THE STATE OF ARIZONA

INDEPENDENT REDISTRICTING COMMISSION

REPORTER'S TRANSCRIPT OF VIDEOCONFERENCE PUBLIC MEETING

Via GoogleMeets

July 13, 2021

8:02 a.m.

Miller Certified Reporting, LLC
PO Box 513, Litchfield Park, AZ 85340
(P) 623-975-7472 (F) 623-975-7462
www.MillerCertifiedReporting.com

Reported By:
Angela Furniss Miller, RPR
Certified Reporter (AZ 50127)

I N D E X

PUBLIC MEETING, BEFORE THE INDEPENDENT REDISTRICTING COMMISSION, convened at 8:02 a.m. on July 13, 2021, via GoogleMeets, Arizona, in the presence of the following Commissioners:

Ms. Erika Neuberg, Chairperson
Mr. Derrick Watchman, Vice Chairman
Mr. David Mehle
Ms. Shereen Lerner
Mr. Douglas York

OTHERS PRESENT:

Mr. Brian Schmitt, Executive Director
Ms. Valerie Neumann, Executive Assistant
Ms. Michele Crank, Public Information Officer
Mr. Roy Herrera, Ballard Spahr
Ms. Jillian Andrews, Ballard Spahr
Mr. Eric Spencer, Snell & Wilmer
Mr. Brett Johnson, Snell & Wilmer
Mr. Mark Flahan, Timmons Group
Mr. Randy Trott, Timmons Group
Mr. Mike Wiley, Timmons Group
Mr. John Strout, Timmons Group
Mr. Douglas Johnson, National Demographics Corp.
Ms. Ivy Beller Sakansky, National Demographics Corp.
Mr. Thomas Bryan, Bryan GeoDemographics
Dr. Moon Duchin, the Redistricting Lab, LLC

P R O C E E D I N G

CHAIRPERSON NEUBERG: Welcome Commissioners, staff, partners, and the public. We will dive right into our agenda.

I(A), call for quorum.

It is 8:02 a.m. on Tuesday, July 13th, 2021. I call this meeting of the Independent Redistricting Commission to order.

For the record, the executive assistant Valerie Neumann will be taking roll; when your name is called please indicate you are present. If you are unable to respond verbally, we ask that you please type your name.

Val.

MS. NEUMANN: Vice Chair Watchman.

VICE CHAIR WATCHMAN: Present.

MS. NEUMANN: Commissioners Lerner.

COMMISSIONER LERNER: Present.

MS. NEUMANN: Commissioner Mehl.

COMMISSIONER MEHL: Present.

MS. NEUMANN: Commissioner York.

I think he'll be joining us when he can, when he gets connected.

Chairperson Neuberg.

CHAIRPERSON NEUBERG: Present.

MS. NEUMANN: And for the record, we have Executive Director Brian Schmitt; public information officer Michele Crank; our legal team, we've got Brett Johnson and Eric Spencer from Snell & Wilmer, Roy Herrera and Jillian Andrews from Ballard Spahr. We've got our mapping consultants; we've got Mark Flahan, John Stroud, Doug Johnson, Ivy Beller Sakansky, and I think Mike Williams [sic] from Timmons will be joining us as well. We also have our special guest, Thomas Bryan who is founder and CEO of Bryan GeoDemographics; and we are waiting on Moon Duchin, the CEO of the Redistricting Lab and associate professor of Tufts University; and Angela Miller, our transcriptionist.

CHAIRPERSON NEUBERG: Okay. Thank you, Val.

And thank you for everybody joining us.

Please note for the minutes that a quorum is present.

Agenda Item I(B), call for notice.

Val, was the notice and agenda for the Commission meeting properly posted 48 hours in advance of today's meeting?

MS. NEUMANN: Yes, it was.

CHAIRPERSON NEUBERG: Thank you.

Agenda Item II, approval of minutes from June 29th, 2021.

(Whereupon Commissioner York joins the meeting.)

CHAIRPERSON NEUBERG: We have (A), the general session, and we have (B), two different executive session minutes from Agenda Item IX(B) and (C).

Is there any discussion on the minutes?

If not, I'll entertain a motion to approve the general session and two executive session minutes.

COMMISSIONER MEHL: I so move; Commissioner Mehl.

CHAIRPERSON NEUBERG: Do I have a second?

COMMISSIONER LERNER: Commissioner Lerner seconds.

CHAIRPERSON NEUBERG: Any further discussion?

Okay. Vice Chair Watchman.

VICE CHAIR WATCHMAN: Aye.

CHAIRPERSON NEUBERG: Commissioner Mehl.

COMMISSIONER MEHL: Aye.

CHAIRPERSON NEUBERG: Commissioner Lerner.

COMMISSIONER LERNER: Aye.

CHAIRPERSON NEUBERG: Commissioner York -- and if we can't hear --

COMMISSIONER YORK: Aye.

CHAIRPERSON NEUBERG: We heard you. If in the future if your sound is -- is, you know, iffy, feel free to do, you know, thumbs up or thumbs down; but Commissioner York said aye.

Commissioner Neuberg is an aye.

With that, the minutes are approved 5-0. Thank you

very much, Val, as usual for your excellent minutes.

We'll move to Agenda Item No. III, opportunity for public comment.

Public comment will open for a minimum of 30 minutes and remain open until the adjournment of the meeting. Comments will only be accepted electronically in the writing on the link provided in the notice and agenda for this public meeting and will limited to 3,000 characters.

Please note members of the Commission may not discuss items that are not specifically identified on the agenda. Therefore, pursuant to A.R.S. 38-431.01(H), action taken as a result of public comment will be limited to directing staff to study the matter, responding to any criticism, or scheduling the matter for further consideration and decision at a later date.

With that, we'll move to Agenda Item No. IV, discussion on public comments received prior to today's meeting.

I open it up to my colleagues.

COMMISSIONER MEHL: A lot of the public comments deal with the procedures and safety and other issues on our public meetings coming up and -- and I think we're very much attuned to those issues. I think as a Commission we're working hard to prepare for those meetings properly.

So I appreciate the public's concerns there, but I think we're well alerted to those issues.

CHAIRPERSON NEUBERG: I -- I agree.

I mean, I really -- there were two things that stuck out to me. One was just, you know, helpful and concerned feedback about ground rules for public hearing, as it relates to safety, you know, health and all of that; and, yes, I agree with Commissioner Mehl that our staff is hearing all of the feedback and taking everything into consideration and -- and, you know, working with DPS to the best of our ability to try to, you know, address all concerns.

The only other thing that I would like to add is I believe that there was some concern from the public about possible communication going on with -- with the mapping team outside of our public meetings and who is having those conversations.

Just, you know, for the sake of all of us being on the same page, there's minimal interaction with the Commissioners and the mapping team because of open meeting laws. I mean, you know, there -- we are all -- we all have the ability and access to reach out to our mapping team if we have questions, we want to learn something, there's something that we don't understand; but in terms of executing business, that all is done in public session and

so there's limited work -- you know, we don't even want to run the risk of serial communications or things like that.

The appropriate communication that is going on is with our staff and the mapping team; there was some concern about that.

The communication that is happening amongst staff has to do with executing the vision and the direction of the Commissioners. There's a tremendous amount of logistical work, practical work that needs to be done. That's why our staff is hired.

And so I just want to reassure the public that -- that, you know, the staff is not working with the mapping team in any way beyond what is absolutely necessary and appropriate.

And that's really what I had on my -- in mind regarding the public comments.

Anything else?

Okay. We'll move to Agenda Item No. V.

We have some guests today to give us presentations on the pros and cons of differential privacy. V(A), we have Thomas Bryan, the founder and CEO of Bryan GeoDemographics. For those who looked at the agenda, we had an attachment that provided, you know, a huge source of information.

And, with that, I would like to turn it over to Thomas, and thank you for joining us.

Thomas, we're not -- we're not able to hear you.

MR. BRYAN: Can you hear me now?

CHAIRPERSON NEUBERG: Yes.

MR. BRYAN: All right. Great. Thank you. Sorry about that.

It's a pleasure to be here, and it's a real pleasure to be sharing the stage here with my other expert to talk about differential privacy.

Let me see if I can find out how I can share my screen.

I'm not familiar with this platform, so let me see.

Present -- can everyone see my screen? It should say "Differential Privacy and Its Impacts."

I can't see any of you right now so I don't have --

CHAIRPERSON NEUBERG: Yes.

MR. BRYAN: Okay. All right, everyone can see? Terrific.

So let's start. This is the agenda that I have for this morning: I'm just going to give you a brief introduction, this is who I am; I'm going to talk a little bit about the 2020 census, the products and the release schedule, basically what's going on right now with the census. I'm going to share with you the Census Bureau's perspective, basically how they came to use differential privacy; how we've gotten where we are right now. And then

I'm going to go back and I'm going to talk a little bit about some of the information that is what they call the demonstration products. The Bureau, throughout this process, has released a couple of data sets that have illustrated examples of the -- the consequences, kind of the fallout of what the data would look like if they went through with this differential privacy process. And then we're going to talk a little bit about findings from the -- what they call the final production settings, whatever the final settings are, for differential privacy that they just set on June 9th.

Does that sound like that works for everyone? Is that an agenda? Is there anything else that anybody would like to hear me cover in the presentation this morning?

CHAIRPERSON NEUBERG: That sounds great.

MR. BRYAN: Okay. Thank you very much.

Okay. So this is who I am.

