthony's for this one.

Anthony Niblett

That's right. When I talk about some of the bail papers, they are involving they utilize some simulation. When I use the word prediction, I'm using it in a very loose way. I'm referring to a few different things. I'm referring to forecasting and, I'm also referring to prediction in broader sense of, we know what we've done in the past, and if I give you a new fact-pattern, and making a prediction based on that, it's not necessarily forecasting what a court will do but I'm just trying to put a probability on this and that's essentially what this is doing. We're just trying to put a probability on this. I can't speak too deeply about some of the simulations. Maybe, Brenden, we can speak offline. I don't know Tony, if you wanted to add more to that. I know that there are differences between that simulation and the forecasting and the types of prediction that we are actually talking about here.

Anthony J. Casey

I think that your prediction is that this is going to happen with that kind of probabilities.

Anthony Niblett

Dazza, I don't know if you wanted to jump in there?

Daniel J. Greenwood

Yeah, just at a very high level. In terms of the question, you know, like simulation is not prediction, I guess. First of all, Brenden and I are friends and so I take this with tremendous respect, but I actually have a different point of view. I think that you can have descriptive models and simulations. And you can have predictive simulations, depending on what you're trying to do. And I know like you and I both go way back with agent based simulations that are exactly trying to be predictive.

But leaving that all aside, I think that the underlying insight here that I hope we all take away, especially those of you who are regulators that may be looking at your 2023 budgets and what you might want to bring online for technology is the capability of being able to simulate something that's pretty close to the cases and the reviews and the administrative actions and the end the economy and everything else here that you're trying to do is really historically important. And one of the ways that I hope that we'll have simulations and models that are incorporated as part of administrative law itself is that when we look at, for example, a proposed rule like okay, the merger review rules going to be we're going to tweak some of the criteria in the future.

Right now, in the US, we have NPRM (Notice of Proposed Rule Making), and it's all like PDF, medieval kind of stuff. Wouldn't it be nice if, instead, they actually have a rule in natural language and also showed us the model or at least a toy model and the simulation so that you could put in questions and get answers like okay, let's say I'm a company of this size with these datasets and these criteria? Would my merger be approved or not approved under the new proposed rule? You can actually get these answers when you express things as models and with simulations and it's a great way to not just measure the performance of existing law and its success against goals that we need to do a better job defining, but you can also be able to evaluate future law, which gets us into the idea of adapting well.

Teodora Groza

Thank you very much for that. Now, I know that I had advertised the last question as being the last but we have two minutes left. And I will use them to combine the last questions that we got in the Q&A section. The essence of the questions is whether we should prefer to use unsupervised machine learning tools because they allow us to avoid labeling the mistakes we did in the past of merger control. What do you think about that?

Anthony J. Casey

My understanding is that letting the past data speak and not labeling is useless unless it's not labeling the mistakes. I think the issue there is there...and I think this

goes earlier to the point of locking things in, you are locking in the way we've done it that's easier to do but you lose the value of potentially identifying them as mistakes, and I'll hit on a few...you hit like, but the important part is that humans have to decide what they want to do. We basically can't do that. All we'd be doing...

Thibault Schrepel

That's a mystery.

Anthony Niblett

Looks like we've lost Tony. I'll just add, unsupervised machine learning is helpful for us identifying some of the patterns, and I'm not sure if that's particularly helpful, it's helpful in the sense that you get to learn lots about what's going on. In terms of making predictions that you can turn into action I think that's where you probably need your supervised machine learning, but that comes with an awful lot of cost and that is we need to be able to tag what an error is.

Teodora Groza

Thank you very much for all your thoughts and for presenting your paper the two Anthony's, Dazza for very insightful comments and questions. Thank you as well to the attendees for the very thoughtful questions, and I hand it over to Thibault now for the next panel.

#3 – Remedies — Antitrust Enforcement and Big Tech

Thibault Schrepel

Thank you very much Teodora. Just a quick note, as you understood, we are very good friends with people at the MIT and Dazza's group. So, go visit the MIT Computational Law Report, you go to law.mit.edu, and maybe you will get a beautiful hat. And they just released a new issue of the competition law report a week ago. It's eyes opening. Go there and give it a try. You'll like it. For the next panel. I'm going to hand it to Maura who is our editor-in-chief, the maestro behind the computational antitrust project. Maura, the floor is yours.

Maura Margaret Carey

Hi, everyone. Thank you so much for attending this conference. Thank you to our panelists for being here today. Thank you to the computational antitrust team and, of course Thibault for bringing us together. I'm here today to introduce our panelists for this session. We have Ron Schnell, who is a computer science expert with more than 40 years of experience in the high-tech industry and in academia. He was the general manager and chief executive of the Technical Committee, the organization that monitored Microsoft under the final judgments in United States versus Microsoft. And he is currently a director at Berkeley Research Group and adjunct professor of Computer Science at Nova Southeastern University. We also

have Doug Melamed, who's Professor of The Practice of Law at Stanford Law School. Before entering academia, he served as Senior Vice President and General Counsel of the Intel Corporation, chair of the Antitrust and Competition Practice Group at WilmerHale and Acting Assistant Attorney General in charge of the Antitrust Division at the US Department of Justice. For the next 20 minutes, I'll turn over the floor to Ron who will present his paper: "Antitrust enforcement and Big Tech: After the Remedy is Ordered".

Ron Schnell

Thank you, it's great to be here. I'm going to be talking about how we can use computation after there have been remedies assigned in a case. In a lot of ways, it's a lot easier. In some ways. It's a lot harder. But what we're looking at in terms of data and such is a lot better defined, I would say. The paper I'm talking about, is called: "Antitrust Enforcement and Big Tech: After the Remedy is Ordered." I have two co-authors. One is Jay Hines, who, at the time of the Microsoft case, was the Deputy Attorney General of the State of York. There were actually two cases, there was the New York et al. versus Microsoft and US versus Microsoft. He was really one of the main people in New York et al. versus Microsoft matter. Jason Nieh, who had some involvement prior to the remedy, but in this paper, he talks about the Bazaarvoice case. And then there's me.

I'll give a little bit of background, additional background about myself. I started out as a hardcore programmer, working at Bell Labs, Sun Microsystems and IBM on the kernel of their version of the Unix operating system. All of those three things were in direct competition with Microsoft Windows, which some people would say is not a coincidence. I also lectured at NYU when I was 14. I was also a researcher at the MIT Artificial Intelligence Lab. After that, I had three startups, the last of which I sold to a public company in 2000. I retired after that for a while, ended up helping out a friend and becoming a vice president at Equifax Corporation, one of the three credit bureaus in the United States. And I was about to do another startup when I got this strange call, asking me if I'd be interested in working on this monitorship and I jumped at the idea it sounded like a really cool project.

It all has to do with the two final judgments in those two matters. Judgments changed or got modified a couple times but was originally entered in November of 2012. And one of the things those judgments said was there was to be an appointment of a technical committee. This was very unusual at the time. Even to I mean, typically, I think back then even it would be a special master or a retired judge who would do the monitorship. There wasn't as much of a concept of having a technical person involved in the monitorship of an enforcement action. But it said right in the final judgments that the TC technical committee members should be experts in software design and programming.

So that was somewhat unique at the time, and it was sort of later on, or at least seen as a great achievement and the judge was very pleased with it, as you'll see. So, what happened was the way that the technical committee was formed, was that the Department of Justice and New York and a group of states would choose one

person, Microsoft to choose a second person and the two of them jointly would choose a third person. The DOJ in New York et al., chose Dr. Harry Saal, who's a very well-known computer scientist and a very successful entrepreneur. He started Network General and was a really big thinker in high tech and Microsoft submitted Franklin Fite, who was an executive, a former executive at Microsoft, since retired and also a great technology thinker. And the two of them together submitted Dr. 'Skip' Stritter, who was an entrepreneur, founder of Nick's Corporation, one of the early employees of Apple, another very well-known technologist.

So then, they were called the members of the technical committee. I know that it was envisioned at the start that these three guys would, you know, look at what Microsoft's doing during the remedy phase and be able to monitor them on their own and sort of make recommendations and allow Microsoft the ability to cure any issues that came up. But, you know, they knew right away that that was not going to be enough that they were going to need more people, they were going to need more resources. I mean, Microsoft at the time, I think was an 80,000 employee, company. You know, at least half of it had to do with Microsoft Windows and things that were directly related to the final judgments in these cases.

They created a corporation called the TC that was actually going to carry out the monitorship. And so, the TC, in a sense is separate from what's defined in the final documents as a technical committee but they're really used interchangeably. And if you look at court transcripts and press articles about our organization, because eventually I was brought on to run the day-to-day operations of the TC. So, the final judgments actually consider the hiring of staff and consultants by the technical committee. I don't think the people who wrote the final judgments thought about forming a corporation or what eventually ended up being me hiring 93 people over six and a half years because the file judgments did get extended several times.

