

Data and Humanity (with Mimi Onuoha and Lam Thuy Vo)

Annie Galvin (AG): Hello, and welcome back to *Public Books 101*, a podcast that turns a scholarly eye to a world worth studying. I'm Annie Galvin, an editor and producer at *Public Books*, which is a magazine of arts, ideas, and scholarship that is free and online. You can read the magazine at www.publicbooks.org.

Natalie Kerby (NK): And I'm Natalie Kerby, digital content associate at Data & Society. Data & Society is a research institute that studies the social implications of data-centric technologies and automation. You can learn about our work at <https://datasociety.net/>.

AG: This is the third season of our podcast, so if you're listening for the first time, I invite you to subscribe to Public Books 101 in your podcast feed and listen back to season 1, which was about the internet, and season 2, about the novel in the 21st century. This season, we are excited to partner with Data & Society to explore the past, present, and future of human life being quantified as data. Natalie is your host this season, so I'll let her take it from here. Thanks for listening.

NK: In this season, "Becoming Data," my guests and I are considering a few main guiding questions. How long has human life been quantified as data, and in what contexts? What are some major implications of humans being quantified or measured as data? How are people pushing back against the datafication of human life, work, health, and citizenship, among other things?

Today, my guests are Mimi Onuoha and Lam Thuy Vo. We'll be discussing the ways that people collect data, how people show up in data, and the inequalities and harms that can result from these practices. Both Mimi and Lam take a creative approach to thinking about data. They have both played the role of collector and collected. Let's get into today's conversation with Mimi and Lam.

NK: Alright, so thank you so much both of you for being here today. I'm really excited for this conversation. So if you could say your name and tell our listeners a little bit about the work that you do that would be great. So, Mimi, why don't we start with you?

Mimi Onuoha (MO): Sure, my name is Mimi Onuoha. I am a media artist and I make work about what it means for the world to take the form of data. And I do that through a lot of different formats, but lately I've been gravitating more towards video, audio, installation, and I still do prints as well.

NK: Great, and how about you, Lam.

Lam Thuy Vo (LTV): My name is Lam Thuy Vo. I use data to investigate the systems that govern our lives and that can take shape in investigative journalism, but also in data visualizations or more whimsical essays around the nature of humanity, as seen through the digital footprints that people leave behind.

NK: Great, so yeah, I think you both have very creative interventions into data, and so I'm very curious to ask you the question what is data? Because I think kind of depending on a person's practice, they might approach that, the definition of data differently. So Lam, why don't you start first.

LTV: Okay. So I usually try to use a friend's definition of data, because I think it really blew my mind when she first brought it to me, and it is the idea that data is an interview done by someone else on a massive scale, right? If you think of the Census Bureau's Decennial Census, you will see ten or so questions on this one form that is getting passed around to thousands if not millions of people, right? And so, to some degree, someone wrote down those questions and devised an interview that is going to be done with millions of people.

NK: Great, and how about for you, Mimi, what is data?

MO: Yeah, well, I have a long history of working as an educator too and so I worked in a lot of community spaces, but also in a lot of universities, and this is a thing, this having to have a definition of data is something that comes up at a lot of these spaces, and so where I began from was this definition that I used to really like from Mitchell Whitelaw, who is an Australian academic, who says that data are measurements extracted from the flux of the real, and that was why I sort of began years ago, that's what I used to say. And I like that because there is a kind of poetry to it and also this use of the word extracted I think is interesting and sometimes does some interesting work within there, but I've kind of shifted to an easier and even more simplistic perhaps definition, which is just that data are the things that a group measures and cares about. And I like that because of these two different sides that it has, where first this idea of things that can be measured, this need to really forefront the quantification that is necessary for thinking about data. But then balancing that too with the things that a group cares about, so this gets at some of those questions of intention. This idea that it is not just that this thing comes, comes out of anywhere, it is not, this isn't oil, it is not data, data is just this existing resource. But know that there is some kind of process or act that has to get something into this form, and so often that is some group being incentivized in some way to do that.

NK: Yeah, I think what comes out in both of your definitions is how much you think about like the people creating data sets or the questions that people are trying to answer with data sets, and so my next question is kind of more about your own personal interaction with data. I think both of you have very kind of personal origin stories about how you became interested in this subject, so I'd love to hear about those. Mimi, can you tell us about your cat calling project and when did it start and who was it for?

MO: Right, okay, so this was the sort of small project, I often call it an intervention rather than a traditional artwork. And it would have been, I want to say 2013, honestly what is time, everything blurs together, but I think it was 2013. And that was, that was really interesting because that was a project that was done for me very specifically, and that is kind of why I say this is an intervention, I really didn't do this with a sense of anyone else in mind and what was happening was that this was, it was a summer when I was just getting cat called a lot and I wanted some kind of response to have when I was being cat called, because the response that I

felt like I was having just was not sufficient in that I was feeling kind of, feeling kind of weird and wishing I could say something back, but then I just didn't at that time.