I used to work at the Census Bureau. I've been working doing political redistricting cases, Voting Rights Act cases, discrimination cases. I have used census data, especially small geography census data, professionally for decades now, including numerous Supreme Court cases.

I'm affiliated with all of the major demographic organizations; I'm a board member of IAAB, and a member -- or a senior member of every other group. So I think, you

know, my experience, my affiliations here, I think I can say that I'm -- I'm speaking for the most part on behalf of experts, professionals, and practitioners in these organizations about the kind of wide-held beliefs and thoughts and concerns.

That what I'm going to share with you are my thoughts and beliefs, but I -- I can say with certainty that they reflect the beliefs and the thoughts and concerns of many of my colleagues.

So let's start just talking a little bit about the redistricting data program. The Bureau's released this slide that says, you know, they want to deliver high-quality data to enable the process of redistricting.

And so, you know, the most important thing that we as practitioners who get out of the census are what are called block-level data, and we'll -- we'll talk a little bit about census geography here in a few minutes. But essentially there's a hierarchy of data that the Bureau releases from the nation, the states, counties, all the way down to lowest level of geographies which are called census blocks, and those census blocks are the pieces of geography that we use to do the redistricting process and that we use to litigate Section 2 VRA cases.

So that for our purpose, the accuracy, the quality of block-level data are the most important things to us

being able to not just do a good job, but to ensure we are complying with constitutional and statutory law.

So the schedule for these data are as follows: We've already had the April data have been released, and then there's this other series of data starting with what's the public law data or PL data that are going to start getting released shortly.

The PL data are going to get released starting in August in kind of a raw -- in kind of a raw form; and the final PL data are going to get released in September.

What's important to know about these PL data is these are the backbones of the census data release, they have the information on and they will be the first data we see on race, on ethnicity, housing unit occupancy status, on recorder's data; and, importantly, because the redistricting process is based on census blocks, this will be the data set that have data down to these lowest levels, these most granular geographic details that we are able then to take and use and craft legislative districts, congressional districts, tenant districts, and everything else that comes after that: school districts, water districts, everything else.

But it's important to note that those districts have to be created from census blocks: census block groups, census tracts. Certainly, pieces of geography bigger than

that just are not granular enough for us to be able to design and draw districts where we can say with certainty "Ah, this district is a majority-minority district, or we know with some certainty it's a minority district."

So that the important thing about this data set for us as practitioners is that they give us accurate data at the census block level that we can use not just to do our job but, again, to comply with constitutional and statutory law.

So this PL data, you know, even though the data are delayed significantly from when they were supposed to be released, they are scheduled, to the best of my knowledge, now they're going to be coming out in the middle of August. I think the one thing I can share with you just kind of outside of differential privacy is that, you know, these data are going to get released in August in what I'm going to call a raw format. They're just going to be these big, cloogy chunks of data that aren't going to be formatted. They're going to be really ugly.

So for the people who are going to be working on the redistricting for the state of Arizona, I really encourage you to make sure you have expertise, you have the code, you have everything in place necessary to process and start analyzing those data as soon as they hit the street in August. They'll be available and ready to use. They just

won't be in an elegant, easy-to-use format that they're going to be presenting at the end of September.

So be ready by August to have that so you can get a head start with the limited time we have to do the redistricting work then. Don't -- there's no need to wait until the end of September to get going with the redistricting work.

All right. So let's talk a little bit about the Census Bureau and how we got into this issue of differential privacy.

So the Census Bureau's mission first and foremost always has been to be the leading nations -- nation's leading provider of quality data about people and economy.

You know, they -- they also have to, as part of their mission, protect confidentiality. It's a critical part; it's a backbone of the census that when people share their information, that the Bureau is going to protect it and make sure that people's personal, private information is not going to get out into the public.

So the background of the census, you know, we have this legal requirement that it be conducted; but there's a tension, right, that the more accurate the census data are, the easier it is that someone could go look at a piece of geography and say, "Oh, well, I know who somebody is."

So if you have a block, for example, that has all

white nonHispanic in it and then there's one Hispanic family in it, and the data came out and said "Well, here's the information about a Hispanic family," people are going to know who that is. So the Bureau is obliged to take steps to protect that family or protect those individuals and make sure that any given block that the -- the identity of those people is not revealed.

But on the other hand, you know, the -- the more that you privatize data, the more that you swap data, change data, conceal data, take steps to privatize it, the less usable the data are.

So if you take all of the data that you have and you apply steps, statistical steps to change the data to protect someone, the less and less usable it is for a practitioner such as myself to be able then to go in and say, "Ah, I made a district, and I'm really certain that it complies with the law, or I'm really certain that I have protected a population that I'm legally obliged to protect or that I'm legally obliged to represent."

So there's really this -- this tension between these two goals, and there's not really a right answer. It's just a middle ground where the Bureau is obliged to both report accurate data, but they're also obliged to protect data.

So no one will dispute that, you know, the Census

Bureau has been at the forefront of designing and implementing, you know, statistical methods for demography; that's their job, and they've had lots of people for lots of years engaged in this kind of work.

The types of things that the Bureau has done historically to protect data have included things such as just suppressing data; just not reporting it in some cases; top coding or bottom coding certain records; sometimes they'll round data; sometimes they'll swap data back and forth, you know, take records from one piece of geography and move it to another and back and forth; or they can inject noise, just put data that are wrong in it in some degree to ensure that the respondent's privacy is being protected.

And so they've used different techniques over time to help protect the data that they are entrusted with.

So over the course of the last decade -- and the Bureau has done this, you know, between every decennial census -- what they did is they went back and looked at the last census and they said, you know, how many records did we collect? And is it possible that using the data that we publicly reported, is it possible with the techniques that we have in place, that someone could have gone in and done a reconstruction? That is, could they just through computing power figure out who people were in different pieces of

geography?

So they put protections into place, but they just wanted to check and make sure that they would work.

So the Bureau is -- is widely publicized that they've went through this great big project where they went and they threw a bunch of computing resources and a bunch of personal resources trying to reconstruct data from the 2010 census. And what they have said and what they have publicly reported is that they were able to, even with the protections that were in place, they were able to replicate, to reconstruct a bunch of the data that was reported from the 2010 census.

So there's two things that the user community would say about that. We -- we recognize that they did it; we believe them that they did it, but there's two issues.

First of all, if somebody did have computing power and did have the wherewithal to be able to go and reconstruct all those records, anybody else at the Census Bureau does not actually have the census data to be able to check and see whether they were reconstructing the right records or not.

The Bureau obviously has that information, and they can check and see if their reconstruction exercise worked or if it didn't work. So somebody on the outside wouldn't be able to do that. They might get some right; they might get

a lot of them right, but they wouldn't know which ones they got right or which ones that they didn't.

I think the second important thing to know about the exercise is that there has not been any cases where anyone actually did go and crack the code at the Bureau and figure out using computational power who anybody's identity was out of the census data.

You know, we've heard rumors of a couple records or there may have been a couple incidents where somebody figured out somebody somewhere in the data set. But there have been no big, widespread cases where you hear about, you know, somebody hacked the Bureau's data set or somebody did a reconstruction exercise and figured out everybody's census records. They just they didn't do it.

And, you know, I think for, again, speaking on behalf of many practitioners is that, you know, today the world that we live in, you don't have to do that. If somebody wants to know who you are, they want to know how old you are, how much money you make, what your family looks like, all anybody has to do is they don't have to break the law, they can just go to one of the consumer household databases, you know, you can go to Experian for example and just give them a little bit of money, and they can tell you all the information about every individual who lives in any piece of geography in the entire United States, and -- and

they can do that with an incredible degree of precision and accuracy.

So for people that really want to know detailed information about you, the -- the motivation to go and put a whole bunch of work and computing power into cracking the code of the census as opposed to just going to a consumer household database and just giving them a little bit of money to know everything about you is not there.

So we think -- the user community mostly thinks that the reason that people have not gone to this effort of cracking the census -- or trying to crack the census is because they just don't need to.

So setting that aside, the Bureau said, "Look, we're concerned that somebody could get into the census and figure out who individual respondents are." Whether people are motivated to do that or not, their responsibility is to protect their data; so, that's fine.

So what they did is they looked at different ways given all the competing power that are out there right now, they said, "Look, there's just nothing really off the shelf that we can use to protect the data that we have." And so they said, after exploring some different options, that differential privacy, this method known as differential privacy is the only solution that can respond to the threat.

They said -- this is their words, not mine. They

said this is -- "We're tied; it is the only thing that we can do to protect the data that we're going to publish."

So they put that stake in the ground.

So knowing that there is a trade-off between privacy and -- the protection of data and privacy and the accuracy of data. The Bureau is tasked with figuring out how much they needed to introduce differential privacy into the data in order to sufficiently protect us.

And -- and this from a daily user standpoint is where, you know, I and many of my colleagues started to get really concerned. You know, the Bureau started putting out these pictures of knobs and slides and scales, and they started talking about, you know, they were going just to figure out how much they needed to differentially privatize the data in order to make them safe.

And so when you think about the accuracy of census data, the quality of census data, and -- and what we as the end users need, the idea that the citizens of the United States have gone and answered the census; they've reported in good faith all the information about their families and they've gone through this whole process of collecting, managing, and analyzing all these data. And at the last step when they decide to report out the data, some statistician at the Bureau is just going to say, "Well, you know, based on our best information, we're just going to

change it by some amount."