There are several sections of the final judgments that had to do with enforcement. I think the one that I started out with the one that's most well-known by the public and that was Section 3 (h), Section 3 (h) talks about a bunch of things, but the two most well-known things and probably the two most important things in 3 (h) were the ability of consumers to choose default programs. So at the time, the complainants in the case were people like Netscape for Netscape Navigator, and real media, you know, media players and things like that, and their complaints had to do with the fact that Windows really gave a lot extra ability for people to use their, you know, the Microsoft products as opposed to third party products and really made it difficult they claimed for them to be able to get customers.

That was the first thing, to allow users to have default programs. And the second thing was to allow users to disable Microsoft programs so that they wouldn't even show up as icons and things like that. So that was Section 3 (h), and it uses the term Microsoft middleware, which is, you know, in computer science, we would look at this definition, say, well, that's not middleware, but this is what the lawyers termed as middleware. So, for six and a half years, this is what I called middleware. It was things like programs that provided the functionality of Internet Explorer, Microsoft, Java, virtual machine, windows media player, Windows Messenger, Outlook, and their successors. Or just, generally speaking, in case there are other

things that don't have these names, Internet browsers, email client software. Network, audio, video, client software, instant messaging, software, etc. Those were the things that we got to focus on at the Technical Committee, and the TC to figure out what we were monitoring, exactly. So, how do you ensure compliance? Right, and eventually, how do you ensure it computationally?

Because we were all technologists at the TC that was the whole point. What we decided, what we thought we could do, was we could use test scripts to ensure compliance with these provisions, because it's so well defined in final judgments. You know, we can check to see for example, if a user were to try to set a default browser to this new browser, would it work? And this is very straightforward. You can do this without human intervention, right? We're simulating human intervention, but it's not really human intervention. And, we had to run these tests test scripts regularly. Because Microsoft Windows changed regularly and the potential middleware changed regularly.

So, we could run these test scripts automatically on every release of a new version of Microsoft Windows. Every release of the regular operating system updates that were part of Windows Update, or if there had been complaints from third-party middleware developers, we could run the test again, related to that. And also, these scripts you know, these programs could be updated based on all the feedback from their execution, from potential complaints or anything like that. So, so really 3 (h) would be quite simple to programmatically test in addition to human intervention, you know, we also do a lot of that as well. I have to be a little bit careful about what I say because the protective order is, you know, what amounts to a lifetime gag order for me that for anything that's not public information, you know, I limit what I say to what's in the public record. That was 3 (h) and 3 (h), like I said, is what most people knew about and what most people talked about.

But what not a lot of people knew about unless they enjoy reading court transcripts and stuff like that was Section 3 (e) and the supermajority of the work that we did was related to Section 3 (e). I had over 40 people at a time working on just the insurance of compliance with Section 3 (e). And what Section 3 (e) said in both final judgments is that Microsoft is required in great detail to document all of their proprietary communications protocols that would work from Windows client to Windows Server and vice versa. And I believe the reason for this was so that a third-party vendor could create these protocol elements on a server so that they were compatible with the monopoly of Windows clients out there. It's sort of you know, if you can't compete on the client-side, perhaps it's easier to compete on the server-side. And by documenting all the protocols, somebody could do that. This ends up being a really daunting task, right?

So, it ended up being an astounding number of protocols. Windows has been around for a long time. The final judgments was not specific about which protocols needed to be documented. So that meant that protocols needed to be documented that hadn't been used in decades. And, you know, it was difficult for Microsoft, especially, you know, they had a great history of creating a lot of wealth and a lot of retirees and for some protocols, there was no one left at Microsoft who even

remembered not what they did, but certainly how they were implemented down to the bit level. It ended up being tens of thousands of pages. Microsoft's first release of what's called the protocol documentation, or the MCPP; Microsoft's communications protocol program documents. So that's, that's a lot of documents. And what we needed to do as Microsoft called us regulators, actually, was we needed to figure out how to verify the accuracy and completeness of all of those tens of thousands of pages. So that's, that's obviously a really difficult job to do. And we again, tried to figure out how to do it computationally because that's what we weren't good at. We were all technical people but it was an enormous scale.

We also realized that reading technical documents is not something that lawyers or judges could do. So, we leveraged our ability as technical experts to just do desk reading. I mean, we could actually try to verify, as best as possible the accuracy of these documents by reading them and looking for obvious errors. You know, keep in mind that since we're talking about communications protocols here, one bit being off would, could be a heat thing, you know, it would make it unusable, right? You might not be able, you wouldn't know as someone who's developing these protocol elements, what is causing an error, you're going to get an error. It's not going to be compatible with the monopoly of Windows clients out there. And you're not going to realize that it's one bit wrong in these tens of thousands of pages of documentation, or at least a subset for this protocol, without a lot of work and a lot of trial and error. So that's something that we thought was really important that we catch as best as possible. So, sometimes desk reading could catch stuff like that, but at scale, obviously, although I would say you'd be surprised at the amount of stuff we were able to find from desk reading. What really, the most of our effort was, was something called the protocol project.

So basically, what we did, and this goes to question in the last talk about simulation, was we pretended to be third party protocol developers, and we, you know, on Linux boxes, we wrote code solely by reference to the MCPP documentation to be a server for these protocols. So, and we did it, you know, very modularly protocol by protocol. And we wrote, we wrote a ton of code that we later threw away that looked at this documentation and did things like SMB, did things like net log on and, and all of the windows protocols, and if something didn't work, or you know, our work product was a bug report against the documentation, which Microsoft would then fix. Those two things are really what we did to ensure accuracy as best we could.

We had another product because we scratched our heads for a while trying to figure out how are we going to verify completeness. Because if something's missing, if a protocol's missing, if a portion of protocol's missing at certain endpoint to an API, how would we know we couldn't do that using desk reading or the protocol project? What we did, we decided to do with Microsoft's help was actually instrument Microsoft's test labs around the world. We actually got some at the time, really state of the art hardware that we installed in Microsoft's test labs. We captured all of their network traffic at wire speed, which is very fast and a lot of traffic. We were using, you know, petabytes of disk space, you know, in the 2000s or the mid 2000 to late 2000 timeframe. And then we turned the protocol documentation from unstructured data to structured data using XML markup. And we

programmatically compared what we captured in the network with the documentation.

And that was our endeavor to figure out the completeness aspect of documentation. That was for me, a really, really cool project and very interesting. The output of all of these things, as I mentioned, there was bug reports on the documentation. We called those and the court called those TDIs (Technical Documentation Issues). So that's what we did for six and a half years is we created TDIs that Microsoft would receive and fix. We obviously we produced tens of thousands of TDIs. It was a crazy number. Because it's, like I said, it's a tough project to create this documentation. And we had to hire more and more TC staff, as I mentioned that, in total, we hired 93 people because we had to scale this thing because it was just so much work and eventually some of the documentation we would suggest, we started to suggest fixes to Microsoft, which they pretty much adopted.

Microsoft talks in their filings, court filings, about the fact that they hired more and more staff to handle our TDIs. At one point, they mentioned that they had brought 800 people onto the project. I can't talk about how they define, you know, a staff member in that context, but certainly they had a lot of people working on it, and they definitely took it seriously most of the time. And keep in mind that the European Commission was also running a similar type of action, and eventually at one point, the European Commission was fining Microsoft, I think, a million euro per day every month per month, I think, until certain documentation issues were fixed. And of course, those documentation issues were the same ones that we were filing. In terms of what the court had to say, I'm very proud of how the court reacted to the work we did.

You know, there were times when Judge Kollar-Kotelly was a little bit frustrated with the rate of TDI fixes. She says in this particular excerpt she talks about, at first, she thought that Microsoft is going to be doing all the work and we would just be monitoring it. And she says here, I will just throw this out. I don't have a full understanding. I understand. I do appreciate the fact these are not simple. These are complex, but I think it's interesting that it appears that TC is able to do what is necessary and bring Microsoft along, not the other way around.

So, this is you know, kind of an astounding thing for a judge to say in open court and it was you know, it was really, like I said, I'm very proud. I was very proud of it. I'm sure Microsoft wasn't happy with that comment. But, you know, I just, I was proud because we were doing great work. And you know, we were happy to be picking up some of the stuff that Microsoft didn't have the resources to do or whatever, but it was such a cool technical project. I mean, most of the people who worked for me loved doing this, these simulations, doing the structured data analysis, and stuff like that, I actually, because I still program I started out as a programmer, and I'm still a programmer at heart.