And so it became this moment where I thought, okay, I can use this distancing effect of a lot of technology, I can use this effect of creating distance, and I can kind of use that to my own advantage. So, I did this thing where I, anytime somebody would cat call for me for that summer, I would give them this little piece of paper and it had my phone number on it, except that it wasn't actually my phone number, it was this number that I had hooked up to a server and that I had pre-programmed these different strings of messages that would be sent to people when they texted the phone number. You could call, no one would answer, but you could always text and then you would receive one of these answers. And so I did that over the course of the summer, and as I said, the whole point of this was to really change the way that I felt was to be able to have a kind of response if someone cat called me that I didn't just think about it, I was like here, take this phone number, and then I could watch this strange interaction play out between the things that I had programmed and whatever the strangers were saying, and I should say, this was early, this was years ago and so it was before bots were a thing. So it was a different moment.

But what was interesting about that, about that project and also part of the reason why I only call it an intervention is that it didn't really resolve because I just sort of realized something else that really just captivated me from it, which was at the end of the summer what I had was this kind of data set of all of my cat callers' phone numbers. And it was one that they had kind of provided just by virtue of texting me, but it was something, this artifact that I just had not meant to create and then what happened was that as I would tell people about this project, people would really focus on that artifact and that became the thing that was most interesting was the data, this data of this cat caller database people would call it.

And I ended up not doing anything with those phone numbers, even though loads of people were like, oh my gosh, you should text them this or spam them, put them on this, whatever, I ended up not doing anything because I was just so, like so really taken with the fact that I hadn't meant to create this thing, but it had emerged through this process, but also that by having the thing, it sort of erased the whole story of collection that brought it into being, and just looking at those phone numbers, there was no way that you could see this whole thought process I had had ahead of time and the feeling of like dread that I just hated having to just hand somebody a phone number and say anything, which is very strange, and it, because that entire, that entire part of it felt like it was missing just by looking at the artifact and that for me was what made me start to think, okay, there is something that happens in this act of data collection, there is a relationship that is kind of built here, and it is sort of erased by just having the data set. If you only get the data set, you don't have to think about any of this, but to not have that is to miss something.

NK: Yeah, that is super interesting to think about how once something becomes a data set, a lot of the context kind of falls away. I'm curious, what were some of the messages that you had preprogrammed that were sent to the cat callers?

MO: Oh my gosh, they were not good. They really weren't. I still remember telling my roommates the messages, and they were like, oh girl, what is wrong with you? So they were just things, some of them were really just overdramatic, like they were like, I wish you knew how

terrible your actions make me feel, which was so dumb. That's the one that I remember the most clearly.

NK: Right, I wonder like, because you like had this database, right, whereas like everyone looking at this data was like, okay, this is a group of cat callers, I wonder after receiving that text message how that person started to identify themselves? Were they like, oh, I'm a cat caller, you know?

MO: Well, they never, I mean, no one told them.

NK: Right.

MO: They never knew, and I told you I didn't do anything with it, but I will say, the majority of messages that I got back were just like people kind of cursing me out or being like, what the, you know, false promises, why did you give me your number? You know, it seems wrong, but I do remember getting a message or two from someone who was like, oh, I'm sorry, I didn't know you feel that way.

NK: Yeah, that is super interesting.

MO: But, who knows?

NK: Alright, so Lam, you have this project that is called the Quantified Breakup project. Would you tell us a little bit about that?

LTV: Sure. I think interestingly enough I came to a similar conclusion to Mimi, but kind of the other way around. In 2013, around the same time as Mimi, I found out that the average office worker leaves behind a digital footprint of five gigabytes worth of data, and I was fascinated by that. And like I think one of the things that I was interested in was the idea that this data exists, but how much can we actually understand humanity through these digital footprints?

And so in 2013, I was also going through probably one of the worst breakups in my life at that point in time, and understanding that there is so much data around about me out there, I realized that I wanted to see whether I could use this data to plot my emotional recovery over time. Sometimes when you are in that moment, it is really, really difficult to pull yourself out, so I was like, how about I look into the data points as a way to heal myself, as a way to really cut through it and to understand both with kindness that I will feel like garbage for a while, but there is a path that I can plot by looking at the data over time and so I started out with like very simple things that were actually almost manual data collection, like oh, when did I go to sleep, right? But then I started looking into things like what data can I get from my Facebook archive? Or what information can I find through my bank accounts? Like I did a whole chart that was called three months of retail therapy, looking into how I was recovering as seen through the stupid purchases I made that were all collected through my bank account. And I think in many ways for me it became a way of like being curious.

So there are these footprints all over the internet of my behavior. What can I actually meaningfully say about them and what does it not measure? What are the missing data sets here? What are the missing components here that would really be necessary for me to responsibly interpret this data? And so I, I basically started understanding the vastness of the data troves out there and the ways in which we could totally misinterpret this information if we don't have the context. If I didn't have the context of me going through a breakup. What would I say about the data of me buying a stupidly expensive fountain pen, you know, like to me it became a very interesting exercise in both looking at the amount of, the amount of data that exists about me, but then also understanding the nature of it. And that is where I think similarly to Mimi, I just became very curious about what is it that people optimize for when they start collecting this data? What is it that they actually want to do with this data and how does that skewed way of looking at humanity limit our view of it?

