

Transcript: 1/02/2021 (checked by Malvika Jolly)

Databite No. 140: Governing an Algorithm in the Wild

David Robinson and Alex Rosenblat

Recorded: 12/08/2020

Alex Rosenblat:

Welcome to Databite No. 140! My name is Alex Rosenblat, I'm a senior researcher at Data & Society and I will be your host for tonight alongside my team behind the scenes, C.J., Eli, and Rigo. For those of you who don't know us yet, Data & Society is an independent research institute studying the social implications of data and automation. We produce original research and regularly convene multidisciplinary thinkers to challenge the power and purpose of technology in society.

We will be spending the next hour together, so let's get ourselves grounded. Data & Society was found in Lenapehoking, the historic land of the Leni Lenape people, a network of rivers and islands in the Atlantic Northeast we now refer to as New York City. Today, we are connected via a vast array of servers and computers around the world, facilitated by the dispossession of Indigenous land acquired through the logic of white settler expansion. We uplift the sovereignty of Indigenous people, data, and territory and commit to dismantling ongoing practices of colonialism and its material implications on our digital worlds. You'll notice a link in the chat to native-land.ca that directs you to more information about occupied lands. If you haven't already, use the chat feature to share your location.

And now it is my *distinct* pleasure to introduce you to our speaker today, my longtime friend and colleague, David Robinson, who is currently a visiting scientist in the Artificial Intelligence, Policy, and Practice (AIPP) in Cornell's College of Computing and Information Science. His research centers on the design and management of algorithmic decision-making, particularly in the public sector. He believes that effective governance of these socio-technical systems will require collaboration and mutual adaptation by the legal and technical communities leading to changes to both institutional and algorithmic design, as well as the generation and use of new

types of data. Through his work he aims to contribute to that effort. Welcome, David, thank you so much for being with us today.

David Robinson:

Thanks, Alex. Let's get started. Thank you, everyone, for joining today. I saw in the chat that Professor Oscar Gandy is with us and I'm deeply honored, Professor, to have you joining us today—your work is an inspiration to my own efforts. Let me go ahead and try and share my slides. Can people see them? Let's see.. Alex, look good? Okay, great.

Alex Rosenblat:

Yes.

David Robinson:

So rules implemented in software deciding things that are both technically intricate and ethically fraught are more and more common. Situations where there are no perfect answers, people have been calling—in policy and in academic circles—for the public to better understand and to shape many of these rules. Could be when someone is arrested: How could their level of “dangerousness” be defined— if at all measured? Or if someone has a disability: How much and what kind of public assistance do they need? Which news updates should be routed to a particular person—which sources can be trusted? There are a lot of debates happening about whether and how laypeople can understand—and can have a voice in shaping—systems like these. The reason that I care about the story that we're going to explore together today is because it shows us that meaningful lay-involvement is possible and illustrates conditions that are conducive to such involvement.

So, today's question isn't so much *which systems* need the kind of engagement by laypeople that is demonstrated here, instead the question is: What can that engagement look like in practice and what can we learn from it? I'm going to begin with a story, then tell you about the challenge of kidney transplantation and kidney disease, then we're going to look at a remaking of the algorithm—a rewriting over a 10-year period—of the algorithm that determines who gets a

kidney in the United States, and lastly, we're going to explore what I would suggest are some lessons that could be learned from this experience. But if you only remember one thing it would be this, I hope: that technology constantly sparks new, ethical questions and technologists need not answer those questions alone.

So, our story begins at this airport Marriott in Dallas, Texas on a chilly day in February in 2007. There's an all-day meeting taking place in the basement ballroom. Hundreds of people who are involved in kidney transplantation—surgeons, nurses, social workers, transplant recipients, and living donors—because, of course, unlike hearts, for instance, most of us have two kidneys and yet a single kidney is enough to keep a person in good health, so there are many living donors for kidney transplants. Right now it's just after lunch, and Clive Grawe, a 54-year-old traffic engineer from Los Angeles, is speaking. He is the first person [to speak who is] neither a medical doctor nor a PhD. He has a rare genetic disease called polycystic kidney disease, which means that over time his kidneys break down and he'll need a new one. What the meeting is about is an effort to revise the nationwide algorithm that matches kidneys from deceased donors to waiting patients. Today there are more than 100,000 people waiting for kidney transplant in the U.S; the algorithm coordinates approximately 15,000 transplants each year. It's often called a "waiting list" but actually it's not a first-come-first-serve queue, it's a matching process that incorporates medical, ethical and logistical factors in deciding who will be offered each organ. This has been done by a nationwide algorithm—a nationwide piece of software—for more than 30 years, and this meeting in Dallas, taking place in 2007, is midway through what turned out to be a 10-year saga that rewrote the algorithm. So, I think this is a very interesting situation and it's one that incorporates in real practice many governance proposals that are frequently proffered today, including, for instance: community input, transparency, simulating the consequences of a decision logic, auditing after the fact. So, the question that animates our time together, today is: What can we learn from this?