So, I took on some of these protocols myself. There are several of them and Microsoft didn't know which ones I was filing the TDIs on, but it was really fun, and we had a lot of pride in our work, and some people had an issue like they did

all this really good work and then we had to destroy it, because that was the way it worked. But at the end of the case, a bunch of the staff people and I went to a giant shredder. I believe it's in San Diego or Orange County, California. This is like one of these shredders you can throw a Volkswagen into and they'll shred it into little pieces, and we literally took all of our stuff and threw it into the shredder and watched it all get destroyed. In a way that was sad, but it still had a good sense of accomplishment.

In terms of computational antitrust in the future. I think it's important even though our paper is titled high tech it's not limited to high tech defendants, right? I mean, there might be a misconception about that but really every company has some high-tech element to them. I mean, they have an online presence. Remedies often include restrictions and requirements related to that online presence, like online advertising restrictions, network safety, and security restrictions, communications requirements, things like that website layout content requirements, doesn't have to be a high-tech company to have this in the remedy for a case, and all of these things are very ripe for computational monitoring. You know, it's very easy, you know, there are libraries all over the place that do things like screen scraping, collect data and analyze them and all of these things can be done that way. You know, I've seen a bunch of cases recently in the antitrust area that have to do with, you know, limitations on search rankings and things like that and all that stuff is easily checked computationally.

And, you know, obviously, we just heard a great talk about machine learning and artificial intelligence and some of the issues with that and some of the usefulness of that. I think for the remedy phase, it's much easier right? I mean, you know, we're lucky if we're looking at remedies because we're very limited. We're grabbing the data. It's very directed. We can grab that data we can sort of learn, we can use machine learning and come up with a computational algorithm to look for something that's a well-defined and a final judgment. I think it's an even more compelling place to use artificial intelligence and machine learning. And in some sense, we did that with the wire capture part of our project. I think that's all I've got. I will open it up to Maura.

Maura Margaret Carey

Thank you so much, Ron. That was a fascinating look at the remedy aspect of the antitrust enforcement process. I'll turn it over to Doug at this point for 10 minutes of comment.

A. Douglas Melamed

[Listen to the audio]

Maura Margaret Carey

Ron, would you like to take a few minutes to respond to Doug's comments?

Ron Schnell

No, I think he's spot on. I think that we can do exactly that. I think by looking at these remedies, I mean, looking at the outcome from what we did, I think I certainly learned a lot and I know that the people who were involved in writing, the final judgments and particularly the sections about the technical committee and restrictions and requirements, they learned a lot. And we talked, and Jay Hines I think would talk if he was here, about the fact that when he was involved in writing the final judgments, he couldn't predict what would happen six years down the line and it was only supposed to be a year and a half.

So, you know, I think it's important to be able to take what happened and think about that in the future. And maybe that's already being done, when these remedies are being written to try to future proof them as best we can. And although technology is really difficult, I didn't know if there's a specific question. There was a lot there. So, if you have a specific question, you know, maybe you want to repeat it and I can see if I can answer.

Maura Margaret Carey

I think we can turn to Q&A at this point. We have one question that came through in the Q&A for Ron. Based on your experience with the Windows operating system platform, do you think that there could be a standardized regulation for all operating systems, for example, Windows, Android OS, and Apple iOS? And I guess relatedly, who is better to do this regulation, the private or the public sector?

Ron Schnell

Right. To me, as a technologist—I'm not an antitrust lawyer – the most important thing is innovation. So, you know, technology companies need to be able to innovate. You know, I tend to believe that, you know, regulation isn't always the best thing for innovation. So, you know, unless you have a company that is stifling innovation and you need to act. I don't think that the operating system should be standardized. Because I think if that happened, for example, Android OS wouldn't be where it is today, as the worldwide leader in mobile operating systems. I think, MAC, macOS wouldn't have a lot of the innovation it has, Linux wouldn't have a lot of the innovation that it has. I'm not a big proponent of regulation unless there's really highly anti-competitive stuff going on that is actively stifling innovation. I think I also noticed that question in the chat.

Maura Margaret Carey

Yeah, we have a few additional questions in the chat that relate to generally the standards for Microsoft to be found compliant or non-compliant. Do you want to expand upon that?

Ron Schnell

You know, we're technologists. So, we used the final judgments as our Bibles, so to speak, and we were just making sure that they complied with everything in section three. I only talked about two subsections here today, but there were others. I personally did things like visit Microsoft partners and competitors and solicited complaints. That's one of the other sections that I had to deal with, but we were very specific in our task, and we reported back to the plaintiffs, which was the United States of America, and I think 19 states, and told them what we found.

And sometimes, you know, I would say if we had found anything that we didn't think was compliant or might not be compliant, we would talk to Microsoft, and they'd have the opportunity to fix it before we would talk to the plaintiff. So sometimes we didn't have to do that. So, we didn't we didn't verify “compliance”, quote, unquote. But our computational tools and our resources we verified whether or not they were complying with the elements of section three.

A. Douglas Melamed

[Listen to the audio]

Ron Schnell

Microsoft certainly did a lot of hard work. And, and they did say and I think we mentioned this in the paper that they didn't realize how useful this documentation from Section 3 (e) would be. It ended up helping them very, very much as a company because a lot of these communications protocols hadn't been documented well, in the past. My opinion is it reduced bugs in the operating system because they had the documentation for themselves to which they could refer in addition to competitors.

Maura Margaret Carey

We have another question that came in through Q&A. The anonymous attendee writes, writing the interfaces to connect to Microsoft platform was a full way to prove their documentation. Did Microsoft not provide test cases, test results and proof of results for proving the platform is open? And you feel that you actually did the test work that Microsoft should have been doing to prove that their products met the requirements of openness?

Ron Schnell

Requirements of openness is an interesting phrase to use here because obviously Windows is not an open system. Right. And these particular protocols were proprietary, they were not open source. Right. So, you know, in terms of the judgment, you know, we would not have looked at any of Microsoft's test cases, because by definition, we couldn't trust them, right, not because they were bad or anything, but our job was to do this independently, so we had to create our own test cases. My opinion as to whether they should have done it themselves. I don't think

they needed to because there was never, you know, prior to these judgments, there was no need for it to work with server elements to work with anything other than Windows clients for most of the protocols.

Maura Margaret Carey

So that's actually all the time we have for Q&A at this point, but I was hoping that both you Ron and Doug had maybe some closing remarks or thoughts that you wanted to share with the group before we wrap up.

Ron Schnell

I'll just say briefly, I think it's a really exciting time for computational antitrust. I mean, as I said in my final slide, we're in a time where every company has a highly technical component to them even if they don't consider it highly technical, everybody's online. All of these remedies necessarily will have something to do with their all...well, I shouldn't say all, but many of them will have to do with their online presence that can be deterministically measured using computational algorithms. I'm excited about the field and excited for what comes next.

A. Douglas Melamed

Computational antitrust is pretty fancy. To me, there is wisdom in realizing that we now have tools and things that are not just technology, these tools enable us to do what we should have been doing in the past that I found so difficult to do for decades, which is actually to evaluate how the law interacts with the economy in various stages.

Maura Margaret Carey

Thank you so much to you both, and I'll turn the floor back to Thibault.

#4 – Competition Agencies: A Space for Discussion

Thibault Schrepel

Yes, thank you very much, Rod and Doug. It has been a fascinating discussion, and I think what we've seen today is that there are potentially two objectives for computational antitrust. One might be to change it all, which is kind of what we heard with the first discussion using competition tools. The second one may be to take the rules such as they exist and to improve them and indeed, as you just mentioned, Doug, to do it better.

Now, I'm very excited for the very final panel of this conference. We're going to hear from top officials and enforcers. They're going to present what they do within their agencies, whether they indeed are going in direction to improve antitrust or if they see that there is a room to actually change the rules. And so, for that, I'm

pleased to introduce Felipe who is the general coordinator of the antitrust analysis at CADE, the Brazilian competition agency. After Felipe, we will hear from Stefan who is the Chief Data and Technology Insight Officer at the UK Competition Market Authority. And then we will hear from Yann who is the Head of the Digital Economy Unit at the French Competition Agency, the Autorité de la concurrence.

The idea for the discussion is the following. We've given them all 20 minutes to present. In the meantime, of course, please feel free to enter some more questions in the Q&A. And without further ado, I will give the floor to you Felipe.

Felipe Roquete

Thank you very much, Thibault. I want to thank the CodeX Center for the invitation and congratulate you and your team for organizing this amazing conference. It's important to emphasize here that the opinions expressed in my presentation do not represent the official position of the Brazilian antitrust authority. My aim here is to give you a brief overview about how we are dealing with the challenges related to the incorporation of new technologies in antitrust activities. I will not go through any technicalities mainly because I'm not an expert in the field but also because I think it will be more interesting to guarantee time to discussion.