NK: Yeah, what I love about both of your stories is because both of you were the collectors, you were able to like see when the data didn't line up with your actual experience, right? And like it became very clear to you what was missing because you are like, I live this and it is not showing up, whereas like I think in other data sets, it is perhaps maybe not always as obvious because the data we are collecting isn't about our personal lives or experience. And so again I think the really important motif in both of your work is not just focusing on the data that is collected on humans, but the humans who are doing the collecting, right? And so and I think Lam, you brought up this missing data set that I think we are going to get to later in our conversation. But first I want to talk about just how people show up in data. So, Lam, you have this project about where you create quantified selfies, and I'm curious to hear about the various kinds of data that we leave behind with the internet and how you have come to see how they represent people, so could you tell us a little bit about it?

LTV: Sure, so there is two, I mean, like if you want to broadly look into the categories of data that exist are, there is one data set that kind of collects things that you do on the internet without even thinking about it much, right? That could be as simple as you liked a photo and that now is a data point somewhere in a large data set about you. And then over time, suddenly all of these companies from like the bank that is looking at your expenditures to the ways in which cookies are collecting your behavior when you are searching online, all of this is information that becomes the data set around things you did when you thought no one was watching to some degree, right? And then there is the opposite of that, which I think is like this interesting performative space on the internet where a lot of people are like making announcements about who they think they are, right? Who they want to be and I think that is a different kind of data set, that is the Instagram post, the Facebook status update, right? And so I think that is a really interesting quest to really understand how do people show up? Well, there is like information that is being collected to build a longitudinal behavioral pattern of you, right, and that is oftentimes optimized to then sell you things with ads. And then there is the stuff that you put out there about who you think you are and who you want others to think you are. And to me that is also fascinating understanding of like yourself in the digital age when you exist in semi-private and semi-public spaces at the same time.

NK: Yeah, I think that is actually a really important point to emphasize because I think so much of, so many conversations around data are about like the data that these huge companies and

social media companies are collecting on us, but you are right, like we are also creating data about ourselves and the way that we want to be read by others. And so –

MO: You know –

NK: Oh, go ahead, Mimi.

MO: Ooh, sorry, Natalie, I just was going to jump on top of what Lam was saying. Something I'm always saying is how data collection is actually a relationship, but it is a relationship that can be difficult to see and I think both of us know what it is like to do these projects where we are both the subject and the object of the data collection, where we are the ones who are collecting, but the ones who are collected, but in fact that relationship for many more people is just more about being the ones who are collected, as opposed to the ones who are able to do collecting, and then it does seem like these questions, some of the questions that you are getting at do arise from this looking at, okay, well, who is the collected versus the collecting in this particular moment or this context? Because if there is a difference between those, and all of a sudden, there are all these other questions that you can ask and so just throwing that, throwing that little thing onto the table too.

LTV: Well, you mentioning, I think the idea of like being both object of the collection and creator of that data, in a both semi-conscious way, right? Like I think that is like a really interesting emotional space that I'm curious about, right? Like the idea of like, how much do we know that when we like something it means something? There is so many unspoken emotional tolls that happen on social media nowadays, like from the fomo that you get all the way to like the ways in which exes are not supposed to behave a certain way online, like what is like the code of conduct that comes with that? Is an ex supposed to like a photo of yours? That is a really interesting question. Can you chalk it up to them being sort of socially inept and sort of not that great with technology? To what degree is this an offense or not? It is a really interesting thing around the idea of like the data that we produce as in, as some sort of indicator of what we think or feel, right? Because I think that is what everyone sees it as now.

And so when it comes to this like idea of being collected about, right, like a lot of it is this ominous machine that kind of just follows you around the internet, and then the other part of like what that social currency of a like of a data point that you produce is now. There is like all these terms that have come up from the social context of social media. Think about orbiting. I think *The New York Times* wrote a big article about it. I don't, I wouldn't pertain to say that *The New York Times* discovered the term, but like the idea of orbiting is that someone keeps showing up in like your social feeds, they used to date you, they no longer want to have anything to do with you, but somehow you can see their little icon pop up as they watch your stories on Instagram or Snapchat, right, like, they suddenly like something that you did or like a post that you are tagged in in a different, someone else's Instagram account, and like there is some interesting social currency that a data point now is, and I don't think we have a common playbook for how we evaluate that, right? We have a lot of conversations about the privacy implications of those people who are being monitored by others, but I don't know whether we had a lot of conversations around the idea of like how that has changed social dynamics, and I'm fascinated

by that. I have no answers. I have a lot of sadness and happiness with it, but also, I'd love to hear what you have to say.