The proposal that Clive Grawe, incidentally, is responding to is that the organ should be allocated to maximize the amount of life years that are saved. That's a very intuitively appealing idea for many, but Mr. Graw argues that, for someone who knows that they're going to need a

kidney later in life, that could be a perverse regime because, in effect, it could punish him for not being young when he needs a kidney—for having taken good care of himself—and the same is true for his daughters who, unfortunately, inherited the same condition. And after his appeal that pointed this out, the attempt to use a purely-benefit-based, life-year-maximizing algorithm was abandoned after this meeting. But, first, let's get some context on this challenge: Ethical challenges depend on what our technologies make possible and as the technologies change, new moral challenges can come up that we need to respond to, and our answers might change over time as the horizon of *what is possible* also changes. Kidney transplant beautifully illustrates all of this and it actually begins—this moral challenge begins—with the introduction of Teflon, the synthetic nonstick surface in the early 1960s. What you're looking at here is a diagram from a 1960 paper that described the invention of a [Teflon shunt](#) so that someone could repeatedly have their blood circulated through an external machine, because one of the key functions of kidneys is to remove toxins from the blood. Believe it or not, dialysis first was attempted in World War II in Nazi-occupied Holland—and it worked quite well *except* that getting blood in and out of the patient was quite difficult, and every time they would do dialysis, they would need a different place to put the needle—so they could only do it for about a month. But after this Teflon shunt was introduced, it became possible to dialyze the same person repeatedly, indefinitely, which was a tremendous breakthrough. It basically turned kidney failure into what could be a chronic—rather than a fatal—disease.

But Dr. Belding Hibbard Scribner in Washington State who invented this device only had four dialysis machines. He and his staff were overwhelmed by desperate patients and their doctors. And doctors, of course, fight for the life of their patient, and so rationing care—deciding which of their patients should get access to a life-saving machine—is not a very comfortable role for most physicians. Really, that's an ethical question. And Scribner decided to resolve that question of who should get the scarce machines by creating a committee, which you see depicted here. Doctors would say who's medically eligible; and then these laypeople would pick from among the medically-eligible patients. This did not go over well with the public—it was seen as too ad hoc. The evidence about this committee suggested that they liked men who were income earners with families to support. A couple of scholars memorably wrote that the Pacific Northwest is no

place for a Henry David Thoreau with bad kidneys. So there's lots to dislike about this approach, but there's also something profoundly honest about it, that says: The technology of dialysis—of cleaning the blood—might be complex, but the hard, ethical choices belong to a wider community and not only to the technical experts. And by the way—in order to get the ethical input that we need— it is necessary to bring laypeople into the fold; to get them up to speed about what these choices really are.

But even with this happening, the best way to solve kidney failure remained transplantation. So, while dialysis was developing, transplant developed around the same time. Believe it or not, the physical act of surgically transplanting a kidney is something that was really resolved in the early 1900s—how to go about it. The really hard part is: if you take an organ from one body and put it in another, the recipient's immune system tends to reject the organ as foreign and attack it. And so the very first long-term success with kidney transplantation was between identical twins. Of course, if you have a transplant inside of you, you don't have to spend 12 hours a week hooked up to a machine. You don't have to constantly yo-yo between a feeling of health and a feeling of toxicity and lethargy. So this was great but, for a while, it seemed you could only do this with very close family members or, ideally, as in this case, identical twins.