The development of new technologies, such as artificial intelligence and blockchain, for instance, brings several opportunities for antitrust agencies, and the concept of computational antitrust is very powerful because it allows us to organize several subgroups into a program of action and research based on problem-solving vision. Computational antitrust, in my perception, should be understood as a toolkit in constant evolution, which offers antitrust agencies a range of alternatives based on innovations. This provides therefore the need to test those innovations, possibly make several mistakes, and consequently redesign our initiatives. That said, it's essential to emphasize that it's not innovation that should drive the agency's initiatives, but it's policy goals that should guide the new techniques and strategies based on those technologies.

In other words, once the agency goals have been well defined, we should look for techniques in that toolkit, that are methodologically aligned with the purpose of the antitrust policy. It implies therefore, that different agencies will be able to test different techniques with different results as well. There is no one size fits all approach as agencies have different institutional capacities are embedded in different legal frameworks and their policy instruments also a variety. I'm going to talk briefly about how the Brazilian authority is dealing with the big opportunities open by the computational antitrust. One of the main opportunities is how technological tools can facilitate the work of antitrust investigators and case handlers.

Automating procedures can exponentially increase agency's efficiency. The analysis of documents and evidence there's a summarization of answers and reports, among others, can result in transversal initiatives that is those capable of being reproduced after the patient by other agencies around the world. Here in

Brazil, our IP unit, in collaboration with case handlers and researchers from a local university are developing a tool to automate the detection of legal arguments presented by lawyers in their defenses in order to enable both the identification of legal reasoning and the standardized treatment of responses. We think it will guarantee not only the due process, but decrease legal certainty in our proceedings. Developing tools based on natural language processing is a challenge, but one that can significantly impact the work of antitrust agencies.

In addition to the issue of efficiency, computational antitrust may be the most suitable way to increase the potential for tax infringements and thus, it develops the power of antitrust authorities. Perhaps the proactive approach is the frontier towards which all agencies will dance walk, computational antitrust provides the tools that each agency will need to identify the best path, and this path will not necessarily be the shortest or the easiest one, but rather, the one that allows each country within its institutional framework institutional capacities to progress step by step. The difference in data availability, the diversity of market structures and bidding rules, for instance, makes cross cutting response very difficult.

For instance, machine learning algorithms for detecting cartels or markets more prone to abuse of market power will need to be internalized and adapt in order to guarantee the results accuracy. In Brazil, we advanced in the development of screening tools to detect cartels, but more complex techniques such as artificial intelligence techniques will require improvements in our team's technical maturity. For example, if the local judiciary is prepared to accept its use, if its use is in line with revelations, so on and so forth. This it's possible that initially, the implementation of artificial intelligence techniques will be more viable for internal use by the authorities. In other words, for intelligence use.

Furthermore, it's essential to emphasize that antitrust agencies will also be affected by artificial intelligence regulations, and its requirements regarding the guarantee of fairness and transparency. I believe that the debate around new technologies for solving antitrust problems should guarantee space not only for the evangelists, those who seek to highlight the opportunities the potential brought by its implementation but also by the apocalyptic ones. Those who emphasize the risks and uncertainties regarding its results. This will provide the decision-makers with a broad comprehension of all variables involved. This relevant to antitrust policy must be embedded in the design of the new computational antitrust tools. Otherwise, we will create an environment of institutional schizophrenia in which we will class these initiatives fully adhering to the antitrust policy, and others that are unable to dialogue with policy proposals. In conclusion, I believe that computational antitrust should be implemented, not from an optimistic or pessimistic perspective, but from a realistic one. Thank you very much for your patience.

Thibault Schrepel

Thank you very much. I think I'm going to give the floor immediately to you Stefan, so we can then have time for more discussions. You should be able to share your screen, if necessary. Perfect. And the answer is, yes. It's working.

Stefan Hunt

Thank you very much for the invite to be here Thibault. It's really great to be speaking with a bunch of people with similar experience with Felipe and Yann, and many similar interests such as yourself and the attendees.

So, my name is Stefan Hunt, as you said, and I run a unit called State of Technology Analytics at the Competition and Markets Authority. I'm going to be really drawing primarily on that body of experience in the next 20 minutes. What I'm going to talk about today, is how we've kind of formed our team, to give you a bit of background about what team's like. I'm going to focus mostly on case number of cases where we've supported some cases that are major well focus a bit on building technology, developing techniques, and then...

The unit began starting in October 2017, when we got permission from the board, to set up a new unit. Our CEO, Andrea Coscelli championed setting up of the unit, and he actually effectively advertised for my role to an interview with the Financial Times in November that year. I personally joined a bit over three years ago in August 2018, and we started hiring for people very soon afterwards. And we really continued hiring the next year, and we've got a great response. And as we started grappling with the big tech firms and other big issues, we started recruiting more. In fact, we started building in a high-level unit actually coming up two years ago now, as well as the best. We've had mostly organic growth, and we're nearly 50 people, which means we're actually really quite sizable when it comes to computational analysis.

In terms of what skills we have, we started off with data scientists and a little bit of engineers. So really focusing on people who can use machine learning and code. And the engineers are particularly hardcore techies who are building up platforms, software engineering, and this kind of thing. However, very, very quickly, literally, just a couple of months in I realized that this wasn't going to be enough. The data scientists and the engineers, they really want to work with data or code, unsurprisingly, and they see the professional domain as being based on this kind of thing, and not competition professionals' par se unlike the economists, who some can be very quantitative, but even then they're happy to be embedded in doing much work that isn't primarily quantitative.

This isn't true for at least quite a number of our data scientists and engineers. This is a bit of an issue; some are more interested. So, we very quickly realized that we needed this additional capability which is data and technology insight. And these are people with hard science backgrounds, but they recognize and they attribute, and they want their impact. They want to be working primarily in the qualitative domain. In fact, one of my engineers she's just moved over actually from head of engineering into this team. But she's got this brilliant technical background, that she realized that she can have the most impact through being in data technology

and insights. I've also already mentioned behavioral hub. We also have digital forensics and e-discovery capability.

Now we carefully prioritize our work from the beginning. We focused initially on consumer and merger cases. I'm actually going to start off with the consumer case. I think you'll find it quite indicative of what's possible in terms of competition. A project we did from the very beginning was trying to think about how do we deal with fake and misleading online reviews. And this project illustrates three main capabilities that we bring to the table when it comes to computational analysis. We, first of all, started focusing on the platforms where reviews were posted and sold, and this was in particular Facebook and eBay, the firm for making technical commitments on what they were going to be able to do take down buying and selling fake and misleading online reviews, and my team's role really was to sit in the meetings with parties. And effectively you can think of us as a pardon my language, a bull crap detector, and it really made a difference, like our ability to understand what they were saying, but also, to be able to challenge was very important.

At the same time, right from the beginning, we started working on identifying problems from data sites that hosted online reviews. And here's an example of some early work. So, effectively, we've been tipped off for a company. This is the red dot right in the center. We then scraped the reviewers of all the seed company then we looked at all the companies that these reviewers had then led, had then reviewed themselves. We found this very unusual networks. For example, in this, you can see these green dots all circled in the center, in this case individuals, all of them reviewing an American asphalt plant, all of them coming from a Korean's firm, which is obviously highly suspicious. Really we had to find that we can identify issues with best network analytics.

But it turns out to be sort of even better than identifying because this original way of identifying actually had already been started by our consumer team, they'd literally been manually going in Excel, looking at individual review, individual review, and then recording in Excel. One of our engineers, his first day on the job on, you know, on his own initiative, he did the scraping. He created this first version of this chart by himself in his first afternoon on the job, obviously, where it's legal processes, where, you know, if you can write stuff in code, you can literally do 1,000 times, you know, 10,000 times, 100,000 times more than might otherwise be possible. It's really a game-changer in us being able to detect problems and identify problems.

Ultimately, the work that we do on scraping and identifying fake reviews was a major part of deciding where to focus. And the third major thing that now we've had on the second stage of the investigation is working with Google and with Amazon, right now, sort of all of the team, this is the third to integrate is to actually really trying to understand the algorithms that Google and Amazon have and obviously, our job is to work with our consumer colleagues to figure out from a legal perspective, are the firms doing enough to understand in order to take down fake and misleading online reviews?

I'll focus on another early case as well, which was our advertising market study, where we were focused in particular on Google and Facebook, but obviously, a significant focus of that is going to be on Google search. Now, we really wanted to understand, in a quantitative and rigorous way the magnitude of Google scale advantage search, right? We really, well, yes. Well, we know there are some, how big is it? Does it really matter? What we did was we went to Google and Bing we got one week of all Google and Bing searches, all searches in the UK. This is half a terabyte of data. And moreover, we used an encryption measure we hacked because it's actually lots of personal data, names, addresses and these kinds of things, such that we actually used the same hash across the two so we can then match with it the same, exactly the same searches that occur on both Google and Bing within a similar week. And this is what we found.