And it's some amount that we won't know. I mean, the Bureau will say, "We're not going to tell you, you know, exactly where and how we applied it." They'll give us some information, but basically the -- the data that we are getting are not the data that were reported by the public by some amount; and by "some amount," that it will be impossible for us to know where exactly these data were changed, where we need to look out.

So the part of the Bureau that's responsible for this decision-making process is called DSEP, data stewardship executive policy committee. It's a mouthful of alphabet soup.

So it's not to say the DSEP is randomly or arbitrarily or ignorantly adjusting or introducing differential privacy into the data. I mean, they're making very educated decisions based upon the best information that they have; but it is still being applied internally subjectively by experts, and we won't know the degree to which those data, once they are publicly reported out to us, the degree to which data are right or which -- or that they are wrong.

The Bureau, they would say "This is great; this is what we wanted." But for us, those of us who are practitioners who rely on accurate data at small levels of

geography, the differential privacy process -- not the reality of the actual data -- is what's going to determine the outcome and the design of districts.

And this, as a practitioner in redistricting, this is what you don't want, you don't want a statistical method of changes to the data to decide what a district looks like. You want to the degree practicable the actual data for the district to decide what the district looks like.

So the question arises as we've gone through this process of watching the Bureau do this: How much of the data get changed? How -- how much of it gets changed? How severe are the changes that are introduced by differential privacy?

So the way that the Bureau measures differential privacy in general is they have this thing called "epsilon." So basically an epsilon is a budget; it's an amount of change that they use to change data at different levels of geography and for different variables. And generically speaking, if this epsilon, if this, what's called privacy loss budget, is -- is zero, that would mean all the data are changed; and if the epsilon is really high, then it would mean that it doesn't change very much. The data are largely going to be accurate, and there's not going to any privacy.

So when I show you the slides and scales and dials that the Bureau are using, really what they're -- what

they're referring to is changing how much epsilon or where this epsilon or privacy loss budget is being applied.

And this is kind of where the private, secret part of adjustments to the data are happening in the background.

We know how much epsilon there is and we get some high-level information about which geographies it's being applied to, but we -- we don't get much insight into the details beyond that.

So what the Bureau did to their credit as part of the process, is they started applying different epsilons and different algorithms to try to privatize the data. So they released different what they called "demonstration data sets" for experts like myself to be able to look at and provide feedback on.

So they gave a version of the data set back in October of '19, and a couple of more versions in 2020, and then they gave one last version in April of 2021.

So each one of these kind of incrementally was a little better, it was changed a little bit and enhancements were made; and so the user community was able to follow along and comment as they were developing the differential privacy -- differentially privatized data.

So when we look at the data, I'm going to -- I'm going to give you some very, very small snapshots of some things that we saw in the data as we went through this

process.

We noticed throughout the different -- different data sets that they had released, that a lot of data were changed. A lot of the data were changed.

And without getting into how much they were changed -- and those are the things we wanted to say, was like, well, you know, out of all the blocks, how many blocks were changed? What percentage of blocks' data were changed?

'Cause when we go back to the 2010 census, you would say when they used their private -- their privatization methodology there that a couple percent of the data had been changed; you know, we've -- we've moved a family here; we've moved some individuals over here, but the data set was largely intact by the time the 2010 census was done and was reported. And we knew the vast majority of the data was reported with the right people and the right place.

When you look at these data in the table here, I want to draw your attention to the far right-most column. So what that far right-most column says and -- and this is for blocks in the Alabama case that we worked on, but it's representative of all states, out of -- when you look at the block level, the percentage of census blocks, the data for census blocks that were changed all up was 94 percent.

When you looked at data for Hispanics, 98 percent of the data were changed; for whites, Blacks, others, again,

96, 97, 98 percent.

So what we saw before looking at any other matrix about how much it had been changed, what we were seeing is that for all intents and purposes, all of the data at the block level that we need to use to do our redistricting work had been changed. All of it had been changed, not a couple of percent.

And that -- that was really eye opening and honestly very frightening to us knowing the importance of each individual block in crafting districts.

So it seemed to us when we saw this is when you're balancing -- trying to balance accuracy and privacy, that the statisticians at the Bureau had swung the ball all the way over to privatizing the data and, as you can see here, there wasn't much left in terms of accuracy.

So when the last demonstration product was released in April, it was our hope that -- you know, we gave a lot of feedback to the Bureau about this -- it was our hope that, you know, the ball would kind of swing back and be more of a balance between accuracy and privacy. Because the way that it looked right now is there wasn't anything left of accuracy for the data we needed. It was everything was made to be wrong so it couldn't be used.

So the April data came out, and there were some inconsistencies, shall we say, that we -- that we saw that

were really concerning to us.

So I'm just going to give you a couple examples of what differential privacy did to data in Arizona and what made us so concerned.

So the first row here we say "Blocks that changed from greater than 50 percent nonwhite Hispanic to less than nonwhite Hispanics."

So these blocks, these are critical blocks; they're cornerstone blocks that help us design districts. That we need to know blocks that with certainty are either a majority-minority or a majority white nonHispanic population.

And out of 242,000 or so census blocks, there are 6,000 blocks that changed from being a majority to a minority white population.

So for the purpose of crafting a district and for Voting Rights Act compliance and Section 2 cases, these is extremely problematic.

But there are more.

So you would think that if you have a block with children in them, right? If there's kids aged 0 to 17, there better be some adults there supervising them.

In reality there are -- and as of the April data sets, there were 1,768 blocks that the way the differential privacy worked in -- in Arizona, all those blocks if you

look up the data for them would have no adults in them. They would just have unsupervised children. So it's -- it's obviously a glaring inconsistency and a troubling one.

There are blocks that had population, at anybody, in the originally published summary file data, there was nearly a thousand blocks that had people in them; they used to have people in them. But differential privacy said because of concerns about privacy, well, we're just -- we're going to get rid of those people; you don't have any people live there anymore.

Another -- another one. This is a real big one. There are blocks that have population, household population, not group quarters. They got people who live there, but if you look at the housing data for those blocks, differential privacy would say there's no occupied housing units. There's no -- there's no houses there holding the people who live there; and that's almost 7 percent of the records.

Again, you can look at this. The other way, there's blocks that had housing units but don't have any people living in it. Again, an impossible scenario. There's almost 1,500 of those.

And then lastly, you know, one of the demographic measures that we focused on that's really interesting to us -- and we use this for public planning and policy. We use it for, like, planning schools and understanding family

structure -- it's this metric called persons per household or P-P-H.

And so what the -- the Bureau said coming out of this differential privacy exercise is that there were blocks that have more than 15 people per household, and there's 7 percent, 17 -- just over 17,000 blocks that have more than 15 people living in those blocks. And I thought of all the people I've known and all the people I've ever known in my life, and I honestly can't think of any family I've ever known who had 15 people living in a household, and the data the Bureau's publishing says that 7 percent of households looked like that.

So we knew as practitioners that probably what was going on is that when the Bureau was taking, for example, in sort of their records when they said, well, there was people there before, but there aren't any people, we think that their algorithm are just taking people, loose, random people that are left over in the process, and are just stuffing them into households because they don't have anywhere else to put them.

So that to us is kind of a systemic indicator that -- that those data are going to be really, really challenging in a lot of ways but, again, in particular would be very challenging to use data like that for the purpose of redistricting.

Here's another way to look at some of the data from the -- from the April demonstration product.

We can look at measures of -- statistical measures of error. So the first column is the percent for the total population; second column is the adult population; third population -- third column is the children's population.

So we can look at this first row. What is the percent error for the total population, for the adult population, the children population.

So the percent error is not that bad overall. This is comparing the percent error between the original published data and the differentially privatized data. So, all in all, it doesn't look that bad. There's about a 1 percent mean absolute error -- or numeric error. The percent errors, however, are significant.

And what I want to point out is there a big difference between the errors for white nonHispanic and Hispanics.

So the percent errors for white nonHispanics are large, the percent errors for Hispanics -- and I can tell you by extension for Blacks, for Asians, for other minority populations -- the percent errors are dramatically higher than they are for whites.

So the concern we have, again, looking at this data for the purposes of redistricting and wanting to ensure that

we are being compliant with the Voting Rights Act, compliant in Section 2 cases, is that the differential -- the differentially privatized data are discriminatory in the sense that they significantly change minority populations much much more than they changed white nonHispanic populations.

So the concern that we have in redistricting then is that we're going to get different answers in a redistricting exercise because one population is being changed much much differently than another population is.

If you're going to change the data and you've got to represent people equally in a redistricting exercise, you would want the amount of error, the amount of change, to be as equal as possible between different population subcategories; and the reality is that the differentially privatized data treat minority populations much differently and much more severely than they treat the majority populations.

As you go down, look at the last two rows of this table, you can see how many records -- how many records have more than a 5 percent error and how many records have more than a 10 percent error.

And you can see that there are lots and lots and lots of geographies that have more than a 10 percent error when differential privacy is applied.

So this again reinforces what we said earlier, the differential privacy is not adjusting some data; it's changing all the data.

I'm going to give you one other picture of something else to think about.

This scatter plot shows on the X axis this number down here is the percent of the population who are children in the originally published data. This axis, the Y axis, is that same number, the share of the population who are children in the differentially privatized data.

So you would expect that children might make up 20 or 30 or 40 percent of the population in a data set. So this diagonal line right here shows where the originally published data and the differential privatized data should line up.