But then, in the early 1980s, the introduction of a new drug called Cyclosporine changed the situation completely. It could suppress the immune system of the person receiving a transplant, so that someone could get any transplant *from a stranger* that could work for many years, potentially. Now, there were still a bunch of factors that would influence how likely a transplant was to succeed and for how long, including [that] there would need to be a compatible blood type between the donor and recipient, and there were conflicts between the immune system of the recipient and the proteins on the kidney—which are known as antigens because they spark an immune response. But, on the whole, many, many people could be a suitable recipient for the same kidney and that created a real logistical and moral question about who would get the organ. So, at first, a process called tissue typing played a very important role in deciding who would be offered which kidney; and this was basically an effort to detect which proteins each recipient would react against, and to prioritize, to receive each kidney, people whose immune

systems were unlikely to reject that kidney. But one of the things that was observed, in all of this, was that the proteins that were causing these reactions were correlated, to some extent, with race. And so, with a much higher burden of kidney disease in the African American community—partly due to significant social determinants of health, so things that cause kidney disease include things like environmental stress and diabetes and obesity, that are differentially present in different communities—there's about a triple rate of kidney failure in the African American community as [compared to] the white community in the United States, *so much more need*. But the biological tests that they were using to forecast how long a transplant would last, tended to—not always, but in general—have an effect of a thumb on the scale toward same-race matching of kidneys. And, of course, the donor pool resembled the overall population, so there was a real disconnect there. There are a number of uncertainties that you might have about this: Factually, how much does this antigen-matching aid survival? And, similarly, how much does this antigen-matching process disadvantage underrepresented minorities in receiving organs? And then, once you've got all the facts, there's some kind of a normative tradeoff involved: How much does the survival and the disparity—how much does each of those factors matter and how to trade-off between them?

What happened over time is that the role of this protein-matching process decreased. And this is one of a number of things we can go deeper into if people have questions. But, just to take a step back, this process of transplants with strangers—how did this develop? Up at the top of your screen, this is the [National Organ Transplant Act](#), which was passed in 1984, shortly after stranger transplants became medically feasible. This says, of course: “Establish a national system through the use of computers and in accordance with established medical criteria to match organs and individuals included in the list,” ie, an algorithm. So then there is a panel that is brought together to figure out how that algorithm ought to work, and it says: “criteria for prioritization should be developed by a broadly representative group,” and then, lastly—this is the current bylaws of the organization that makes these rules—which say that transplant candidates, recipients, organ donors, and their family members must be represented and must hold at least one quarter of the seats on this board that approves the algorithm. And, by the way, there's a whole governance process with notice and comment, with proposed changes, the rule is

transparent, annually extensive data is published about who's getting transplants and who's not and which *parts* of the different rules are causing people to get—or to not get—transplants. But circa 2003, although race equity has been improved somewhat from an earlier baseline, the system is extremely complicated—what you're seeing here is a quarter of it—and it's not making best use of the organs, in the sense that it would often happen that a much older recipient would receive a younger and healthier kidney, and a younger and healthier recipient would receive an older and less robust kidney which meant that they would likely need a second transplant later. So, there were some opportunities for greater efficiency that were clearly being missed, and the system had become complicated and hard to understand *even with* all the transparency.

So, how did they make the new algorithm? I'm about to summarize a decade of debate in a few minutes. There are a number of other changes that I'm not going to highlight here that include transplant across blood types and a change in how waiting time is calculated, but the central debate— the big question in this decade of debate—was: How to balance between maximizing between benefit and making the system fair?

Here's a cover image from 2009 from the midpoint of the debate from the leading journal of transplant medicine. There were a series of three different proposals over time that were seriously considered for the new Kidney Allocation Algorithm: The first one, LIFT, is the one we saw Mr. Grawe, the traffic engineer, arguing against at the beginning of our story. And what this graph depicts is how much benefit in added survival to transplant candidates get, given different ages of the candidate?

David Robinson:

The distance between these lines on the left shows how much added survival you get if the recipient is a young person; and, [on the right], how much added survival you get if the recipient is an old person. And so if you maximize life years from transplant, then, basically, you end up giving to younger and also healthier recipients with fewer comorbidities. Here, what you can see is the results of a simulation that the Scientific Registry of Transplant Recipients did—these are not the people who administered the algorithm, but a separate auditing and analysis group, a

different nonprofit organization that plays this role. The pink was the old rules; the light blue was the new rules. And this is, on the bottom, the age decade of the people who would get kidneys. And what you can see here on this slide is that the fraction of transplants going to people in their twenties would triple and the fraction for people over 50 would approximately fall in half. So, the admirable clarity around the age impact of this proposal led the committees that were looking at this to reject the proposal.