MO: I agree with you. I think that there is this additional layer that I think about, which what you are saying is we think, which is actually something my partner really put me onto, because my partner was doing a lot of work around disinformation and misinformation and it was just this thing of how, you know, in real life, you have, when you are having a conversation with someone, you know, there is what you say verbally and then there is everything that is nonverbal, and there is so much information in the nonverbal, just as there is information in the verbal, and there is this like information in the combination of the two. And my partner would talk about people sharing things on social media and how that is not just an act that is like this explicit what is it that you are sharing? It's not just about the content, it is also about how you want to be perceived, what it means to be somebody who shares this particular thing. Just understanding that there are more, the different layers to this, where it is not just about the explicit and it is the most human thing to not just be about that.

LTV: Yeah, and I think what you were saying about like when we post something, we adhere to it and we say something about ourselves. I think one of the things that is also a bit interesting within the larger context of that is it is a demarcation. If I post something I'm reading, it is me drawing a line between what political camp I'm in and what political camp I'm not in, even if I post it with zero comments, right? Like our political news consumption is now, it is a very posturing community building act, right? Like when I share an article let's say, even from let's say *The Guardian* or like *BuzzFeed News*, where I work, it does not even matter whether this article solicited complicated feelings, whether I was looking at it in an exploratory manner, the fact that I'm posting it in a semi-public space, right, that then builds sort of like a little coalition that then other people can adhere to, and so if someone likes that post, that means you are my political group, you are my little village, you are part of this group now that we are all participating in and I think that is actually like a lot of the polarizing filter bubbles that people have been talking about just using the like button, the heart button on Instagram, or posting like if I post the picture of a queer friend who is transitioning, that is a clear political statement, by which I then demarcate myself from one group or another, right? So it is a really fascinating thing where everything, because of the nature of, the performative nature of the internet, where can we even just have exploratory fluid spaces?

NK: Yeah, that kind of brings me back to something I wanted to ask you, Mimi. You wrote this piece for the Walker Art Center about how the original conceptions of the internet subscribe to this universal us. And you kind of make the argument that the universal us ignores the complex histories and messiness that comes with being human and I think kind of what Lam and you are both saying here is that people use social media and the internet to kind of perform an identity of themselves, right? And so I think that is already like telling us that there is no universal us online, but I'd be curious if you could talk to us a little bit more about the article that you wrote and the argument that you are trying to make there?

MO: Yeah, I am just always very interested in this question of us and we and who really gets seen as the we of a society? And I think that we is just extremely interesting, because it becomes this kind of narrow proxy for highlighting and revealing a much larger kind of matrix of power.

So this ability to like say we and not have to consider who is included in it or by nature of what you are saying, excluded from that grouping can be just very revealing I would say. And so in that article, I am talking about this moment that the world is in, which is this moment of injuring coloniality, which has emitted so many different groups and types of humans together, but in ways that are not equal, and the internet was built to top that same world, and so the thing that I'm always sort of holding and working in technological spaces is the degree to which that reliance on that vision is acknowledged. And the internet at the time of its founding was such a good moment to look at because it simultaneously held these kinds of spaces where it was like, oh, there is this chance for kind of different pockets, for different ways of maybe being to sort of peek out, but of course, with anything that is new, it is always important to note what is kept the same and what is carried over and one thing that was carried over is that same like epistemological perspective, where the we is very clear and very, it is very clear who is held within it.

NK: Can you give us an example of an instance when someone might have referred to a group of people as that we in relation to data?

MO: Oh definitely, yes, let me ground this a little bit. I think in that article I talk about John Perry Barlow's "Declaration on Cyberspace," which was published in 1996. Not trying to hate, but there is, you know, that's where is he like, oh, yeah, cyberspace is going to be this fantastic land, like we are without our bodies, we are free, we can do what we want, look at us, you can't stop us, this is it. And of course you know the story, anybody who is online now, or most people who are online now really have a strong sense of no, no, we didn't leave our bodies behind. That, it all showed up, it is all here and in fact, our experiences of this space are determined by all of these other things that we carried with us. The internet is such an interesting space, because there is just so much, I'm endlessly fascinated, there is so much potential, there is so, so much potential in it, but also, even within these spaces that seem very novel, the same power structures actually are kind of carried over and reformatted to still fit. Unless you are actively trying to push against those.

NK: Absolutely. So I want to move the conversation now to what is missing, because I think what is clear throughout this whole conversation so far is there are narratives that get excluded, there is data that doesn't get collected purposely or not, and so there is kind of this almost like alternative track that we don't always see because of the cultural and economic context in which we lead our lives. So Mimi, I know you have this project on, where you have created a library of missing data sets. So can you tell us about the project and why you created a library specifically?

MO: Yes, I'd love to. So, library of missing data sets. This kind of came out of this very small observation for me, which was just that there would be these systems where lots of data are being collected, and then there would be something that was missing or a few things that to me were very obviously missing. And around the time when I started this project, it was, I want to say 2016 perhaps, and at the time what I was thinking about was police brutality. It's funny because it is, you know, it was 2016, and obviously it is still relevant now in 2021, but police brutality and people who had died through police brutality. People who have been killed by law enforcement agents. And at that time, there wasn't really a clear data set around it, and I found that so illuminating because there is just so, so much data when it comes to thinking about

policing and justice, the justice system I mean, and so many other things kind of tied to that. And yet there was this thing and it didn't exist, there was nothing there, and I ended up just being really, really fascinated by that and I should say, by the way, that this is no longer a missing data set. There are groups working really hard, who have been working hard to try to collect and make sense of that, but the fact that it has taken a lot of effort and from very many groups to do this when to be honest it is something that could be very easily collected by law enforcement agents, just by the police, but the fact that it has taken that much kind of speaks to it.