If it's 30 percent in the original data, you would think it would be about 30 percent in the differentially privatized data. But instead what we find is you read across the bottom row -- 10 percent, 20 percent, 30 percent. If these numbers are 10, 20, 30 percent here, what the differentially privatized are saying if you go up, is that there are cases where if the kids made up 20 or 30 percent of the population before, these dots here would say: No, no, no, they're 70 percent of the population; they're 80 percent of the population; they're 100 percent of the

population.

And these are places. This isn't census blocks, these are -- these are places.

So there are places where -- in Arizona -- where there used to be some adults and some children, and what I told you earlier that there's cases -- there's geographies where all the adults are gone and there's now children. That doesn't just happen at the census-block level, that happens at the tract level and at the place level as well.

So, again, if you're trying to represent and understand the geography of Arizona and you look up a place and you see that the place is entirely populated by children, it's really hard to make good policy and infrastructure decisions, let alone representation decisions for those places, if you look and see that all the data for that place is compromised of children.

Okay. So on June 9th, the Bureau came out and said, "Okay, we have our final number. The April data was close but not quite. We're going to raise the epsilon, the earlier number we talked about, we're going to make the epsilon much much bigger."

And we thought, we suspected that because of the Alabama case they may have been feeling some pressure, they were getting a lot of feedback about all these crazy inconsistencies that we were seeing, so we thought, "Ah,

they've seen the light; they're going to raise the privacy loss budget way, way up, and they're going to fix all of these inconsistencies, and the end result is we're going to get data compared to April that are going to look much, much more reasonable." So when the press release came out in June 9th, we were really heartened; we were really optimistic.

So they put out these charts that showed the improvement in the percent error between the original data and the differentially privatized data between the different demonstration data they put out.

So if you look back here on the left, October 2019, this great big blue mark says: "Oh, there was -- there was a lot of error in places." True.

If you look at this bar, there was less error but still some; and then less yet; and then less yet; and then less yet. And so now they've pointed to the production settings off to the far right where there's virtually no error and they say, "Ah, you know, now, you know, at the place level incorporated places have very, very low errors."

So we saw this chart and we were very hopeful. This looked really good.

And then we looked at the same chart for counties. So, again, we had a lot of error at the county level in the beginning demonstration product; and then it went down; and

then it went down and went down; and now "aha," again this looks -- the geography of counties, this looks really, really good.

But then the geography that we are concerned about, because the census first and foremost is used for redistricting, is: What happened with the blocks? What happened with the census blocks in this final production run?

So the Census Bureau presented at a demographic conference in Texas, and -- and I asked them, I said, "Will this new epsilon, this new privacy loss budget fix these illogical inconsistencies, the cities that have kids entirely as their population?"

And their answer was that they had reduced it -- and this was a quote, they said it, "They reduced it; without additional changes it's not going to completely eliminate it."

So we were really really hopeful when we heard that, that they were actually going to reduce some of these things that were demographic impossibilities. But then when they released the data, we saw in this table down here at the bottom, that not only had some of these demographic inconsistencies not been resolved in the final production setting -- and this is what they're actually going to deliver to us out of the census -- some of these things got

much, much worse. So they didn't do what they said they were going to do.

So when you have occupied housing units more than household population, it went up by 50 percent; everybody in the block under 18, the number of blocks with that situation, doubled; the number of households with more than 10 people in them, that went up by 50 percent.

So at the block level -- and this is their final answer. Their last attempt at this was to make the block level data much, much worse while they made the data for other levels of geography that, you know, for redistricting we don't care about, they made all those other data better.

So when we dug into this, we figured out why.

So these -- these tables, side by side -- there's not a quiz on this -- if you look at the bottom line, this last number, they actually shared with us the privacy loss budget and how much they spent on blocks.

And what you can see is that the numbers for all the other geographies, for states and counties and tracts, those all got bigger, that is more accurate; the only piece of geography that got worse, much worse, were census blocks. And, again, that's the geography that we need, that we rely on critically for doing our redistricting work.

So their last effort was to make those as bad as they could.

So we thought about this; we discussed it at great length, and, you know, we kind of, you know, in the -- in the groups that work on this, the people who do redistricting, the people who think about it, and the best analogy that we could come up with was that, you know, the Bureau found that they had to destroy the data that we had in order to protect it. You know, all -- they have made all of the data at the level of geography that we need to do redistricting and to litigate Voting Rights Act cases wrong.

Those data are critical for so many purposes beyond redistricting.

And so what's going to happen now, since it looks like, you know, the Alabama case is over; there's not much more to say or do right now. It's that redistricting and Section 2 cases are really going to rely on experts and what experts say much more than what the data say because we're not going to be sure when you craft a district that it is what you say it is since we know that those building blocks are all -- the data for all those building blocks are wrong.

So kind of the last point I want to make and kind of the key takeaway here is that there is another source of information that we use as part of redistricting that we use for other purposes called the American Community Survey.

And I would never consider advocating using ACS at least as guardrails for doing a redistricting exercise

because they're very difficult to use, but they are -- those data are based on the 2010 census, which is the last correct stake in the ground that we have for a lot of these areas, and they're in 2010 geography.

So to the degree practicable, I'm advising clients and we're designing ACS data to kind of help know, as one other point, one other stake in the ground, what is going on in some of these blocks and tract blocks and neighborhoods what another set of data says is actually going on there.

So in conclusion, you know, the use of differential privacy, we understand the need; but the use of it, the extent of the use of it to make all the data that we need for redistricting is going to be very problematic, will be very challenging for states; and our fear is it is going to create a great deal of uncertainly and create an environment where litigation and the opinion of experts is what is going to decide the political representation of the country, not the data that the Bureau publishes.

Thank you very much.

CHAIRPERSON NEUBERG: Thank you, Thomas. I mean, that -- that was highly informative, provided a lot of information for us to absorb.

I'd like to open it up to my colleagues if you have follow-up questions.

Please note that we will have the opportunity to

discuss the implications of what we're learning in Agenda Item No. X when we have the ability to discuss with our counsel the implications of this. So for now, please, you know, I encourage you to limit your questions to the academic, you know, content of what we're learning.

Questions?

COMMISSIONER LERNER: This is Commissioner Lerner. I'm actually --

MR. BRYAN: Hi.

COMMISSIONER LERNER: Hi. Thank you for that presentation. I appreciate it.

MR. BRYAN: Yes, ma'am.

COMMISSIONER LERNER: I don't know if this is possible or not, but I would be interested in hearing the other presentation as well and then I'm sure I will have additional questions as part of it.

Or I can ask some now, but I just thought maybe once we heard the second one, that might help.

CHAIRPERSON NEUBERG: You know what, Commissioner Lerner, that -- that's an interesting, outstanding question and I don't know if -- if we actually planned for that with our presenters to stay around for group questioning with both of you; or, if we need to, based on time and -- and commitment, separate these presentations.

And please be honest, you know, presenters.

Thomas, I mean, do -- what's your schedule like?

MR. BRYAN: Yeah, I'm happy to stick around if you want to save questions 'til the end, or I can answer some now and some later, whatever is your preference.

DR. DUCHIN: For my part I'm happy -- I'll be here.

MR. BRYAN: Hi, Moon.

CHAIRPERSON NEUBERG: Commissioners, you know, other thoughts on -- on the process of this?

Thank you so much for your flexibility, you know, and accommodating, you know, our style.

COMMISSIONER MEHL: This is Commissioner Mehl. I agree with Commissioner Lerner, I think I'd like to hear the second presentation and then -- then do questions given that Thomas is willing to stick around, which I appreciate.

MR. BRYAN: My pleasure.

CHAIRPERSON NEUBERG: Okay. If there's no disagreement or different way of thinking -- last chance to chime in -- welcome, Moon, thank you so much for joining us. We're looking forward to your presentation and please take it away.

DR. DUCHIN: Okay. Let me ask the same questions before because now I can't see you. Can you see my full screen in screen mode?

CHAIRPERSON NEUBERG: Yes, I can.

DR. DUCHIN: Excellent.

Hi. My name is Moon Duchin, and I want to talk to you about privacy, census data, and Arizona redistricting. I'll give you an overview and show you some experiments that are Arizona specific.

And I'm in the happy situation today of delivering what I think is sort of good news, which is lots of different ways of looking at the data from a redistricting point of view that I think are really reassuring. So I appreciate the chance of getting to hear the previous presentation and understanding what some of the concerns are there, and I'm going to try to address some of those concerns as I go.

I'm also happy to take clarifying questions as I talk, so please don't be shy to interrupt. But I can't see you, so you'll have to speak up out loud.

Just, I'll start with just a word about who I am and the -- and the group that I run.

So we're called the MGGG Redistricting Lab; we're part of the Tisch College of Civic Life at Tufts University. I'm a mathematician and a math professor here at Tufts and founded the redistricting lab over the course of the last five, six years to work on issues related to census voting and redistricting specifically.

So we're a nonpartisan, scholarly research group

first and foremost. We do community mapping support, so we actually have, for instance, designed software that helps take public input and set up public portals in I think it's 10 states at this point, most recently your neighboring state of New Mexico, and we work closely with -- with Doug Johnson your -- your line drawer.

So it's -- it's a pleasure to be talking to you today.

We also do map evaluation and lots of other things. So I'm happy to take questions about the work at the lab.

Our main funder over the course of the last few years has been the National Science Foundation, which for instance awarded us a data science grant called Network Science of Census Data.

So this is kind of exactly what we do is really try to understand census data deeply. This particular differential privacy study that I'll be talking about was funded by the Sloan Foundation. And I want to mention my collaborators: Aloni Cohen, JN Matthews, and Bhushan Suwal; and in addition Mark Hansen, Denis Kasakov, and Peter Wayner, who all helped us understand all the different moving parts in the census data.