Second, we had a proposal called 2080, that used something called longevity matching. There was a problem, before, that some people would have much worse chances of getting a kidney under the LIFT proposal. So now what we're going to do is say: "Everyone has the same chance of getting a kidney as they had under the old system, and we're just going to change *who gets which kidney*—and the youngest, healthiest, 20% of kidneys we are going to give to the youngest, healthiest group of potential recipients. And then for the rest of the organs, we're going to age-match so that someone who's in their 50s gets a kidney from a donor in their 50s, and so on." This was a much simpler system, much easier to understand, but it was ultimately rejected as being age-discriminatory to have that age-matching piece.

Thirdly—and finally—the proposal that we live under today got rid of age-matching for that bottom 80% and went back mostly to waiting time. And the rules were different for these KDPI numbers [Kidney Donor Profile Index]—that's the quality of kidneys—so there were four different, basically, quality of kidneys that each have a different sequence. And, by the way, they change the way that waiting time was calculated: it used to be you would get extra credit for how long you had been waiting—which is only one of the factors—based on when you were first added to the transplant list, but they realized that this really rewarded people for having access to healthcare and punished people who had less access to healthcare. So they changed this system so you would get waiting time from the time you started to need dialysis—which is a more objective medical indicator of medical need—as opposed to "When you were able to get yourself listed?" And so in the new system, one thing you see is that the waiting time calculation, there was a large population of people with less healthcare access who had gotten onto the list later in the course of their disease and got a tremendous boost from the change so that—this is

basically the per-person likelihood of getting a transplant—and you can see that the racial gap between candidates of different races, in their unit likelihood of receiving a transplant in any given month, greatly narrowed after 2014, when the new rules were implemented.

So, I thought that was the story that I had for you—a pretty clean, pretty simple story about this deliberative process leading to this incremental improvement in the algorithm—but then I encountered a surprise twist, and that was [a lawsuit filed by Miriam Holman](#) in 2017. So what you're looking at here is a map of attachment areas—the zones in which transplants are gathered and first distributed. And basically, Miriam argued [that] she was in New York City and needed a lung but because New Jersey's a different zone, a lung donor who was quite close to her [geographically], potentially that lung would go to someone further away from the donor, somewhere in New Jersey just because it's in a different zone, [rather] than to Miriam in New York City who needed it—and, of course, unlike kidneys and dialysis, if you need a lung, it's life and death immediately.

And so she basically said these geographic borders are arbitrary—that everything is based on including the Kidney Allocation. And in this ten-year process, geography wasn't discussed, hardly at all, in the thousands of pages of material I read and in the series of interviews that I did with people who were involved in this process. And, indeed, that same data analysis group—what this is is a look at factors that are not supposed to matter to how likely you are to get a kidney—so, for instance, here's education, here's age, here's ethnicity. And this largest bar, the factor that makes the largest difference—that shouldn't make a difference—is donation service area, that's which geographic zone you are in.

And so Miriam files her suit in November of 2017 and *within days* the courts force the Organ Procurement and Transplantation Network (OPTN) to eliminate its geographic zone. So whereas much smaller changes were debated for a decade, this transformation of the system—using concentric proximity circles around the hospitals where the donations are—was implemented in less than a week. And you can see here—this is for kidneys—that the median wait time across groups of states varies by more than a factor of two. What this shows us is that morally-important differences can be offstage when we're having these detailed debates about

algorithms, and this was an illustration of that.

So, I'll quickly give you a few of the takeaways that I am left with and I'm eager to hear what impressions and questions you are left with by this material. One thing is, as I was saying, algorithms direct our moral attention. So, this dial between favoring, basically, maximum benefit by transplanting into young and healthy people who have the most to gain, versus fairness, of giving everyone an equal chance—that's a dial you can turn—and we saw that in the progressive different versions of the algorithm. But there are other things offstage, like what about these geographic zones that no one had the political will to reconsider during the algorithm development process? Or, how do people get onto the list in the first place? It turns out that many people without strong access to healthcare never end up getting listed for transplants even when they could benefit from it. By the way, this whole process and algorithm are about addressing the shortage of kidneys to transplant into waiting patients, but what about things that could remove that shortage? Things like, for instance, artificial kidneys that are implantable—which is something that's being worked on just a few miles from here at UCSF—or transplants from animals, which, again, is something that people are researching actively.