And so I started this whole project, which is on the one hand kind of art installation and that's the library of missing data sets and then also kind of research project, and all of it is about all of these places where data do not live. And in it, I'm thinking about how, you know, a lot of my practice looks at absence and it looks at removal and silence and that the, I always think the easiest way to make sense of the system is looking at what seems to not be included within it, and noting the way that these siloed things are not distributed evenly. And so in the library of missing data sets, the actual installation is this filing cabinet and inside of it are all of these different data sets, things that are not collected, things that the larger public does not have access to, and there are all these different reasons for it.

A lot of what I try to do in that piece is really make clear through the organization of all of these different filing folders, which as I said, they all have the title of a missing data set, but then there is nothing inside of them, so you can kind of pick it up and look, but you will see nothing, because the point is that they are missing. A lot of what I'm trying to do is get at this pattern of absence, is to not fetishize the things that are missing to say they need to be filled, but to really think, okay, well, what is the pattern behind it? Why is it that there are so reliably things that cannot be collected, that cannot be made into data? Or that folks are insistently saying that they will not let be made into data. What are these patterns and what do they speak to? And I think that becomes a kind of way to hold this, this thing that I think we've been talking about a lot in this conversation. This weight, like you said, Natalie, of these different alternatives, different modes of being or living or different modes of whether, how you are included or excluded, it becomes, that piece becomes a way to hold a lot of those different ones.

NK: Yeah, you know, one of my follow-up questions to you was going to be, are these data sets that you think should be collected? And I feel like, kind of through your explanation just now, I almost feel like that is not quite the right question, although, I am still curious. It is like regardless what you are pointing out are patterns of things not being collected, right?

MO: Yes, exactly, to, I think to want to collect everything, and that does happen. Sometimes this piece gets shown and that has been a reaction where folks have reached out and said, yeah, just great, tell us all the things, let's do it, let's fill it, let's go. Like, no, no, no, no, that's not the point. What I'm trying to do is pull us back into a kind of structural understanding, which is really, really hard to do, but really that is what I am doing is not saying everything needs to be collected, but what I'm saying is that there are clear reasons why these things are not and each one of them, there is something, there are these patterns to it. And, you know, some of the those have to do with just that some things do not fit our systems of collection, you know, some things refuse quantification, they refuse being made into metrics, or there is this imbalance between who has the incentive and who has the ability, so that is like the case where I was talking about,

like police brutality and people killed by the police. The group that has the incentive to have that is the group that doesn't have easy access to actually have, to seeing that data set.

And then I think another, another reason is this one that is a little tricky I think to hold sometimes, but it is just that sometimes there is just an advantage to not having the data set be present, and the reason why I say that's tricky is because anytime something is missing, it actually is an advantage to someone, so a really good example of this is municipal ID cards. That is one of my favorite examples to use is that there are some around 2016 this really came into a kind of public news space, but there are some counties that municipal ID cards precisely to both protect undocumented people, but also, you know, to give this kind of municipal identification for people who would want it in the city. And so some cities do this. When they collect the information for these municipal ID cards, they will keep the information and then some cities when they collect it, to collect it so they can give it to you, they don't actually keep any of the information, and so in that moment, when they are not keeping the information, they are doing that as a strategic gesture of removal, where they now know that if the federal government should try to come to them and say, give us the, give us the database that says who took these cards, if you have that database, you could run that against some other databases and really quickly figure out who is undocumented, and in fact that did, that was a process that almost happened here in New York.

Or happened to some degree, and so there are loads of, there are some really interesting examples of different, like indigenous groups, there are things even in my own culture, I'm an Ebo Nigerian, where you can kind of look at it this way, where something is missing because it is protected and a group kind of realizes how it can be used them against them by the sort of dominant group, and so they say, no, we're going to withhold this.

NK: Yeah, I think that's such an important point and I think, I hope I'm remembering this correctly. I think in a presentation that Ruha Benjamin gave at Data and Society, she was kind of referencing the work of our data bodies and how they had documented a lot of their like organizing strategies for pushing back against the extraction of their data, but they also were like, we're not going to share all of these strategies because we don't want to reveal this because then the strategies won't be effective anymore, right? So Lam, I really want to give you a chance to respond, especially because I know that as a journalist, you are someone that is often doing investigations to create those missing data sets, right? Especially when it comes to like inequalities in social systems. So, I'd love to hear what you have to say.

LTV: I mean, I think I would like to complement some of the things that Mimi said. I think in addition to certain things just not being measured, right, like is that some of society's most vulnerable people have probably the most punitive data being collected around them, right? Like there is the idea of an eviction for example. I've been doing a lot of reporting around eviction records, and like understanding that that data is being used in punitive measures on some of society's most vulnerable like communities is something that is really problematic and there is so much more data about human beings who have been evicted in that realm that is not counterbalanced by things like, I don't know, we don't measure how often, how many jobs this one woman may have been like applying for, right? Like all of these other things that may counter that is not being collected.