Let me take you through this chart. To be clear this chart would not have been possible without a mix of my engineering team to actually get the data we wouldn't be able to deal with data of this volume otherwise, but then my data science team to do the analytics. Let's take the red line first of all. This is looking up all of Google Search searches and then ranked from the most frequent to the least frequent; the most frequent is on the left. And then we say, those individual searches that most frequent, what percentage of those searches can Bing see? And of course, the most frequent searches that Google sees, Bing can see all of them. And this will get into, say, the median search of Google we can see that actually it's around 25-30% of those searches that Bing is actually able to see. And as we get lower as we move into the bottom 30%, the tail queries, this is still three in 10 queries, right? It's very important for search, right? Yeah, we can see that it's actually 0.95% of those exact queries is what Bing is able to see.

Let's flip that around and look at Google. And we can say, with Google, you can see that the line is much, much higher. In particular, if we focus on the tail queries, we can show that on average in the bottom 30% of queries, actually the percentage that Google can see exactly the time of Bing searches is over 30%. We've got over 30 times the amount of data that Google is able to access, using base 10 that's one and a half orders of magnitude in terms of better data that they have and really helps us understand the nature of the scale advantage Google has, not possible without a team like mine. But it wasn't just that there's a variety of other things. So, one was really getting the nature of technologies. There's some really complicated technology in terms of tracking that we have to be aware of when it comes to advertising: understanding the role of privacy-enhancing technologies and understanding Personal Information Management Systems. That engineer I referred to earlier, who's moved from the engineering team to data and technology insight. She was the one leading all of these things, really making sure we had the detailed and she got a background in software engineering. She's really a technical she was able to understand these things very well.

But it wasn't only that; we also did a whole bunch of stuff on data architecture too, architecture for privacy. I'm probably not going to focus on the behavioral aspects in this population. Let's move that forward on that particular study. We did this work on digital advertising in January of this year, we launched a case, very unusual

antitrust case. The Google privacy sandbox case, this is where we're prospectively concerned about the changes that Google is making, central again in making and will be making to how people are tracked over time. And the concern is actually the amount of information that Google would have that each individual will be much higher than other parties and that would give them a whole bunch of advantage compared to other party in the market.

It turns out there are lots and lots of different, really quite technical aspects to this. For example, the main way that Google is going to put this into buckets is usually federated learning with cohort. This is an unsupervised machine learning technique that has quite a lot of detail. That's really technical, really important to understand, especially federated learning, you'll learn that it's on our devices, the information that you can send. But it's not just that; there's this technology called TURTLEDOVE; there's lots and lots to do with technology changes to do with conversion tracking. There are lots and of lots of stuff they're doing to make sure that companies can't hack the way that they can track people over time. It's really important in all of these things to make sure we really understand the technical details so that we can do our job. We've only had 1.2, full-time equivalent of the team based on this, this team about seven people, but what we're able to bring to the table is so different in terms of capability that we're really able to catalyze the rest of the resource previously, frankly, we'd probably be relying on economists to do this technical work, but these individuals can do it, do much more of it with much less of their time and speed up the whole enterprise of doing this particular type of study.

So, in the interest of time, I'll skip through the mobile ecosystems market study. We actually came out with a very big interim report yesterday on mobile ecosystems. I'll skip that but the point here was going to be this, there's a lot of very important behavioral analytics you can do on this subject yourself. But I think it's important when it comes to casework to highlight the role of merges. So actually, there's a lot of opportunity for pulling data and technology capabilities into merge teams to make sure that we can understand issues in the digital markets properly. Best example of this is Facebook's acquisition of Giphy.

So, one of my data scientists joined that team and was crucial for understanding the main data and foreclosure barrier haul, one of the three main data hauls, yet again, very small amount of his item was able to very significant catalyze the other colleagues who'll be handling this particular merger in place. And it wasn't only Giphy, so we've had people really involved with a whole bunch of stuff to do with Cloud Essentials, Tableau, algorithms in the case of Amazon delivery, or COVID SmartBox, which we are trying to understand the technicalities of effectively an ID remedy for private parties where they would open up a bunch of code to others. And it was really important with technical people who could assess whether those commitments were really deliverable or not.

But it's not just cases. Filipe referred to the work on document review they've been doing in Brazil, and the one of the things that we were doing, we've got a whole new pipeline for how to deal with submissions of documents into our new evidence

Submission Portal. And were able to take in millions of documents and automatically and immediately verify whether the document's in the right format or not, and actually reject submissions that aren't like that. And so that was our engineering team leading on this. But more than that, now we've got our own web thing that we built. We actually add in a lot of our own peptide-based enrichments and the metadata, actually each of the files, so on the left is an example of a text length, which is a similar algorithm to PageRank, which was the original algorithm that Google used when it's trying to rank different websites, quality of search, which can be used actually summarize documents. It identifies the key phrases within documents. We're actually able now to put that within our submission portal to make sure that every document we have now has capability to actually take out the key phases. We've got a number of other data enrichments where we're trying to improve what was in the document. We've also built a Resale Price maintenance detector, which was launched at the end of last year. So that allows us to get data and analyze when we think it might be...

In terms of technologies, it's also been really important to develop links. But we published a paper in January this year on algorithms - how they can reduce competition that harms consumers. It tries to lay out different things at home, where we think algorithms could be relevant. And then right now, what we're trying to do is make sure we're building out the capability to deal with the various different types of harm. For example, I mentioned on online reviews, that's an example of ineffective platform oversight. There's more general category of harm. We're trying to make sure that we understand now, what are the kinds of techniques that we need in order to be able to understand and improve where we find the algorithms aren't sufficient enough, and also thinking about a variety of other ways of harm and how we can build up relevant techniques. Another thing that's important is horizon scanning as well, think about and understand different issues that might be coming along the pipeline, mostly we're thinking about this as a qualitative piece of work, and we're thinking about starting to work on technology requirements, but where I think there's potentially something that we could do with some forms of automation that's what I'm thinking.

Some reflections, we're very much conscious about our new technologies' perfection, professionalism, and it's not just us, it's great to see Yann on the call, as well, we know that the Canadians, the Australians have in the last few months made some significant rejigging and increase expansion in what they're doing. The commission has just had a significant upgrade of its technology team, and a whole variety of other agencies as well. We've got these new technologies profession, we're working with lawyers, economists and others, but it's important to understand what we're really talking about here is quite a spectrum of different skill sets. It's important to understand and know what they are. We technically work across all competition cases, but the fact is, the vast majority of cases we're working on are the digital ones. I think there's a chance to radically change how we approach cases. So, for example, I gave two examples where there was a really small amount of resources that we were able to put in, but really catalyzed very significantly and changed how cases work and are being done.

But it's not only casework; we've been much developing technology, our own technology, as well, and techniques. We're using a wide variety of tools and from engineering, lots of different kind of pipeline technologies, especially in AWS, lots of different machine learning techniques as well. We think we're really using the right tools because we've got a wide variety of people with a wide array of expertise. I guess one thing I would caution from some things I've seen on some academic papers is some focus on techniques that I think sound quite interesting and certainly we're well aware of, but I think we really can't see ourselves using those techniques at all in the next few years. Because we don't think that we really don't think they're applicable to the use cases in practical competition law and economics. We'd be very happy to discuss some more of that if people wanted to.

And lastly, the last one I want to focus on so what we're talking about here is if I was going to be a management consultant, they call this normally the digital transformation because the language of the organization's if we talk about the digital transformation of competition agencies, there's actually the opportunity for us to benefit from the same factors that firms have been benefiting from themselves as well. So, to the extent that we can build these capabilities in different agencies, you know, we can really get network effects as we build our own capabilities and we're able to share them, obviously going to be most important when it comes to digital products that we've developed, where they can be potentially very low marginal cost in sharing them together. You know, we have in some ways already started sharing code with a few other agencies. It could include the platform, we've got our own bespoke platform, could include scraping libraries that we have, not going to be protesting. We're very excited about the opportunity to work together with other agencies and co-build technology. We will have shared GitHub libraries that we'll be working on. We're very...sorry. We'll get heavily by talking about some of the problems with intelligence, which are really hard. We think that can be you have to work with other agencies that allow us to make more progress on some of these problems.

Thibault Schrepel

Beautiful, I was typing questions. You are getting me very excited, right now. But I'm not asking questions. Right now. I'm first going to give the floor to you Yann. And I think you may have some new projects you want to discuss. So, I'm very excited about this. And then I will be asking my very complex question to the three of you.

Yann Guthmann

Okay, let me share my screen. Hope it works.

Thibault Schrepel

It works.

Yann Guthmann

Thank you for the invitation, Thibault and CodeX. I would talk about the French variant. I'm Yann Guthmann, the head of the digital economy unit. This unit was created in September 2020. So, we're very new; we're kind of a toddler unit, you can say. In the future, maybe we could have our own cases, but currently we're kind of at talks stage, we're only four in my unit. So, we are very much smaller than Stefan's unit where they're 50; I'm very jealous. And so, our resources are better used assisting case team than being on cases and so, the use of data and scientists, assisting current case team.