The second lesson that I took from this process is that debate creates opinion. I think we often treat public input as if it were out there waiting for us to go detect with some kind of opinion-detecting procedure or machine. But what happened in this case, as I read the evidence, is that in hearing one another out, stakeholders gradually developed a shared belief in the legitimacy of the long and costly deliberations in which they were all engaged. In the end, over a ten-year period of learning and debate, a community of people *changed its mind* about what was reasonable to do and gradually converged on some sense of agreement. For instance, after the LIFT proposal, community feedback decisively rejected a pure focus on maximizing total benefit, and when that extreme was ruled out, the window of acceptable policies was narrower, and the debate focused on how to improve benefit without radically reducing any group of patients access to organs. And I don't want to sugar-coat this: consensus was not achieved—as, I think, it never could be in a situation where not getting a kidney can mean death [and] the risks from dialysis remain high for every patient. It's not that consensus was achieved so much as a

kind of earned, mutual acquiescence to a particular set of values, compromises; people were sort of worn down until they were ready to accept the inevitably imperfect compromise that they had played a hand in creating and that was arrived at.

A third lesson: experts are critical. We can talk all we like about transparency, but transparency all the way to the level of understanding requires more than just disclosure—it requires expertise. Even with lots of transparency, there's a ton to learn, and firsthand knowledge of how these systems work is not something that can be replaced by any set of documents. Take it from me, having gone swimming in the voluminous documentary record: conversations with participants were an irreplaceable catalyst for me in even beginning to understand how these processes work.

And finally, the shared process of the analysis function, here, from the Scientific Registry of Transplant Recipients (SRTR), created a sense of shared awareness about how the system was working and about how proposed changes might impact the system and that, I think, enabled an informed debate to unfold. Without some shared sense by some trusted analyst, you would have ended up forcing debate participants to reinvent the wheel on an extremely effort-intensive [endeavor]—they do, every year, hundreds of pages of reports. This figure, for instance, this was the 13th in a list of 30 such analyses they published in a year that I was looking at. So, with that, let me just come back to that one thing that I'm hoping will emerge from this, which is that even though technology does constantly spark new ethical questions—technologists, engineers, the technical folks do not need to answer those questions alone. And with that, I'll thank you and let's turn to discussion.

Alex Rosenblat:

Thank you so much, David, that was a remarkably interesting presentation. I'm thrilled to hear it. I've been eager to learn more about this project since you discussed it with me long ago and I'm so glad to hear it come to fruition. I have a couple of questions to start us off, but then I will turn to the audience Q&A. You described a deliberate process for inching progress in stranger transplants around Miriam Holman's case, but her case also demonstrates how personal

narratives and legal interventions can accelerate changes to transplant governance structures, such as by challenging the arbitrary legacy conditions or overruling geography. I wonder if you could comment a bit more on the lessons we can take away from this remarkable history of multi-stakeholder processes and how we can apply it to present political challenges over governing algorithms?

David Robinson:

Yeah, that's a really interesting question. One of the things that's true about any kind of a new process that we set up, like creating a new committee to publicly give input, whether it's transplant algorithms or courtroom or whatever, is that it's ultimately subject to the intervention of the rest of the power structure. So, a court can come in and force a change or a legislature can come in and say: "Well, we created this committee but now we're revising its mandate." Or the executive branch can say: "We're interpreting this differently." Ultimately, all of those other levers of power are still going to be present. And I think that one of the balancing acts if you're in a committee situation—they try to achieve consensus so hard partly because, if they don't have agreement, then whoever has lost can go look for another of these ways to intervene. So, the idea that we can predict with total confidence which bureaucratic mechanism is really going to resolve the values debate, I think this cautions us against that.

Alex Rosenblat:

Thank you. One of the things that fascinated me in your astonishing history of stranger transplants is how social justice was centered in this governance, multi-stakeholder debate. For instance, there are real medical differences at the population level between people we group into different racial groups and that affects kidney-matching for marginalized populations. I wonder if you might further comment on how we can best continue to center social justice issues in questions of algorithmic governance today?

David Robinson:

It's interesting. Antigen matching is such an interesting case. It turned out—so if you have a

perfect match, like identical twins—it's still the case today that that confers a tremendous increase in the likely length of success of the transplant. But for levels of compatibility that are somewhere in the middle, the evidence is much more equivocal about whether it makes a difference at all. And, I think, for the doctors involved, there was a temptation to say: "Oh, we're going to use this medical fact to decide who gets the next organ." It sort of resolved for them a tension that is a very un-doctor-like thing they might otherwise have to do, which is make an explicit ethical judgment about who should come first. So there's sort of this desire to *invest*, in some cases, these medical facts with a greater clinical significance than they may, in fact, possess. But I do think, also, that this is a situation where the lived experiences of the people who are in the transplant system play an important role. So, for instance, the transplant physicians who drove these changes to de-emphasize antigen matching—many of them were African American physicians and transplant nurses and others.