Let's take a very simple story of a single mom, who is a woman of color, oftentimes black, this is the general demographic of what you see a lot in eviction court, right? Like, let's do a simple data profile of her. I bet you I can find a lot more data sets that document something bad that she did, then I could find data sets around the efforts she has made to take care of her children, to take care of rent, to take care of her car bills, of her insurance bills, and so on. There is all of this other countering like human measures that she must have taken, and I have talked to many women like that in my job, to rectify her own situation and to save herself and her children. Not only do we miss certain data sets, the way in which it is being collected oftentimes as data is very lopsided and places a much larger and disproportionate weight on "negative data sets," right? Any brush with like law enforcement, that is a data point. Any brush with like child services, that is a data point. Any application for like Unemployment benefits, that is a data point. Any eviction, that is a data point.

Well, what about all of these other ways in which this person has tended to the health of their family and the health of their community? We don't see that. And that to me is always something I try to counteract. I cannot prove to you that of the 400 people who are being evicted by this one corporate landlord, all of them are doing well. But I can show you the actual like goings-on of this one apartment block and like kind of counter that missing data by collecting data around what efforts have been put into the community to help each other out. What kinds of costs have these women incurred by fixing things that their landlord wouldn't fix, right? Like these are all different types of, hopefully balancing data sets that could then also explain some of the larger trends and anomalies that we can see in the negative data sets that exist about them.

One of the things that I find particularly troublesome is that not only do we have missing data sets and data that shouldn't exist, sometimes we have downright dirty data sets, and like what I mean by that is like it comes down to there is a great paper that Rashida Richardson did about police practices. And one of the ways in which she like looks into data and how flawed it is that police records and criminal, like criminal records, actually have bias baked into them, and then suddenly by looking at them through statistics, we have this idea, oh, this is a clean data set, it is completely unbiased, there is nothing in there. But in the meantime, we have now seen enough occurrences in the media around the country to understand that there is bias in who gets policed more than who doesn't, right? And so understanding that there is not just missing data, but that the data itself, the idea that something that was collected with technology is infallible, that to me is just as problematic as like the missing data sets.

All of this is to say I think this goes at a concept that Meredith Broussard kind of beautifully point as techno-chauvinism. It is the idea that just because it is data and it uses scientific processes doesn't mean that it is infallible. And techno-chauvinism is this like belief that it is science, it is data, it is technology, oh, that means that it is infallible, that means that I have to just blindly believe in that. I wish people would just kind of think about this more intently because at some point someone came there who did not look like us probably, who did not lead the lives that Mimi and I did, and like decided in a stroke of like a line of code what was supposed to be like what is considered moral and not, what is considered a risk or not, or what is considered someone who needs to be persecuted or not, based on shitty data.

NK: So I think this brings me to the next question that I would like to ask, and so Mimi, you have kind of coined this term algorithmic violence, and I think it kind of gets to a lot of what Lam is talking about when she is talking about social systems and government systems that end up making decisions about people's lives, right, and that turns into this very individual experience. So could you define algorithmic violence for us and just tell us a little bit about the term and maybe give us an example of it?

MO: Sure, yeah, I think about algorithmic violence as the kind of violence that a social algorithm or automated decision making system inflicts by stopping people from being able to meet their basic needs, and I don't even need to provide an example because Lam just did, so many fantastic ones right there, particularly when talking about eviction, talking about just the justice system in general. I love also citing Meredith Broussard's kind of techno-chauvinism is such, that is really kind of at the heart and I will say, algorithmic violence, something that I think is just important about it is that, it is understanding this as something that lies on top of other forms of structural violence. So this is on top of racism and racial capitalism. This is on top of the same like, like what we have inherited from colonialism. The way in which these automated decision-making systems, which are built on the data that comes from those same systems, now inflicts an additional layer of preventing people from being able to do what it is that they need to do to just be able to live their lives. It is no surprise that there are so, so many of the people who have been, who really held down this field in talking, people who Lam was talking about. It is not surprise that it is so many Black women who have held down this field, folks like Safiya Noble and Meredith Broussard and Ruha Benjamin and Simone Brown and there is, Latanya Sweeney and it, those lists, you mentioned Rashida Richardson, the list just goes on and on, I think it makes sense, because it is folks who are at this particular intersection of seeing the ways in which lots of different, lots of different structural violence can be inflicted and then different systems can fail you if you fall into those systems. Or fall into those cracks.

NK: Yeah, Lam, I want to give you a chance to say anything else you wanted to.

LTV: I think to even zoom out further and I think Mimi and I have like, oh, I miss, like we should just have a drink after this and like hang out.

MO: Yes, let's go.

LTV: On a rooftop, somewhere six feet apart, but like with very rigorous conversations.

MO: Yes, masks.