I want to particularly highlight JN Matthews, who did a lot of the experiments on Arizona that I'll be showing you today.

Okay. So I'm going to take a look at district drawing in Arizona and took two particular counties as case studies; so Pima County and Navajo County. Pima being larger than a congressional district, Navajo County being relatively small.

Both counties fairly diverse: Pima largely with white and Latino population; and Navajo largely with white and Native American/American Indian population.

And I'll talk as we go about different kinds of districts. Congressional districts which this time around will have nearly 800,000 people in them; and then I also wanted to consider very small districts, smaller than the ones that -- that, you know, the statewide body is tasked with drawing.

For instance, let's look at the county commission of Navajo County where there's about 20,000 people there.

Especially -- before I go on I'll say, one of the themes that I want to kind of riff on from the previous presentation, is the theme of the role that census blocks play.

So what's fundamentally important for redistricting, of course, is to understand the aggregate properties of the districts that you'll be drawing. And as I go on, I'll -- I'll -- I'm not going to get super deep in math, unless you ask, in which case I'm very happy to.

But the theme I want to hit over and over again is what makes differential privacy be called "differential," is that there are these different levels of aggregation, and by making lots of small changes at the finest level, you end up with so much cancellation that you have really good accuracy by the time you get to districts. And that's what I want to say is like the good news that I want to deliver is that the -- the whole design of differential privacy is many small changes to the smallest units so that by the time you aggregate up, things look really good. And I'll give you a number of demonstrations to go along with that.

Okay. So both of the two counties as I mentioned that I'm focusing on, Pima and Navajo, have significant diversity, so here are some core factors to show you the Native American population and the Latino population in these counties.

So I'll start with the risk. So you heard a little bit about that and I'll kind of give you my -- my point of view on the risk and the difficulty of the kind of attack that this disclosure avoidance is supposed to protect against.

So what we're seeing here is a reconstruction of Navajo County. So you heard earlier that it might be possible with great computing power to potentially reconstruct -- and I just want to give you a sense of how

much computing power or how, you know, difficult or hard that it is.

So we completely reconstructed Navajo County in the ten days or so since being asked to join you today. The run for Navajo was done in under six hours on a kind of laptop the typical college student is issued, so we're not talking about fancy computing power; and what we were able to do, what you see here on the right, is recover a complete person-by-person list, which gives the census blocks, the JOID, the ethnicity, the sex, the age, the race; and then that last column is the number of people that answer to that description in that block.

So what I want to emphasize is -- so this was -- this was quick and if you did it on the whole state, we could get the whole State of Arizona completely reconstructed in a few days.

So it's -- it's -- this is to say that the possibility of reconstruction, that is, of being able to reassemble the original microdata without names but with all the rest of the information isn't theoretical; it's extremely easy.

And here's the way that this works, the -- the census data the way that it's released is in an aggregated form. So for in the P.L.94-171, for every block you get the total number of people answering to various descriptions.

But that means if you -- if you think about the different queries such as how many Black women age 52 to 55, if you think of all the different queries that you could ask, by making a sequence of queries that gives you various equations; and then you just solve the equations just like you did in, you know, let's see -- maybe that will be ninth grade where you get a few linear equations and then you can simultaneously solve that equation. With the powerful solvers that -- that we have access to, you can get an exact solution quite -- quite fast.

So this table, let me be clear about the accuracy and about the sort of status of this table. This table is a perfect solution to the problem of, you know, it exactly matches all the aggregate data. That's what I mean by "perfect solution." So it's a hundred percent consistent with the aggregate numbers that are released.

Now, that may not be the original microdata because there could be more than one set of information that's a hundred percent consistent with the aggregate numbers that are released, but just to be clear that's the only source of inaccuracy is when there's more than one table that would fill in those numbers.

So then, you know, this is called reconstruction. To actually figure out, like, who those people are and what their names are and what their phone numbers are, the thing

that you would need to do is pair this with, as you've heard, easily obtained commercial data.

But the interpretation of the Census Bureau is that this data that you see here on the right-hand side of my screen, this is what they're under statutory obligation to keep private. And so that's the -- that's the kind of situation that the Bureau found themselves in, is that these days it's really quite an easy task to take aggregate data and reconstruct person-by-person data of this kind and they're under an obligation to protect.

I should mention by the way because this is something that comes up a lot in conversations around differential privacy, why would this be sensitive? What's -- what's so personal about this -- this kind of information?

So there I'll just sort of drop a couple of, you know, kind of a bookmark that we can come back to later if you want to talk about it, but one thing to say about that is there are sensitive pieces of information here.

For instance, one example that comes up when I -- when I talk to folks out in the field is, you know, landlords could use this data to figure out how many people are living in a unit, and there might be more people that are living in the unit than is sort of allowed on the lease. That's the kind of information that might make people,

especially people in marginalized communities, think twice before filling out their census form honestly.

So that -- that's an example just from this simple data set.

But as you -- as you heard from my colleague a moment ago, the -- there's another important census data product called the ACS or American Community Survey, it has a great deal more sensitive information about income and about education. It's a long form; it has all kinds of information on it which makes it such a gold mine for social scientists.

Initially the Census Bureau had planned to start privatizing ACS data right away. They've now postponed that while they work through the kinks of how well this is going to go, but down the road they'll be needing to protect that ACS data as well.

So that's the risk and that's the need.

Okay. So this new idea that's being applied here is called differential privacy as you have heard.

And I -- let me sort of say one more word about that term "differential."

So, again, what's -- what I think is differential about it, this is a guess as to why the originators called it that, is that you're -- you're changing the numbers a lot at the lowest level which is protective of privacy, but also

protective of accuracy in aggregation. So the differential, I think, is about the different levels.

The name might imply that I think is maybe a little unfortunate is differential privacy as in more for some group than for others, which I'm going to try to go through and -- and sort of reassure you is not what we're finding when we look at the data.

Okay. So what's the idea of differential privacy? It's that for privacy you add noise. So here's kind of a cartoon image of that. If you want to anonymize something you fuzz it, right. It's the same principle that might make you blur out a face or logo in an image if you want it to make it harder to read, you make the exact numbers fuzzier; but on the other hand, you can still get a decent large-scale picture of what's going on. So that's the underlying idea here.

So what are we going to do; we're going to take every single little pixel, every census block, and we're going to draw a random number from a probability distribution and add that number.

So all I want in here, you know, you were promised pictures of knobs, so I have my own picture of knobs here.

The idea here is that you'll draw -- so in this picture you see these curves in red and green and blue, and they tell you kind of when you draw from those probability

distributions, see how they're all centered at zero. You're picking a random number. It might be zero, might be two, might be negative three, you're picking a random number, and you're adding it to the number using your original data. So you put all the numbers; you introduce noise.

This is why, by the way, as you heard in the last hour, you're going to see an enormous amount of census blocks whose numbers are changed. That is the design of differential privacy is to add a small random number everywhere in such a way that by the time you aggregate it up, it canceled out and gives you really good accuracy.

Okay. So what's fundamental and important about differential privacy?

And the reason why you see the knobs and scales and sliders and all the images is that it gives you control. What it does is by giving this privacy loss budget setting, this number called "epsilon," and then some subsidiary settings I'll talk about as we go, you get to sort of target where you have the greatest needs for accuracy and kind of think about how you distribute that -- that sort of budget over the different elements in this extremely complicated census histogram.

Okay. So let's say -- so the census created an algorithm for doing this called "TopDown," and here's the two main things I want you to know about the TopDown

algorithm. One is that it uses the geographical hierarchy from top to bottom, that's why it's called TopDown. I'll give you an image of that on the next slide.

And the second important thing that is crucial to know is that it's got two phases. The first phase I've already described, you had a small random number everywhere. But then after that there's this stage called postprocessing -- processing or postprocessing where you make the number satisfy various probability constraints.

And some of that, you know, again circle back to things we were hearing about in the last hour like areas with all children, so these are the probability constraints and this is part of why the Census released those demonstration products, was to ask people what looks weird about this data; we will then create what are called "invariants" to go back and address that.

So, for instance, you might put some inequalities down that say "I don't want to see any negative numbers in the population." That's a pretty straightforward example.

Remember, we're adding numbers that can be positive or negative to numbers that can be small, so that can produce -- in principle that could produce some negative -- well, it certainly will produce some negative populations.

So then you can put down an inequality constraint that says, you know, I never want to see a negative

population because that just doesn't make sense; that's not plausible.

Likewise, you can say, you know, I don't want to have, like, 15 children and no adults. Those are all invariants; those are probability constraints that you can add.

So after adding the random noise, there's a second stage where the data is minimally adjusted to satisfy all the plausibility constraints. So this is why the Bureau had all those different data releases that you were hearing about and opened it up to the various user communities to ask the question: What do you see in this data that's not comfortable? And then they layered in additional invariants to make sure that the data wasn't going to come out nonsensical for various uses.

Incidentally, this is a good time for me to say so, in my work, in my research, I'm very focused on redistricting, on different levels of redistricting, so this presentation is about research -- it's a research project that took over a year to really think about how TopDown would impact the redistricting use case.

I do want to acknowledge here that there -- as you heard and as I absolutely agree -- there's a huge number of applications for census data, and I make no claims to be able to talk about the impacts of TopDown on all the various

kinds of social science that use the census, but I do feel like I have a very reassuring message about the use for redistricting specific.