And so, I think that representation matters and that the debate over which factors are relevant, for instance: does a community-level difference in representation matter if a single physician is saying: "Well, each of my patients is as important as any other." I think from that point of view—and I encountered some of this in interviews—there's a limited patience with this kind of group-level analysis, because I think, to some, it seems to smack of differentiating how much we value different lives. And, of course, people who would argue for such thought would say: "Well, we *already* act as if different lives had different value, so we have to be aware of it in order to respond to it." I think that debate is very much a part of the transplant conversation.

Alex Rosenblat:

Thank you. I'm going to turn to questions from the audience. Our first one is from Betsy Cooper—thank you so much for joining us, Betsy. She writes: "I can't help but think about the upcoming COVID-19 vaccine distributions and how algorithms may play a role in who gets them first. Any lessons to share from this case of kidney transplants that might inform the vaccine distribution?"

David Robinson:

That's a great question. And thanks, Betsy, for tuning in. I think that explicitness is valuable. I think it allows us to understand the tradeoffs that are involved. There are different reasons why you might prioritize different groups, healthcare workers, and the medically vulnerable. I think vaccine risks are maybe a little less acute—waiting a month to receive a vaccine can, of course, make a life or death difference—but when I compare that to “Do you get a heart that you need as a transplant or not?” or the more acute care question of rationing ventilators, I do think that there's a little more of a sense of time to be deliberate when it comes to rationing vaccine doses. But I would just say that explicitness is really valuable—and I've been comfortable to see a significant amount of different healthcare authorities, different countries saying: Here's who we're putting first *and why*.

Alex Rosenblat:

It seems that there is something common with vaccine distribution and the use of dialysis as a safety net for renal failure, and against the questions of urgency and priority in receiving a transplant.

David Robinson:

Yeah.

Alex Rosenblat:

Our next question is from Kelly Owens. She says: “Thanks for this great talk. I'm wondering if you could speak a bit more about how these types of algorithms—or systems—can be gamed and how that can affect equity. As you mentioned briefly, savvy doctors know how to get their patients on transplant lists long before they need transplants and try to game the system to do that. How do policymakers and developers think about gaming when creating these kinds of algorithms?”

David Robinson:

Gaming is such an interesting word, because it carries this normative charge. Like, “gaming” is

behavior change that gets you an advantage that you're not supposed to get. And so part of, I think, a debate about gaming is: Where do we draw the line between acceptable and unacceptable ways of seeking a transplant? I'll give you a concrete example: in the federal law—the National Organ Transplant Act—there is a protected right for each patient to be listed in multiple geographic zones. And so, if you're the top end—the really, really top end— of wealth and resources and means, you can pay to have the extensive—call it \$1,000—workup done in hospitals in a half-dozen different jurisdictions, and then fly wherever you get an organ first. So, for instance, [Steve Jobs went to Tennessee](#) is a famous example—a recent example of someone going where they needed to go to be on the right list. Whereas most patients—and particularly patients of social disadvantage—they're going to list, if at all, in the place where they live, and they're not in a position to move. That's not considered gaming because it's not against the rules because that's what the rules are—they let you multi-list. But I think the rules are, themselves, one might argue, unfair in that way.

Alex Rosenblat:

Thank you. From Catherine [\[indiscernible, 1:17:16\]](#) we have a question about how much visibility, if any, patients get into the algorithm: is there transparency around the decision factors and waits that are shared with patients and their families?

David Robinson:

Oh, boy, is there ever! If you mean disclosure, there's as much of that as you could possibly want, I think—including exactly what the formulas are. And, by the way, they're formulas on top of formulas on top of formulas, because each kidney has a quality score that is itself a prediction about how long that kidney's going to last, and then they group the kidneys by quality level and have different allocation systems for different quality tiers—and then there are different lists. So basically, every time a kidney becomes available, there's a match run that's looking at: Okay, who is listed who either [has] the right amount of immune-matching or [is] nearby? There's kind of this whole complicated series. If you recall that tree with all those leaves on it with different rules that I displayed earlier, it can be very difficult to translate the five-page single-

spaced list of different rules into a kind of gestalt sense of where the kidneys are actually going. That's one of the roles that this SRTR group plays that I think is very helpful—that they help to visualize and summarize and make tractable this blizzard of facts that are available about who's getting transplants.