LTV: To everyone who is listening, but I think one of the things that I talk a lot to my friend about, is the idea of like documentation as a colonizer's force, right, like the idea of bureaucracy, the ways in which something that is valued is documented. The same way in which like for example my grandmother and my grandfather met on a colonial plantation. They were like workers on that colonial plantation. What does this mean when their history is not documented, but the history of this other person is documented?

Even when you look into Simone Brown, you mentioned, she wrote this like incredible book, *Dark Matters*, which at some point she talks about what is being collected, what kind of data was being collected about Black folks during the time, like very early on was their assets as slaves and as property, right? They were not considered human beings. It is the same way the Census did not count Native Americans in 1790, but then divided race into categories of white folks, slaves and other free folks, like there is a very inherent almost colonizes gaze to the act of documentation. And I think when you look at different cultures and histories, think about Native American cultures that actually pass information down in an oral and verbal fashion, right?

There are so many non-written documentations that actually are much more a participatory way of looking at information over time. If you think of history as documented through data versus, and documents, versus history documented through oral tradition, right, there is a very participatory, communal and community-based understanding of that information versus he has a clean slate of spreadsheet that details every single person's like function at that point in time. I don't know if like one is particularly superior to the other, but in an occidental Westernized look at society we oftentimes prioritize the written documentation, which includes data, right, above any other ways of understanding and documenting history. And to me that in itself is already like an interesting question. Just because it wasn't written down, just because it wasn't written down into a spreadsheet, doesn't mean that it matters less.

NK: I'm very interested if you could like talk a little bit more about what you mean when you say participatory, what does that look like?

LTV: Right, like I think there is an element of like understanding like if a, let's take a very simple thing, cooking, right, like this goes into my heart as well, the idea of like the way like my grandmother made pho is the way that my mother made pho, is the way that I learned how to make pho. These are all three different ways, like this is something that was passed down from generation to generation, but it was always participatory and it was always sort of catering to the circumstances of that moment. Understanding how to read a fricking, like the people who are in that room and then figuring out how to understand that information in a very contextualized, in a very community-based way.

Data and documentation is oftentimes written down in the most sterile of ways, right? That doesn't make room for like any sort of different way of contextualizing that. It is not necessarily like something that has served a purpose, and it oftentimes one of the things that I'm interested in is the idea of data stewardship, so when it is written down, similar to what Mimi was kind of like collecting this like data set of like cat callers, what the intention of the data set was is not what it ends up being used for maybe, right? The same way for example the intention of like one of the things that I'm interested in in policing is the idea of like someone documented their life in a certain way and used the hashtag to mourn the death of a community member, that is a very contextualized understanding of like expressing grief in a community. When you strip out the communal aspect of that, this little hashtag and this is something that the NYPD does, becomes a way for police officers to listen on potential gang members, because the person who died may have been affiliated with a gang. Data stewardship. Suddenly you take something that is a right, that is a ritual, that is a community-building aspect. You document it in like a very sterile format

and put it into like a Facebook archive or whatever it is, police officer picks that up and puts that kid into a gang database. What does this say about the fallibility of data stewardship right there?

MO: I think, Lam, I just am vibe-ing off of what you are saying so hard.

LTV: I'm glad you are.

MO: Yes, you know I am. And I think just one thing that I would like to point to, just to your point about, I think you just said it so beautifully. This kind of, I think this is a form of violence, this sense, this view that says only that which is written, only that which is documented, is valued. That that is kind of the central, it illuminates so much. And there is something in, just what you were talking about more oral kind of cultures, you were talking about a lot of indigenous groups. And I know for me, as I said, I'm an Ebo person, I'm from the Ebo group, which is in Nigeria. And a lot of what, you know, there is so much, because it is a culture like proverbs really matter a lot, the spoken has a very, has a huge significance.

Something that is very interesting about things that are passed down by being spoken is that the person knowing something is never removed from what is known. So, that is never, this thing that we talk about where there is the artifact in data, and then the context that it is produced in, with something that is spoken, the context and the thing that is produced, the thing that is said, they are always tied together, you have to, you know, you are with the person. And then they say it. And so there is no confusion about that. There is no separation. And I just, I agree with Lam. It is not a question of what is better, it's a question of what does it mean that this focus on the documentation, on the artifact, is overrepresented and is now made to be a value that the whole world in the same coloniality sense is now pulled into and so that creates these rifts and I think there is a violence to that that Lam is describing really nicely.

NK: I like that you both pushed this out of this binary participatory versus written or documented, because I think that is where conversations often stray and instead focus this on how, you know, documentation in and of itself can strip data of its context, you know, remove it further from the people it is about and how that can be a violence in and of itself. So in this series, we have been trying to end each episode by looking to the future, and two themes that have run throughout this conversation, for me at least, are morality and power. Right? So a lot of the data collection practices that we have talked about today are not legal and thus conversations about the harms that they perpetuate often fall into questions of morality. So I'd love for each of you to tell me how you think about morality and power in the context of data? And I'm hoping this will provide us with some frameworks for negotiating our future relationship to data. So Lam, why don't you start?