Okay. So what is this census hierarchy?

Well, basically the Census collects -- and actually creates -- various kinds of pieces that cover the territory of the state. So here you're seeing counties subdivided into tracts, subdivided into block groups, subdivided into blocks. That's part of what's sometimes called the central spine, it's kind of fundamental nesting set of geographies made by the Bureau; and then there are various kind of geographies that are off the spine like something with the census called county subunits.

Importantly for you as folks who are really focused on redistricting, districts themselves are off the spine; and so part of what's going on here is what TopDown does is it starts at the big geographies, adds noise and adjusts them and then moves down to the next geography; noise, adjust; noise, adjust all the way down to the bottom. So this the TopDown process.

So the key question that my research team asked and I think is of, you know, primary salience for you is districts get drawn after all of this happened. So if a process is carefully calibrated to give you pretty good accuracy guarantees on the spine, what's it going to do for

districts that are off the spine? And that's what I want to show.

Okay. So let's see some experiments. And for the purpose of these experiments I'm going to use a simplified model -- I should say the Census Bureau has been amenably transparent in many ways, although they have changed their description of what they're going to do repeatedly. One thing they did that was especially helpful was to actually not just describe their code but to release their code.

And so I think our team might be one of the few groups outside the Bureau itself which can actually run the Census Bureau code. So we don't have to rely on the data demonstration products that they released. We can grab a new geography and run TopDown on it.

But for today I'm going to show you a simplified model, a toy model as we say in math, that we nicknamed "Toy Down." And here's a link MGGG.org/dp where you can see the paper where we describe some of differences and some of these kinds of experiments in greater detail.

Okay. So I'm going to ask a number of questions and here's one for starters: So do districts lose population or gain -- you know, does the population level change enough in the district that you might have to worry about, for instance, knowing that you've got a majority-minority district?

So there are a lot of different ways you could address that question, but I'm going to address it in a way that this cartoon animation suggests, which is I'm going to draw random districts. I'm going to draw lots of random districts, and then I'm going to take those random districts, look at them in the reconstruction, look at them after noising, and ask how different are they.

All right. So I get to actually do the experiment and see what happens.

Okay. So here's a first indication. So epsilon as you heard -- remember, when epsilon is small, the lower epsilon is, the noisier everything is; the bigger epsilon is, the more accurate everything is.

A lot of us who have been following this differential privacy conversation around the census for the last, you know, year or more expected the Bureau to land on something like epsilon equals two, three, four, that's what we kind of expected.

So this first illustration shows you for districts drawn in Navajo County, how much did the American Indian population change when you noise the district repeatedly?

So this is 16 different runs. You see all these colored columns, these are 16 differently, you know, applications of noise; and what you see within each column is a hundred random districts that were drawn by this

algorithmic process.

So you can already see that even though the population of the individual census blocks might change, by the time you get up to even these very small districts -- remember these are districts of just 20,000 people -- you're down to half a percent error in the Native American population. That's already with epsilon one or two, you're typically getting an error of like under a hundred people.

But as you heard, the Bureau recently announced it's not using epsilon equals one, it's not epsilon two, it's using epsilon 19 -- over 19, 19.6. And if you rerun all this, you're down to, you know, nearly indivisible levels of discrepancy in the Native American population of a small county commission district.

Okay. So let me illustrate that again.

So looking at Navajo County, looking at county commission districts with a population of about 20,000, these plots show you the discrepancy that's introduced by this TopDown style of differential privacy.

We made a hundred random districts, this is just a repetition of what I said before. We noised them 16 times, we measured how much the Native American population poll changed; and what we find is that even with epsilon equals one, the typical discrepancy is under 500 people. With epsilon 19, the typical discrepancy is under five people.

So I think this illustrates that your ability to know that you've drawn a majority-minority district with epsilon equals 19 seems to be really kind of well-preserved by this style of adding noise.

I'll just say just a little bit more about what you're looking at and, again, happy to be interrupted with questions.

But so from left to right you see the columns in the different epsilons: Epsilon equals one, epsilon equals two, epsilon equals 19.

The rows here are different ways of placing that epsilon budget at the level of counties, tracts, block groups, blocks, right, of the different levels of geography. So an equal split over the level; a top heavy split where you put a lot of the budget at the county level; mid heavy; bottom heavy; block group heavy.

And so I just want to say, you also heard in Thomas' presentation from a moment ago that one of the adjustment that the Bureau made had to do with this split, how much budget do you put at what level, and that they had actually backed off the bottom level, the block level; and actually one thing that I want to say is that what we found in our study, is that if you were to look at only the tiniest districts you might like there to be a lot of budget at the bottom level, but for you, you care also about large

districts; you care also about congressional districts. And for good accuracy at the larger district level, it actually pays to have some budget higher up in the hierarchy.

So in our paper we ended up recommending that the Bureau consider an equal split over the levels as a, you know, part of an understanding that redistricting is fundamentally multilevel: Large districts have to be drawn; small districts have be drawn. We think that actually pulling back some of that epsilon from the bottom level was quite good for a district.

Okay. I want to show you a different way of slicing this that I think is also really interesting and really maybe actionable for you as you think about your process.

So now on the left I'm showing again random districts, and I'm showing the epsilon equals 19 cost. So this time we looked at different ways of constructing the districts.

So what should they be built out of? Should they be built out of those itty-bitty census blocks or what if you built them from larger pieces?

So on the right you see again random districts that are built out of block groups rather than blocks, which you're still able to do even for county commission districts while being well within legal bounds for population balanced

districts.

And check out the outcome. Is that the already very small errors when you built out of blocks just all but vanish again when you build out of bigger units. So this is quite interesting, I think, because it is a finding that I didn't necessarily expect going in.

But if you built some bigger units and then just use the smallest units to tune, there are various reasons you might think of that in a good redistricting practice; but now in the error of redistricting privacy, it will also give you more population totals on your districts. So it's another reason to consider building from bigger units and tuning with smaller units.

Okay. Let's ask the next question. So that demonstration was focused on the Native American population in those county commission districts. But I want to address the question, you know, these different racial groups don't live in isolation, they make up the whole, and one of the things that the differential privacy algorithm does is make sure that everything adds up.

So what happens to the overall racial composition of districts?

Okay. So here's a way that we went aro- -- went about answering that question.

So from the pool of a hundred random districts that

I referenced in the last experiment, we noised each -- we just grabbed four: So random districts number 2, 9, 13, and 46 just arbitrary. Because, if you look at these, they have pretty different racial compositions. So you can see the white, Native, Latino, and Other composition of these random districts, and just chose them to be different from each other.

And then we grabbed epsilon two, an equal application, and just noised them, and this is what we found.

So this is random district number 2 noised 16 times. And if you look, those pie charts look identical, it's because they're pretty hard to distinguish with the naked eye. So the numbers are printed here and, of course, I'll make these slides available so you can study them at any greater detail later.

But, again, by the time you get up to a district of just 20,000 people, the -- the demographic breakdown of the district is extremely stable. And this is even for epsilon equals two, again the Bureau is actually using epsilon 19; much more accurate than this.

Here's random district number 9.

Here's random district number 13.

Here's random district number 46.

And, again, you see the visual constant showing you

that you get pretty striking consistency over the district.

In particular, notice this district had a very slight majority of white people; and you can see that in all 16 runs, that that slight majority is preserved. The numbers are changed a bit, but, again, by the time you get up to districts of this size, you can trust the numbers to be really quite accurate.

Okay. So let me pivot and talk about the Voting Rights Act because that's another topic of fundamental importance for a line drawing body like this Commission.

So what is -- what is important to do if you want to be able to enforce the Voting Rights Act when it comes to redistricting?

Well, as I'm sure you know, to bring a Voting Rights Act case first you need to demonstrate some kind of threshold tests. So these are called the *Gingles* factors, you have to be able to show that it's possible to make a majority-minority district -- you don't necessarily have to draw one in the end, but you have to show that it's possible in order to launch a challenge; and then you have to show *Gingles* two and three, which amount to showing that voting is racially polarized.

But what you see in this picture that I included as a illustrative thought, this is a real picture of the precinct of Chicago in the 2015 mayoral race between Rahm

Emanuel and Chuy Garcia, and it's showing you that the more Latino population was in a precinct, the higher its support for Chuy Garcia. So this is the kind of demonstration that you need to illustrate that different racial groups might have different preferences in this.

Again, why you need this racially polarized voting or RPV is one of the threshold elements that you need to do Voting Rights Act enforcement.

So there was a lot of concern in user communities that differential privacy would mess with RPV. So let me show you the nightmare scenario.

So, here, little skull and crossbones to remind you this is danger danger.

What you'd worry about is that in blue you'd have the original data, the unnoised data. And then what -- what we did to make this picture was we noised it; we added a bunch of random noise to the data; and the pink dots are the noisy data. So they're just the blue dots but moved a little bit sideways.

Let me quickly mention why sideways, because in this plot you're seeing the demographics, in this case the Latino percentage in each precinct; that's on the X axis, and the share of the vote that goes to particular candidate, that's on the Y axis.

So differential privacy because what it changes is

the census data, it only moves things sideways because it only changes the demographics but not the vote. Right? We don't add noise to the votes only to demographics.

So essentially what you'll see in these next few pictures is you'll see these scatter plots, and then you'll see some sort of sideways pink movement, that's the noise being added to the data, and then we fit a line to the data; and the -- the nightmare is that maybe when we fit lines to the noisy data, which is what you're seeing in red, they're really different from what they would have been.