Alex Rosenblat:

I can only imagine the stress of watching a data visualization of where the different kidneys are going at different times, or other organs—liver or lung geography. Our next question is from Amanda Lenhart: Is there any way to apply pressure to create this type of process to private applications of algorithms in healthcare? Thinking here about how insurance companies may ration or modulate access to certain types of care.

David Robinson:

Oh, that's really interesting. Insurance regulation is such an interesting field. There are state boards, for instance, that control the rates of different kinds of insurance. Actually, I'm thinking of Barbara Kiviat and [her work on car insurance](#) and basically taking morally-salient factors out of the credit predictions that are used to set car insurance rates, so legislatures have said: “If you defaulted on your loan was because you had a terminally-ill family member, then it can't be held against you.” That's car insurance and not medical, but I think it's a sort of parallel phenomenon in some ways, which is ethical regulation—or ethics-driven intervention in how insurance decisions get made. And I do think that this could be a useful model. One thing to say about what makes the transplant case special—and might make it hard to “transplant” this bureaucratic procedure into other domains—is that organs are not for sale. They're a non-market resource. They're donated. And so having a system that inspires public confidence is seen as mission critical. And I would say, by the way, that it's not obvious to me that our policy of not ever rewarding someone financially for donating the kidney is the wisest possible policy. I find the debate about whether or not to have, for instance, single-pay or government incentives available, I think is a very interesting—and separate—conversation.

Alex Rosenblat:

We're moving now towards AI and machine learning and away strictly from algorithms in our next, upcoming questions. From Edna Egal we have: "While the most recent update to the kidney allocation algorithm seems to address black box concerns, how do you think AI will affect regulatory and governance mechanisms for autonomous machines and complex algorithmic systems like this one?"

David Robinson:

I'm optimistic on this—I think—for the following reason: no matter how complicated the algorithm is, there is a human somewhere that has to decide how it's going to work, whether or not its choices are acceptable. So, for instance, if it's a self-driving car type of a scenario, somebody who works for the company that makes self-driving cars is iterating over different adjustments to that algorithm, is assessing, in some way, whether it's doing the right thing or the wrong thing. There has to be some place where some human is deciding what's acceptable. And so, I think the drive to regulate is, in essence, the drive to open the door to wherever that conference room is—and that whiteboard, where people are sketching, "Do we want it to work this way or that way?" And my sense is that no matter how complex the algorithms get, there's always a whiteboard somewhere—it's out there. The challenge—for regulators and for people who share this interest in how these things are governed—is to figure out how to get into that room.

Alex Rosenblat:

I love the idea of starting each interrogation of algorithmic opacity with the question: "Where's the whiteboard?"

[Laughter]

Our next question from Annmarie Giblin is: "In your opinion, is algorithmic governance possible in an automated machine-learning environment?" You address this a little bit in your previous response, but I think this one's a bit more narrowly-tailored.

David Robinson:

I do think that this—“there's a whiteboard somewhere” is a part of my answer—so there's a way for humans to understand what's happening. The other part is: true understanding requires infrastructure. It's not enough just to disclose some stuff. You've got to have a community of experts that knows how to understand what's going on. You've got to have to have efforts at sense-making, to use kind of a Data & Society-flavored word. That's where people are analyzing and are saying: “Okay, what does this mean? What is this telling us?” And I think one thing that's very helpful in the transplant case, and would, I think, be helpful in other cases is to have some of that analysis work—because it is work, it's effort and resource-intensive—be done once and then shared widely. So that if I want to participate, I don't have to reinvent the wheel and employ my own building full of statisticians in order to discern what's really going on.

Alex Rosenblat:

I'm going to combine two questions that sort of capture the gist of each other. From Jumana Abu-Ghazaleh and Kamala Hayward Rotulien: “Thank you for your talk, and we want to know a little bit more about how the ethical questions can be addressed at the level of algorithm development or, in other words, how should the people in the room with the whiteboard be trained? What should they know? What qualifies them?”