Regarding my team, we're four, myself, two data scientists and one lawyer. And we have four objectives. First one is to develop new tools investigating methods, technical tools allowing the authority to deepen its understanding of a digital economy. The second objective is to work in cooperation with the authorities involved in digital matters. And so, we worked on several cases like Google never even heard part of Google cases. Also, Google News Corp, for example. Our third objective is to act in close cooperation with other competition or regulatory authorities in Europe and beyond, and also with the relevant service within the government in France. Our fourth objective is to develop contact with the academic world and institutions. And for the rest of our presentation, I will focus on the first objective that is to develop new tools.

To develop new tools, I think there's two ways to do that data unit. First one is case handler centric. And in this case, usually you answer to a specific need, the case handler has a problem, and you want to provide a solution, and there's immediate application. But what's really important is that you can have very good feedback from the case handler to improve your tool. Usually, you're working under a time constraint. And one very important thing is to focus on the user interface. One point of attention I'd like to mention is that we conduct an in-house survey inside the authority and we found the case handlers can be afraid of digital tools. They fear that they will lose more time than actually gain it with using this kind of tools. And so, we need to appear. And so that's why it's very important to focus on the user interface. You need the case handler wants to use these tools. So that's one way to do it.

And there's also another way that is technology-centric. In this way, you have no case handler, and usually it's more about experimentation or exploration. Sometimes projects might be useful directly or indirectly but you don't know when you start this project.

One thing I want to mention that there's a lot of skill over effects as a new surface kind of project as define a working relationship, for example with the IT team on loss Tech's non-creative projects, this way we were able to work hard to build virtual machines, for example. And so, when a very urgent case arrive you already know how to build virtual machine with the IT team. One thing I'll mention is that sometime the IT team is kind of afraid of us because we want to use a lot of very complex things. We need a lot of authorization so they can be afraid and so it's very important to have a nice working relationship with this unit.

You have freedom of cooperation, it's usually a long-term project that can also involve any shift of activity. That means that when you have a lot of cases where you need to support the antitrust unit, you can put this project on standby and they can be inter-project and when there's low activity, you can put more resources on this. And in the technology-centric project you don't need to focus on the user interface; you need to focus on proof of concept. You need at the end of a project to have something to show to some people usually your chief. And so, I think that these both types of projects are needed for any data unit. And you can see that one is the upper way of doing things and the other one is the Google way. And so now, for the presentation I will present two projects, one case-centered centric and one technology centric.

So first one, screening tool, it's kind of essential for data scientists and you answered with a very specific need a case handler wants to monitor one of several websites. We can all agree that it is very time-consuming for a human and it also can be very boring. But it's very easy to do it with digital tools. And so, at the authority, we use a Python library that's called scrapy for the back end and HTML for the pitch. What's also important with the screening tools is that you need to avoid a black box feeling because you can't set it to be just under where the screening tool is working. And in two months, you will have the results. It wouldn't work. And so, we created a dashboard using the dash technology. That's the same one that I use on the website, for example for the dashboard of the COVID pandemic, for example, with the same technology.

We also created a demo project that is available to all the case handlers at the authority so like they can see an example of a screening tool. One more thing is also very important that we add feedback from one case under the screening tool. And you need to focus on simplicity. Because my data scientist that worked on this tool at the beginning, he was full of ideas, fancy indicators that you can forward series and so on and so on. And it was a disaster, a total failure because the case handler didn't understand what was presented to them. And so, on the screening tool, you need to focus on the simplicity easy indicators to understand, and you can save the case analysis of the end report at the end of the screening period. So that's very important to understand. That's one quick application of the case-centered project that we did at the digital economic unit.

And now for the technology-centric approach. One new project. It references between decisions so it's an ongoing project that started with one simple idea, opinion and this is in the computational field reference of previous cases that's the principle of jurisprudence, that's how it works. And so maybe by detecting and combining all the quotes from all the authorities to publications, could we identify the important ones. So that's the basic idea, very simple. And so how did we implement it? First, we gather all the opinions and decision of the authority from 2009 to 2020. It's a little bit less than 600 opinion and decisions. We chose 2009 because it was creation at the authority at this current form. For that we were concerned of competition. So that's why it's arranged to 11 years; we think it was sufficient data.

After that, we created a tool that can identify the references to publications of the authority inside these documents. We focus only on the authority publications. Why? Because it's easier that's the answer. We know the matter the format of the authority decision, and so it was easier to create a tool to notify them one way for the future maybe we can design a tool to also recognize for example, references to a commission with some French jurisdictions. And so, we gather more than 10,000 quotes identified representing less than 5000 publications so that's the end results and all these opinions and decisions can be found on the website for the authority so we use only publicly available data.

What are the early results, the most quoted decision is a decision from 2007. That's concerning French Telecom in the sector of internet high speed. It was quoted 85 times by 41 publications. And there's two main reasons why this decision is quoted. First one is the Telecom relevant market but only for about 1/5 of the application. And the other reason is because it's a standout for the integration practice. So, this first result is very good because we identified very important decision that set the standard out for the authority for the integration practice. We were kind of happy. One other result, for example, most references by a decision. It's a very recent one from 2020 December regarding practices implemented in the mass retail food distribution sector by Carrefour and Tesco but in these cases more than 90% of our references are related to business secrecy. It is confusing, we don't understand much. You need to regard all the results are not very interesting.

One more thing we can do is that we can do some math because you can create a graph of the authority's publication where the publications are the nodes of a graph and the references are the edges. And so, you should be familiar with graph because this presentation had a lot of them and so what are the question that can be posed is the final graph connected? Is there a path between any two decision of the authority or what is the partition of the underlying connected subgraph? Because if your graph is not connected, you can't divide it between connected sub graphs and so what is the main distance between two publications? Those are some question that can be asked.

I wanted to check because we have a picture of the final graph. So, beware. You can see that we have one big, connected subgraph at the center, and a lot of tiny subgraphs and so we were kind of disappointed because we were hoping for some separated subgraph that maybe can be mapped to some specific sectors or maybe specific practices but now everything is connected. We need more work to identify what is interesting in this work, and so that's why it's an ongoing project. There're a lot of clusters, and you can see all the references plotting on some points, and for the size of the image the final one it's more than 13 megabytes, which kind of huge for an image and if you want to zoom on one part here, you can see here, some publication here. And of course, from this publication. You can see the relationship between some decisions of opinions. What we hope to do in the future, it may be an internet graph - you can browse on your browser on the website. And when you click on one publication, you can have all the data related to that. So that's what I'm hoping to do. And we hope that by doing more work, we can identify all the important decisions and maybe can help new case handlers for example, identify

what decision they need to read to understand specific practices or specific markets. So that's all for my presentation.

Thibault Schrepel

Thank you very much. We have quite a few questions in the chat. I also have quite a few questions myself. I've tried my best to cluster those questions into three different segments. Some of those questions may be directed to just one of you, but of course feel free to jump in if you have something to say. Some other questions are directed to the three of you.

The first is for you, Yann. You mentioned the fact that the case handlers may be afraid of using computational tools because they have the feeling it's too complex. I'm wondering how did you deal with that? Was it just a matter of improving the interface as you mentioned, or did you also have to organize training sessions? And I suspect the question may be also relevant for Stefan and Filipe.

Yann Guthmann

You need to optimize the interface, but we didn't do any training sessions, but on the opposite we do a lot of communication internally at the authority. We communicate a lot about what works, how they're working, and we're doing a lot of demos in internal webinars to show how it's working and what can be done with this tool. Because one other thing we found that is the case handler has no idea what these tools can do and so they usually don't have needs and don't have ideas about what can these tools do. You need to go to them and to present a lot of the work you're doing, you cannot work on your own like a black box. We need to show internally what you're doing. And at the authority, we have implemented internal webinars to show our work, for example,

Thibault Schrepel

Have you seen that the attendance is growing? Is it easier for you to get people to attend the webinars?

Yann Guthmann

No, we have a kind of core audience of around 40 people, but it's very good.

Thibault Schrepel

It's good. Yeah, it's very good. Stefan, I know we discussed that in private so maybe there are some things you may want to share there?

Stefan Hunt

I had a thought which is building on what Yann just said, which is I mean, much of building up these teams. I mean, there's just a lot of very practical skills in terms of

how to engage with colleagues. I guess one of the other questions was about kind of getting buy in, how to make sure you do that. So, for example, in my first few months in the job, I was lucky to have a monthly meeting with our CEO, so he said some of the slides were very visual that I presented. So, getting really visual slides in front of the CEO really early to demonstrate what was going on. Lots of numbers. Try to show the numbers in general show the impact that these tools can have and obviously, presentations, when we get the clients to actually give a presentation. Basically, they're going to be much more persuasive about the value of the impact on their work than we can possibly explain.