Why is this a risk?

Because the slope of that line is telling you how much difference there is between the preferences of two groups, in this case Latino and nonLatino. So a line with a lot of slope is telling you that there's a big difference in the preferences. Maybe the scary scenario is that adding all that noise will flatten out those lines, and what was really polarized won't look polarized any more because of all those inaccuracies that were added by the data.

This would really be a risk for Voting Rights Act is because we might be unable to test the merit of VRA claims, those threshold factors of racially polarized voting. Maybe you just lose all the signal to the noise.

So here's what we found when we tried this.

So, again, these are experiments done in the last

few weeks. So let's look at Pima County, and we'll look at Native American support for Biden in the 2020, which as you know, statewide was rather close.

And so here's what you're seeing in these pictures. You see how those lines of fit are fairly flat? What that's telling you is that in Pima County, Native American support for Biden wasn't that different from nonNative support for Biden. It's only a small difference is what's inferred from these graphs.

But interestingly, so the top row is epsilon equals one, middle row is epsilon equals two, the bottom row is the actual rough choice made by the Bureau, epsilon 19. What you can see is that pink fuzziness essentially disappears by the time you get to the bottom and the line of best fit doesn't budge at all.

Okay. So what is this saying? This is saying that this method I'm showing you here which is just linear regression in the context of racially polarized voting is called "ecological regression." Ecological regression, it's findings are not budged at all by the application of TopDown style differential privacy.

Okay. So let's look at Latino support for Biden. And now you see maybe a little more slope, so Biden got a little more support in Pima County among Latino voters than nonLatino, but it's still a fairly mild slope.

And here is white support for Biden, now the slope goes a little bit down saying that white voters in Pima preferred Biden a little bit less than nonwhite voters did.

And, again, as you look at the top row, epsilon equals one; the middle row, epsilon equals two; and the bottom row, that noisiness all but disappears, and the finding is extremely stable.

Okay. I'll summarize that in a moment, but first I want to show you Navajo County.

So Navajo County is just way, way, way smaller as we saw early on, and so you might worry that with so many fewer data points -- I believe there's only 14 precincts actually in Navajo, if I remember right.

So with so many fewer data points you might worry that it would be really hard to get a signal out of a method like ecological regression.

Well, what you see is that on epsilon equals one, equals two, epsilon equals 19, that noisiness all but disappears, it's extremely stable; that same high slope that you see in the blue, which is the unnoised data, is preserved in the noised data.

Okay. So there's no threat to the finding of racial polarization that's imposed by this style of differential privacy.

I'll mention that Navajo is quite different from

Pima. Here, look how sloppy that line is. It's telling you that the Native support is inferred to be extremely high for Biden, and the nonNative support is pretty low.

Here's the picture for Latino support, this is pretty interesting.

Here the slope is so steep that the point estimate is off the charts. This is telling you that this method concludes that Latino support, Latino voters did not prefer Biden in Navajo County, which is an interesting finding that you might not have guessed going in.

And similarly for white support.

Okay. So the point I'm trying to make in these demonstration isn't to do with a particular method like inferring racial polarization, it's simply to ask the question: Some of the established techniques for doing RPV analysis, will those established techniques be threatened by differential privacy?

And here are the finding -- and we've now tried this across many different geographies here in Arizona, in our original study we did many places across Texas, many different elections, and so far we're finding as long as you use kind of reasonable practices for your inference, here it's ecological regression that's weighted by the number of cast votes, as long as you're using reasonable updated practices, we're finding no threat to the ability to detect

racial polarization.

I'll quickly mention, as some of you may know if you're RPV experts, the state of the art technique for racially polarized voting isn't this regression, it's something else called "ecological inference." We ran those tests, too, but they don't produce as pretty pictures. So for our demonstration I'm showing you ER, but we also checked out the EI results and found similarly stable outputs.

Okay. Here's a summary. And, again, this is epsilon equals two. Looking at numbers, the unnoised numbers in the lowest of 16 noisy trials and the highest of 16 noisy trials, you know, going in I expected you might see things changing by 5 percent; and instead you just see remarkable accuracy even at epsilon equals two; by the time you get up to these districts and epsilon equals 19, it will only be better.

And then here's the result for Navajo, where the largest minority group is Native American, and you see here the stark difference between Native and nonNative support for Biden that's inferred by this method and, again, remarkable stability across 16 noisy trials.

Okay. So I'll just wind down so that we can move to questions for both presenters, but one thing that I hope you're wondering is how realistic are these experiments?

They sound reassuring, but how close are they to what's actually going to matter on the ground?

So I'll -- I'll just remark, our studies used -- our original study used the census code release from July of 2019. Since then they've made a lot of announcements, as you've heard, about changes to their methods; so some of those are in response to the pushback they got from end users who were really worried about the -- the changes in the data.

So let me detail a few of the analysis -- a few of the differences between our experiments and what the Bureau is actually going to be doing.

So one thing I used the simplified model called Toy Down; they're using a more complicated model called TopDown. In every instance their model is more accurate overall. So when I'm telling you something reassuring, the truth should be even more reassuring because TopDown is consistently more accurate than the simplified model I used for the demonstration.

A second thing which was a recent announcement, which is I showed you the random number distribution, they slightly changed the distribution from something called the gaseus (phonetic). All you need to know about that is that it has thinner tails which means you're less likely to change things by a lot.

So every single one of these recently announced changes is because move in the direction of more accuracy as you also heard in the last presentation.

Here's a big one that I was pretty excited about. So the census hierarchy that I mentioned, the central spine goes from county to tracts, what are called block groups to blocks. The Bureau actually announced that they're going to tweak to something they're calling "optimized block groups." Took a little while, a little digging asking a few people to figure out what that meant.

What that means is the original block groups were not that well tailored to fit into cities and towns, and optimized block groups are slightly tweaked so that they fit cities and towns better. This is good news because using optimized block groups means you can expect better accuracy at the municipal level. The reason they did this again is feedback from users that said, hey, we really need our city and town totals to be pretty accurate, so can you try to work that into your budget allocation, and this is the Bureau's response.

And then finally, lots and lots of work has been done tuning what's called the workload and the invariants. Basically, many of the ways that the Bureau is handling households and other kind of structure in the data reporters and so on, they've been refining their exact ways of, like,

making those plausibility constraints so that you don't get silly looking answers like lots of kids and no adults, and all of that is just going to produce better accuracy at the end of the day.

Okay. So all of these make the discrepancies substantially smaller than what I've already described.

Finally, I'll just end with what I think are some take home messages.

So, one, the privacy risks are real; they're not abstract. The reconstruction is readily at hand.

Two, there's kind of some no going back to the way things used to be. So the previous disclosure avoidance methods from ten years ago and more such as data swapping are what I would call opaque, ad hoc, and underpowered. Opaque meaning we don't have an exact accounting of what was done. So with data swapping, you heard about suppression. Data swapping would be taking information from one block over here, taking information from another block over there and just exchanging them. Like literally swapping the numbers in order to protect privacy. The problem is that we don't know how much swapping was done, how big the changes were that were imposed by it. We don't have any way of kind of taking account of the impacts.

So what's fundamentally important about this differential privacy move is its transparency. Is that

ultimately we get a complete -- we get a complete accounting of what was done with what settings, with what invariants. I'll tell you, less I sound like a cheerleader for the Census Bureau, there's been things that have been frustrating about the rollout, like the story changing so much over time so it feels a little bit like moving targets when you're trying to study it, but they have a commitment to ultimately making public -- even if those decisions were being made in kind of a staged way -- ultimately we'll have full access to -- to code and to settings.

And that just creates a lot more transparency and a lot more ability to understand how much we know and how much we don't know, which was opaque, ad hoc in -- in the earlier regime.

And underpowered meaning that the Bureau concluded that its old techniques of disclosure avoidance such as swapping couldn't protect against the kind of reconstruction attack that I showed you at the beginning.

So, you know, the big take home message is that for every geography we considered, whether it was large county, whether it was congressional districts, whether it was small county, whether it was municipal, we concluded that the census data are clearly completely adequate for every redistricting application that we considered; and, you know, we want to sort of be clear that this only applies to the

places that we've looked at. But I hope that one thing our team has done is hopefully laid out a methodology that could be applied for a lot more places.

So as I mentioned, we were invited to kind of look at Arizona just a few weeks ago, and we're able to take those methods and apply them in a new place, and I hope that people who are here in this meeting will get in touch if they want help getting set up in our system to do these kind of studies in other places and to ask other precise questions.

So we find no threat to Voting Rights Act reinforcement or reasonable population balance or the ability to draw majority-minority districts as part of VRA in any of the places that we looked.

And here's some kind of nice note to end on I think, is that in addition to kind of reassurance that I hope to have provided about the core tasks involved in redistricting, the studies suggest that it might be time to update best practices for redistricting.

So here's a few points that came out. I'll highlight, though I've discussed them already in the last -- course of the last hour. One is seems to be a pretty good practice to build from bigger units and just tune for the smaller units. Many people have been talking about that as a good practice for years but now we have a better reason

which is better accuracy when we do that.

Weight your regressions. This is a pretty simple point for math and statistics person for social sciences. If you're trying to fit alliances and data, you should probably weight the data. But we found, and that's something you could check out in the full study that I linked earlier, that if you don't weight the regressions that the small additive changes made by differential privacy can really throw things out of whack.