David Robinson:

Right. I think that's a great question. I feel like “impacted communities” is kind of a bad answer to this question in a way? There's a grain of truth to it, but I think it's complicated, like “Who is the community?” is a hard question. So, you know, people will say: “You know, you should hear from patients.” And then we could ask: “Well, which patients are you likely to hear from? Who has time to go to a community meeting on a Tuesday evening? Probably someone with, relatively, a lot of access to care and resources and so on.” And, similarly, I think even getting input from doctors in the case of transplant that is useful, or from nurses, there's almost this kind of—I think, this anthropology piece. You have to go among the people that are where the thing is happening in order to even have a well-rounded opinion about *whose input* is needed at the table. I don't think you can easily discern *a priori*—from far away—whose interests are at

stake or what the situation looks like. So, I would suspect that this debate over who belongs at the table is, itself, something that needs to happen kind of particularly for each given system. And I think if I were being parachuted in as a data scientist into a completely unfamiliar context, what I would look for, or look to generate, would be an informed conversation about: “Okay, what is this like for the people who are involved in it? Who are impacted by it? Who are those communities?” Almost a snowball sampling—where you might have some idea about who might have something to say about this, you talk to them, and part of what you ask is: “Who else might have something to say about this?” And you sort of build some picture.

Alex Rosenblat:

We only have time for a few more questions. I'm going to ask two of them combined and return, finally, to Leif Hancox-Li's question which takes us back to the centerstage of COVID. Before we get there, the burning questions we have right now are around how to build successful collaborations: “Is there something this story teaches us about how to structure a successful collaboration between technical and layfolks?” is a question that Ben Blum-Smith is asking, and Amy Chen is adding: “Is there anything unique, you feel, to the culture of the world of people working on kidney transplants—policymakers, advocates, patients—that you think made for this type of collaborative and progressive process?”

David Robinson:

I would say one thing about transplant and kidney medicine is that people on dialysis have lots of time where they're in the medical establishment—physically, 12 hours every week—plenty of time to think about and work on—there's a lot of facetime that people have. And these committees, by the way, that were meeting to deliberate—there was a fair amount of in-person meeting that happened. I think face-to-face, it's harder to dehumanize or abstract away from what somebody else is thinking or saying; these were sustained over many years, these discussions. Sometimes, for instance, I would talk to an activist who would work on this, and find out that their personal physician was one of the physicians who was doing policymaking. So it's not a huge universe of people. And I think the sustained interaction, over many years, of

particular people with each other and the relationships that were part of that was important to this.

Alex Rosenblat:

So, all we need is time and face.

[Laughter]

Our next question is from Leif Hancox-Li: “I like the idea of algorithms directing moral attention. Going back to the COVID vaccine situation, what do you think are some offstage questions that the dominant way of framing distribution has neglected?”

David Robinson:

That's a great question. I'll stall for time by saying, I always feel disappointed when STS folks say: “This raises values issues or has assumptions.” And I always want to know: What are the assumptions? It's a great challenge for all of us. I would say there are questions about who bears the risk of going first that, maybe, are not as much in focus. There are, I think, questions about what to do with the greater safety that's created once a significant part of the population has vaccination. For instance, will there be heightened pressure *then* for people to go back to work in ways that are still hazardous for them, because the hazard has been diminished? And I think, also: How long do these initial promising results hold up for? But I think, instead of, individually, trying to do that, if I could make all of us take a few minutes and think about: “Well, you know, what are we taking for granted here?” I feel like it's a fruitful exercise.

Alex Rosenblat:

I think, in particular, your work on how trust is facilitated in the medical and governance establishment to equally or equitably distribute these valuable and yet non-monetary assets—these kidneys and lungs and liver—and think about how that might appeal to vaccine distributions, and how to galvanize people to both trust in the process and in the vaccines, themselves. Those would certainly be thinking points for the future. I think we're rounding out

on our Databite today. So David, I want to ask if there are any final remarks you want to leave us with?

David Robinson:

Just expressing strong gratitude to everyone who was here. There will be a paper of some kind. One of the challenges has been how to position it in a discipline or what exactly to publish, but I'm very hopeful that, in the next six months or so, the substance of all this work will be citable and public. And I'm sure Data & Society can help to keep interested folks in the loop, in one way or another. And I would just say that if you want to talk about this set of work or ideas—I'm thinking about them very actively and always appreciate that—so, feel free to reach out.

Alex Rosenblat:

Thank you. I have had the early pleasure of reading this paper and it's absolutely fascinating. Everyone will be very excited to see it in print. For now, I'm going to conclude our session today by thanking you, David, for your excellent work and presentation. And thanking CJ and Rigo and Eli, who have helped to put on this amazing event—thank you so much for all of your efforts. And, thanks to you all for joining us today. This was our final public event of the year. We hope that other years are better and stay tuned for more exciting programming ahead. The paper for this presentation will be coming out in the spring, as David mentioned, and if you need to cite it in advance, please contact events@datasociety.net. Thank you everyone. Have a great day.