Thibault Schrepel

Great. Beautiful pictures, big numbers is the key, right?

Stefan Hunt

Yeah. Also, anecdotes, anecdotes are really important. Like I gave the anecdote of my engineer Tom, who built that first version of the chart, on his, literally his first afternoon on the job. You know, things like that. And then there's lots of other physical anecdotes we had of little things. For example, one of my data scientists, one of the teams was really interested in is pricing data for a particular product. That was going to take them around two weeks or so to literally, manually download all these different things and put them together. As I came to ask the data scientists whether they could help, it turns out I audited HTML code, HTML code for the site. It turned out there was a download button that had been hidden. So, you just unhide the download button, and literally say two weeks of work in about you know, 10 to 15 minutes. So little things like that can really make a difference, people remember.

Thibault Schrepel

Great. Filipe, I'm curious about your experience. I mean, the expertise is one thing, but the adoption is another, isn't it?

Felipe Roquete

Yes, one of the more successful strategies we used here in Brazil, we tried to consult case handlers and investigators to consider about actual problems that we could help them just solve with new tools. Then they gave us insights and inputs to solve actual problems they have and that was a very successful approach at the beginning of the project.

Thibault Schrepel

Alright, so showing how to help with what they do already, instead of saying you will be able to do even more. My second question is as follows: how do you make sure that the top management will actually give you the budget to hire the people you need? Stefan already mentioned that you need beautiful graphics, big numbers and anecdotes. But yeah, how do you approach the Heads of agencies?

Yann Guthmann

Well, since we're a newly created unit, we kind of get a token of goodwill. That means there was a need for the creation of my unit. And so, at the beginning, people were eager to have the help of a unit. And after that, you need to deliver, you need to help them you need to get time for just answer with tools or to help them understand a very complex digital market. And we kind of delivered internally, the case handler we're very happy with what we provided and so we are building on that. And now, the case handlers come directly to us to ask for help. And so, we're kind of goodly established inside the authority.

Thibault Schrepel

Stefan, I think you kind of had the same experience, right, trying to convince the CMA management...

Stefan Hunt

Yes, that's certainly a little bit more of what Filipe was saying right now, find out other people's problems. Yeah, just before my data scientists arrived the initial set, we did a whole bunch of kind of vague collaborative exercises. We like the whole of the variety of the rescue organization, lots of kind of posted notes on walls, trying to come up with all the different problems we could potentially tackle. So we actually had 35 different possible problems around the different agency again before we really had had a team and that was really helpful because I didn't take those to our CEO who is very decisive, and he's like, I want you to do 3737 and 17.

Obviously, the team did buy in because we developed with them, then he's got buy in, and that's when I started doing this. Let's go back with all the visual, the visual things. And then so he can see straight away what we're doing. And then just we'd keep feeding back to him. Very well. He actually stopped doing all the visual stuff after a few months. He didn't need that immediate feedback. But then it was much more about bringing based on what Yann said, I mean, it's all going to be based on what you deliver.

So, it was actually quite a while of writing. It was just a couple of years before we really did the first...I think you've got Bill Kovacic on here. So, Bill, it was actually couple of years before we first went to the board with all this but we really thought incredibly carefully about that paper and redesigned it very well, so it was quick for them to read but conveyed a lot of information really well. And it's pretty interesting as you create these different pieces of communication, how useful it is to reuse. So, for example, actually, we've recently been sharing the original strategy paper that we did a bit over a year ago and then we did a refresh a couple of months ago, and sharing shareable virtues from those, with really all of the other agencies that would like to see them. And so again, this talk about what we can do in terms of helping catalyze other agencies and of course, then helping them catalyze us. So, we were thinking very hard about influencing the CEO throughout the lifetime of the unit.

Thibault Schrepel

Yeah, great. And so, regarding the adoption, something that you kind of mentioned without being too specific on purpose, I think, is the type of tech that you wouldn't use because you think it's not ready or will never be ready. And indeed, it might be very hard for an agency to explain how the machine came to deliver a specific result. Is it the type of tech that you had in mind? And I guess the question is goes to both of you, Felipe and Yann, is there a space where you think “well, I will not go there, because I'm not sure if I can deliver, and even though I may be able to deliver results, I need to be able to explain how I came to the results”.

Felipe Roquete

Well, we are still evaluating the challenge of implementing artificial intelligence. We're still waiting for a Brazilian regulation currently analyzed before Brazilian Congress.

Thibault Schrepel

Great, thank you. Yann, anything you want to share?

Yann Guthmann

For the moment, we do not have the opportunity to use AI in actual cases, but we are thinking of using some techniques like NLP but for the technological-centric project. First, we want to experiment with it. We want to understand it better.

Thibault Schrepel

Yeah, makes sense. Stefan, can you disclose the tech that you had in mind?

Stefan Hunt

Sure. Okay. I will get there, I'll get there in a second. In terms of techniques, I think we'd be happy to work with unsupervised techniques. So, for example, the PageRank algorithm, I was discussing summarizing documents that's another unsupervised technique. The reason we feel it's okay to use is because nothing's really resting on it. It's just helping the case teams, documents, and have a quick look wherever they think it may be important or not, but it doesn't affect their judgment of importance or not at all. The key thing is, is the document important or not, and how would it be used? And we just thought it doesn't really pertain to that. We've used a bunch of different supervised machine learning techniques. We've got some named entity recognition techniques that we're using. Interestingly, we just started building some of these we originally did it, we had a COVID taskforce where we're doing lots of complaints.

That was when we built some of this tech. And then turns out, you can actually deploy it for document review some of the code so that's a supervised technique. So, there's a whole variety of techniques you can use it really just depends upon the

task at hand. In general, actually, quite a lot of the stuff we've done is deterministic. There's not necessarily using machine learning. It's really making sure the code's worked making sure you've got sufficient quality controls for that. The stuff I was referring to was a gentleman started talking about agent-based modeling, for example. And there was some stuff I saw on the use of machine learning in another paper.

Agent-based modeling is actually just technique is of course, fine. I mean gets used in evolutionary biology all the time, and it's very effective. In principle, it can be used actually I think within markets. I've seen a couple of good attempts to do it. There's a really good mortgage paper about six years old, another one that's about four or five years old on the bond market, and they use agent-based modeling but it's really hard to use agent-based modeling within some market systems because assumptions that you need are normally just too strong and you just don't have enough data. And all these systems are too complicated. I can't see us using any agent-based modeling not in the next five years.

Thibault Schrepel

You're getting me worried because you mentioned COVID as a way to foster tech adoption [laugh]. I know that Brexit also played a role because you suddenly had more cases. So, if all we need is Brexit and COVID for computational antitrust to be adopted, I'm not sure we really want that. As to agent-based modeling, I did a brief recording on the subject and I'm also working on it. And I think it may depend on whether you want to use agent-based modeling to predict for sure what's going to happen on the market or if you may want to use it in a way to document uncertainty.

There're lots of study explaining how agent-based modeling can be used to predict the future such as the effect of a merger. But it relies on many assumptions, so, maybe it depends on what you want to do with it.

Yann, you've showed us a fascinating project using network analysis. I know it's hard. Let me share some thoughts. First, people will tell you that it's not because the case has been quoted several times that it is important. Second, how have you done the labeling? Who did that? Was it you, or was it the lawyer in the team? And which software did you use, was it Gephi, or was it another software?

Yann Guthmann

We didn't choose any labeling because we just take the decision and just identify automatically the references. At the core of the project, there's no labeling. After that, it's very useful to add data and metadata. So, we use the one from the website of the authority where our communication team usually input some data on the website, what is the main practices and what are some of the juristic aspects and so we're using this data. So, it has already been labeled and after that, we need to crunch the results, for example, to understand why this decision is for most quoted.

And so, the example presented, I was the one to do it. We need a way to let the machines do our work but we haven't found it yet.

One other aspect that is annoying us is the time aspect. Because usually the more ancient decisions are more quoted, all the recent decisions can fly under the radar that way. And so, we need to think of it and define maybe another matrix, maybe the number of quotations by year. There's a lot of things we need to work on, that's why it's early results.

Thibault Schrepel

Of course. And it might also be that you have the opposite problem in which a case is being quoted by another case which is then cited by subsequent cases. This means the first one kind of disappears, although it's actually the basis for the case law.

We do somehow have only 10 minutes left, and I have five questions. So, moving on to the procedures, I have a question for you Stefan. You obtained Google and Bing search queries for just one week and I think you had between three and four billion entry points if I recall correctly. It seems to me that that is the type of data gathering the European Commission couldn't do, right, to get access to this data without a proper investigation behind it. And so, I'm wondering as to whether you think that's all agencies should actually push to obtain a similar power to just be able to go to the private companies and say, "Okay, we want to see the data".