LTV: I think it is not necessarily the data itself, it is what we end up wanting to do with it and how we evaluate what that goal is. To some degree I think the overall argument is that the act of data collection, the tools that we use to do it with, they are neutral, right? There is nothing moral about writing a line of code that then picks up a piece of information and puts it in a spreadsheet. What do you do with this and how do you understand limitations of this and honestly, if you really want to use data, how do you, what are you optimizing for? What is, what to you constitutes the health of the society versus what makes you the most money, Facebook, right,

like it is the, it is this really interesting question of having just gone with, let's collect all the data that we can and just kind of optimize it for like capitalistic gains to a point now where we have to really consider why, what are we optimizing for and how do we even measure that if we want to use data, right? If I want to use data to allocate school lunches to neighborhoods that need it more, that is not an amoral choice, right? Like that is hopefully something that we can do with the information that we gather, right, like through the Census. To me, it is always sort of an index of like metrics that I like to use to get to a point where hopefully society is healthier. If that is something that we want to optimize for I am all for like finding different ways of using data in ways that also make room for errors and us changing that course of action. But I think we don't even ask ourselves the questions, what is it for? What is our optimal outcome? How do we nurture with data versus punish with data?

NK: Yeah, I think it is interesting that you use the word optimize also, because I feel like that almost is a bit of a techy word, right, that you are trying to set and view these different questions into like what it means to optimize. Mimi, how about for you, how do you think about morality and power in data systems?

MO: Well, I think that there are some baseline moral questions. There are some questions that have a normative answer, I suppose. Or are normatively constructed. So, do we think that people shouldn't die of hunger? Do we think that we shouldn't, people shouldn't face violence for things they cannot control? Things like that. That is a moral question, to which I would say, yes, yes, very much so. You know, do we, there are a few of, these are some big questions that exist, and there is a moral question there, which is that basically do we believe that people should be able to live these lives that are full?

Now, there is a lot more, to just really oversimplify, I tend to think that when it comes to what gets in the way of this, that is all a question of power. That is what that is. Really, a lot of the things that many of this in this space are investigating are really questions of power, and so something I do think I see a lot is that those get confused so that people will say, I will try to say this in like a way that is diplomatic, but people will, these get confused. People will take something that really is about power and they will reframe it to be about morality. And in doing that, they allow themselves a kind of choice or eliminate a kind of responsibility, so, you know, a company like Uber can talk about wanting to make things morally better for the people who drive for the organization, but doesn't have to think about making them employees, or paying them minimum wage, you know, it doesn't have to think about some of these other things in which they would lose something, and so I guess that is why.

The thing about power is that power, for a group to gain power often another group has to give up a little bit of something. I know people don't want to give things up. Morality I think often becomes more comfortable, for people to use that framing, because it doesn't necessarily involve giving something up, and instead, it frames someone who has power as very good, whatever that means. It frames them in this positive light as opposed to being like, okay, there is something that needs to be taken away. So for me, I tend to think about things with the lens, with a sort of power-based analysis. That is often how I approach this. This kind of work that I do. But that is not to say that I don't think that there are any moral questions here at all. It's just that the moral questions are at a far lower level.

NK: Yeah, I think you hit a really key point there in the way that we kind of grapple with power amongst each other, in organizations, wherever it may be, corporations, is like someone has to give up power in order for other people to gain power and to step into that, and I think that is a really nice place to leave off.

MO: Well, it's a take, I should say, let me just get it right. I mean, that's why people say, you got to give it up or it has got to be taken.

NK: Right.

MO: It is better given up, but –

NK: We would hope they give it up, but. Alright, well, thank you both so much for chatting with me, this is a great conversation. Is there anything else that you want to add?

MO: No, this is great. Thanks, y'all, and Lam, thank you.

NK: And that's our show! A huge thank you to Mimi Onuoha and Lam Thuy Vo for sharing their thoughts about data and social life.

Please rate and review the show in Apple Podcasts and subscribe to the show there, or in your favorite podcast app.

You can follow us on Twitter: @DataSociety and @PublicBooks. If you have thoughts about this podcast, feel free to share them on Twitter and tag us. And if you like the show, please tell a friend, or even a few friends!

Next time on *Becoming Data*, I talk with Shaka McGlotten, a Data & Society fellow whose work focuses on the intersection of black study, queer theory, media and art, and Chris Ramsaroop, an activist and co-founder of Justicia for Migrant Workers. Together we discuss how data has—both now and historically—been used to categorize and exploit workers. So I hope you'll join us for episode two about data and labor.

This podcast is a production of Public Books, in partnership with the Columbia University Library's Digital Scholarship Division. Thank you to Michelle Wilson at the library, for partnering with us on this project. This episode was produced and edited by Annie Galvin, with editorial input from Kelley Deane McKinney and Mona Sloane. Our theme music was composed by Jack Hamilton, and our logo was designed by Yichi Liu. Special thanks to Data & Society Director of Research Sareeta Amrute and Director of Creative Strategy Sam Hinds, and to the editorial staff of Public Books for their support for this project. Thank you for listening, and I hope to see you next time.