Stefan Hunt

I mean, we get in the context of market value, so we had tangible legal power. I do think market studies are very useful tools, because the commission is moving in that direction with their sort of focus on market investigations. I think that could be potentially...I may not be as maybe as well up to speed on that. I am aware that's bad. But I think that could potentially be a very useful tool indeed. Yeah. So, we use market studies in complement of our competition cases. It can allow you to focus on the bigger picture.

Thibault Schrepel

Indeed. And so, one of the new amendments in the DMA is apparently moving towards the direction of forcing the European Commission to publish annual report as to the state of digital competition. So, this could be a way for the European Commission to get those market studies. Another question I have, and I will start with you Filipe, concerns procedural fairness. This is an important subject. I think it's fair to say that when an agency is facing one of the big tech companies, they do have computational power. So, it's fine. But in the case of private enforcement, this could be a problem. Felipe, are you discussing within the agency the change of procedural rules to kind of frame and constrain your work, or do you have total freedom to just use compositional tools, regardless of those issues?

Felipe Roquete

Well, in the absence of the federal regulation of artificial intelligence and other technologies, we are free to use any technology we want. But of course, we are concerned about how we could translate the results to our decision makers and for other stakeholders as well. For us, it's a challenge. But nowadays, we can use anything we want.

Thibault Schrepel

Okay, Stefan and Yann, I'm curious as to whether you do have those discussions regarding the procedures and how to frame your work so that it is fair to the companies being investigated?

Yann Guthmann

It's a very important question. In my unit, we do legal analysis. It's very important to have at least one lawyer in my team, because she can be the one that does this analysis.

Thibault Schrepel

Great! Stefan?

Stefan Hunt

Sure, it's really important. I mean, we have for example the Regulatory Investigations Protection Act in the UK, also within data protection legislation to do a data protection impact assessment. So those are risk assessment tools that we would use to assess what we're doing. I mean in particular, we've had less of an issue with methods but more of an issue with data, especially when we're taking any public data, we really want to make sure we've been through those procedures.

Thibault Schrepel

Yeah. All right. I think this is a question that we'll talk about for quite some time. I just want to move on to the last part of my questions regarding the interaction between the agencies and private companies. Felipe, I'm curious as to whether your agency will consider using a tech designed by a private firm. There are quite firms right now on the market developing computational antitrust solutions. Is it something you will consider or is it a no go because it's coming from the private market?

Felipe Roquete

Today, it's a no-go. We can change in the future, but today we are trying to develop everything in-house.

Thibault Schrepel

Alright, great. Stefan and Yann, I think it's a very important question. So, anything you want to share, Stefan?

Stefan Hunt

In principle, we would be willing to use tools from elsewhere. We haven't found any, beyond document review, that we think would particularly meet what we need, and we've been focusing on building our own bespoke capability. But document review is a great example here. It's a class we thought about buying some stuff off the shelf. I mean, I don't know whether you'd call a competition agency staff, you've got continuous active learning built into the document review software. But then where this is why I talked about those enhancements that we've made, the e-evidence review portal so I think mix of those things together. So, it's not necessarily a build or buy; it can be very much a combination.

Thibault Schrepel

Great. Yann?

Yann Guthmann

Yeah, I think it very much depends on the tool because some very technical tool can be used and there's no problem. But if you're using some unsupervised machine learning that you can't explain afterwards, and also you need to be able to explain your methodology to the judge at the end of a case. And so that's one big problem. If you can't understand what's going on, how can you explain it afterward? And I think that for all unsupervised machine learning techniques, we need to develop them in-house because it's kind of too dangerous to externalize it.

Thibault Schrepel

It's funny because we have a mirror with the question of liability for the private sector. If AI decides to collude with another company, then the question is, who's liable? And for you, the challenge is to be able to explain why the system reacted that way. Very interesting. Stefan, the next question is just for you. You've mentioned GitHub. I'm not aware of the possibility of making GitHub private. Does it mean that you're willing to work with companies towards providing them with such tools to enable better procedures?

Stefan Hunt

In part, we literally just started to work with a couple of parties right now. What we haven't done stuff is where we're sharing things with other parties in the market yet, and there's certainly the possibility so we did build one piece of software that does a particular type of local area analysis that can get used in merges quite a lot, so we've bought some software that really significantly increase the speed of doing

that. But the issue there is we as an organization, just really reigning back from that from doing that type of analysis, we're doing so much than we actually used to.

And it's just not worth us maintaining that software. However, it's a good example, because that is something where, in principle, I think we would have been willing to open up that software to everyone in the market. It could, you know, I mean, relatively defined and our system is about kind of establishing facts. So, in that particular case, actually if the same software's available to everyone could allow merger cases to proceed more quickly. Now, for the right kinds of tools. Yes. Obviously, anything to do with screening and detection, no.

Thibault Schrepel

No, of course. Yeah. All right.

Yann Guthmann

I wanted to add that at the French authority we have a private GitHub.

Thibault Schrepel

Oh, you do? Okay. Yeah. Very interesting.

Yann Guthmann

I'm using that especially for the project I presented.

Thibault Schrepel

Yeah, that's very interesting. I'm just giving you about 30 seconds to conclude and convince people to vote for you, in a Presidential debate fashion. There is one more remaining question if you want to address it. Rafael is asking if the tools can be used by students for hackathons. I guess you kind of answered the question right now. But in any case, if we could start with you, Felipe, is there is a final thought you would like to share with us?

Felipe Roquete

I think it's a very interesting time. The debate CodeX and you organized is very useful. Thank you for the opportunity.

Thibault Schrepel

Yes, thank you very much. Stefan?

Stefan Hunt

I'm really excited to be speaking on this panel, and to be starting...we've been internally focused for the beginning that was really important. All the reasons we've discussed about building up support and building the unit. We're starting to focus more, we've been talking to other agencies actually throughout. But even more recently, the last year we've been increasing the amount we speak to other agencies. Yeah, I'm really excited by the fact that we're now speaking more externally. I think there's a really great opportunity, as I said, for us to work together. We can either be...we could be making steady progress. We could be kind of a local train, or if we can work together, we can build up these networks. We can all benefit from the low marginal cost of showing digital products. Maybe we can be the TGV, we can be going much faster. And I hope that together we can make sure that we're building TGV.

Thibault Schrepel

Yes, thank you very much. It will also be very interesting to see how companies are going to compete and develop adversarial systems but that is something for next year's conference. Thank you very much.

It's now almost 9pm here in Paris and Amsterdam. So, thank you to the three of you very much. What I would like to do now is to conclude the conference by sharing a few thoughts. And I promise you, it wouldn't take more than just five minutes.

We've reached the end of the Computational Antitrust First Annual Conference. Before anything, I do have to say thank you. Thank you to the CodeX team, to Roland Vogl for his support. Thank you to all the speakers. Most of all, thank you to all of you who came in great numbers, in fact, by the hundreds, which was a beautiful surprise. And thank you to the entire team at the Computational antitrust project, Teodora, Alex, Maura, Alexandra, Yann, Glen, Alessia, Kirill and Ben.

I am pleased to report that the conference has been recorded. The videos will be made available in January, along with the transcript of our discussions, so make sure you subscribe to our newsletter to be alerted of the publication.

Now let's just take two minutes to reflect upon what we've heard during these three days. On Monday we talked about the objectives of computational antitrust. We have delimited the scope, and it appears that computational antitrust is particularly useful to detect anti-competitive practices. It can also play an essential role in connecting competition agencies to market reality. We have heard that computational antitrust poses technical problems and questions concerning fundamental rights. Lawyers have an essential role to play.

On Tuesday, we explored several projects that will serve as pilots for agencies and companies, the majority of our discussions yesterday revolved around machine learning, and, more specifically, the computational antitrust ability to connect antitrust agencies, not with the outside world, but with themselves, their own case law. We discussed the absolute necessity of creating a representative body of data, which by default, as we mentioned, is complicated when the data is jurisprudential. That said, computational antitrust can provide a better understanding of doctrinal

movements, case law, and we saw together that it could also be used to draft legislation and court decisions.

Today, finally, we have analyzed how competition agencies take action based on the research published with our project. We discussed merger control as a natural candidate, as agencies and companies have a common interest in improving it. The same is true for antitrust remedies because they give rise to costly follow-ups on both sides.

My personal take is the following: I think we cannot afford not to explore and build computational antitrust capacities. We have explored many places where to start such as merger control, cartel detection, etc. I think it's fair to say that in two years, will come the time for us to discuss how computational antitrust will change not only the methods but also the substance of antitrust, getting us to explore complexity theory, evolutionary economics, and more.

Overall, I'm confident that the future of antitrust law is bright. We have seen during the course of the last three days that we can create convergence in a highly polarized field. Convergence, in my view, requires developing the right tools, to gather the right data, and also to stay kind with each other. So, take care of yourself and, if you can, of someone else, too. Bye, bye.