And then, finally, let's talk about population balance. So it's the practice typically for legislative and smaller districts that the population deviation should be no more than 10 percent from top to bottom without a good reason. However, over the years the case law that's evolved around congressional redistricting has been to impose much much tighter balance on the districts to the point that most states have no more than a one-person deviation between the districts. That's just become the practice over the years since the 1960s when the courts first, you know, put out that edict, one-person, one-vote, right.

Well, what this study suggests is that if we were to look at the sources of accuracy/inaccuracy in the census data, and if we were to remember we typically redistrict in a way that is supposed to produce districts that would last ten years, the zero balance habit it might have been little

excessive; and now with the introduction of this new explicitly noised data, it may give us finally a reason as the redistricting community to transition away from that zero balance habit, and to understand that a little bit more discrepancy between the districts might be best suited to, you know, the fundamental things that matter when we draw districts.

Okay. I'll stop there.

CHAIRPERSON NEUBERG: Thank you. Again, an incredibly informative, very helpful presentation. Thank you so much.

I think the Commission and our public and all of us are learning a lot.

I'd like to open it up to questions from -- from my colleagues.

COMMISSIONER LERNER: So this is Commissioner Lerner. I found both presentations really interesting, and I guess sort of a natural question is why the different perspectives that exist? 'Cause we saw from Mr. Bryan, you know, basically saying the inconsistencies are very strong and striking and can really change the way -- you know, what we do, it could really affect what we do; and then Dr. Duchin, you sort of give us the other indication.

So just for me being -- my statistics is rusty, so

I would just be -- and I'll start maybe with Mr. Bryan here, like, why the discrepancies in conclusions between both of you -- with both of you?

MR. BRYAN: Yeah, for sure. I mean, and, you know, Dr. Duchin, that was a great presentation. I've heard some of those points of view before, I mean it's really great this holistic open view of all the different ways of looking at this. It's a super dense, complicated, hard subject and there's not a right answer how far you swing the pendulum back and forth between privacy and accuracy.

You know, I think my position is just based on, you know, years of experience in doing a hundred -- like actually boots on the ground doing a hundred of these, you know, projects where you look and say, well, you know, the Voting Rights Act, you know, the *Gingles* positions we spoke about requires a majority-minority population, right.

And oftentimes in these cases you have to figure out whether you can make one majority-minority district or two majority-minority districts, and knowing exactly where minority-dense populations are so that you can either add them or subtract them so that you have some certainty that you've actually got a majority-minority district, is really important.

So, you know, and I -- I've heard the argument and I understand that, you know, in aggregate, you know, if you

aggregate a thousand blocks, you know, it's going to look great. But, you know, we don't have in -- in litigating these cases, we don't have the luxury of saying we want everything in aggregate. You know, we have to get down with a scalpel, look at some individual pieces of geography and say we want this one in; we want that one out. And that can be for a lot of reasons, not just for the math to make sure you've got a majority-minority district, but sometimes you got to respect incumbency or, you know, voting precincts. There's going to be other things that can dictate whether you've got a specific piece of geography, one specific piece of geography in or out, and you need to know whether that piece that you're adding in or out with some certainty really is what it's to be purported to be, right.

I think it's really that concern that gives me my point of view.

I think the other thing is -- and, again, you know, looking at detailed numbers, precise numbers, and needing to have accuracy for individual pieces of geography was highlighted, you know, when we did, you know, the Alabama case. We were in the analysis and what we found is that there are districts -- not just in Alabama but across the United States -- where there are congressional districts that were designed and reported to be majority-minority populations; it was designed and reported to be a Black

district or a Hispanic district.

Now just barely, but it was. And when we ran those same districts using differentially privatized data, they were no longer.

So whether you -- even in aggregate, whether you use differentially privatized data or actual data, that alone can change whether a district is a protected district under the Voting Rights Act or not.

And so that is -- is obviously of great concern to us and that -- that specific case was the type of thing that is going to end up deciding the formation of those districts based on expert's opinions and whether somebody can actually go on the ground and point to a block that may or may not be included or, you know, that will decide whether that district is majority-minority or not. Because what's going to happen is people will go and litigate this and look at a block and say "This block has a hundred Black nonHispanics or this block has a hundred Hispanics," and the data is going to say -- from the census is going to say they're all white; and then when you debate what the boundaries of the districts are going to be, you're going to have people going to blocks that are obviously completely wrong, and say if you take the reality of what this block is, it would be a majority-minority district. If you just take the reported data which people are going to be able to show is wrong,

it's not. And that's going to be the crux of fight after fight after fight in the courts over those districts.

DR. DUCHIN: A couple -- a couple of remarks. So one is I think that you didn't hear any factual disagreement, I think, between the two presentations.

MR. BRYAN: Right. Thank you.

DR. DUCHIN: Yeah, I didn't hear any factual disagreement between the two cases.

MR. BRYAN: Yeah.

DR. DUCHIN: What I -- what I heard instead was a kind of a very different focus, whether what we care about is the accuracy of the block or whether what's salient, what's important to us as people trying to do redistricting is the accuracy of the district.

Differential privacy is specifically designed so that tiny errors on the little pixels add up to something very accurate by the time you get to the unit of the democratic representation, which is the district.

So that -- that's one observation I would make just about kind of putting those two presentations together in your mind. Is -- is the question is, do we have kind of as people who are trying to do the best practices in redistricting, should we be alarmed that blocks change?

So I think that -- to help you kind of reconcile the two in your mind, I think that's the question.

And what I'm pointing out, for instance, in the Navajo County district that I showed you in some of the experiments is that by the time you're getting to the settings that the Bureau is actually doing, the change in the number, for example, of Native American residents in a county commission district is like five people; and I think that if -- you know, that that's extremely protective of all the kind of legitimate interests that -- that we have in thinking about redistricting.

So the Voting Rights Act specifically, let me just make one little nuance there which I'm sure is well known to all of us.

The condition that asks you to draw majority-minority district, you have to demonstrate that it's possible to draw a majority-minority district. The courts have been absolutely clear over and over for many years, that at the end of the day the Voting Rights Act remedy, the districts that you draw does not have to be majority-minority. And, in fact, you can get in -- in kind of constitutional trouble if you target any particular demographic percentage in the ultimate districts that -- that you draw at the end of the day.

So -- so that's why people call *Gingles* one a threshold condition, you have to show that it's possible to draw a majority-minority. I'm sure we completely agree

about this because this is a pretty clear saddle of law --

MR. BRYAN: Right.

DR. DUCHIN: -- and that's not threatened at all, the ability to draw a majority-minority district is not threatened at all is what these demonstrations show.

COMMISSIONER LERNER: Thank you. Thank you. It helps clarify just to hear that.

MR. BRYAN: Thank you.

COMMISSIONER MEHL: This is Commissioner Mehl with another question.

As a Commission we're going to receive the census data and we don't have an alternative, and so I'm not clear, is it -- is there any reason we should hesitate to use the census data; and, if so, what would be the alternative?

DR. DUCHIN: I can -- I can take that.

MR. BRYAN: I have a point of view. Go ahead.

DR. DUCHIN: Okay. Well, some states have started to release their districts already you might have seen and -- and, for instance, they're using ACS data to do it.

So unlike the decennial census which just comes out ten years, there's a -- there's a new American Community Survey every year and there are five-year rolling averages and some states decided to use various kind of estimation products to produce their districts.

But I think on the basis of every single geography

that we looked at, we can expect the decennial data to be absolutely of high enough grade to vastly outperform any alternative.

So, you know, we -- my group, we don't have a dog in the fight; we went in looking to see is the census data going to be good enough to use? And I'll tell you actually, I was quite surprised with the number epsilon 19 came out. As I mentioned, I was expecting something more in the neighborhood of two, three, four. I think of two, three, four as moving the data around, and epsilon 19 as, like, barely breathing on the data.

MR. BRYAN: Right.

DR. DUCHIN: You should think of this as sort of a dry run for later when the Census might want to dial up the data protection. Right now they are trying to get everybody comfortable with what they're doing and the changes in the data are so slight that we find just no risk at all to anything that you want to do.

You should continue to think, in my -- my view, you should continue to think of the decennial census data as the gold standard for redistricting.

MR. BRYAN: Thank you. Yeah, and so I can dovetail with that, you know. When they -- I was surprised, too, with that when they came out with that gigantic epsilon.

But, you know, what we found is that they packed

all the benefit of that into higher levels of geography and they actually took more away from the blocks than where we actually began with.

So that's -- you know, when you get the error, the error in essence has been passed on to smaller cases of geography that we have to use for the purposes of redistricting.

So about the ACS, I will say this: I was actually at the Bureau, and I helped work on the design of the ACS; I'm very familiar with it. The ACS is built on 2010 geography, census geography; and geography changes, it's changed dramatically for 2020, and it's based on the 20- -- 2010 census data, and it's best on estimates that were derived from the 2010 census.

And we, unbelievably, are still using the ACS data right now to litigate Voting Rights Act cases at the doorstep of getting the 2020 data.

So those data that we have right now, we have down to the block group level through something called the DOJ file. So the Department of Justice gets a special tabulation of data down to the block group level that has citizen voting age data in it, and it's what we use to measure voting, minority voting strength and one-person, one-vote matrix in our VRA cases.

So what we have done and what we're forced to do

because the Bureau doesn't release it, is that we actually make up commits; we take those data, we make block-level estimates of EVP, block-level estimates of minority voting strength off of that DOJ file, and that's what we use and the courts -- it's an ugly business, but the courts have accepted it, and we use it, and we won in those cases by that data.

So what we are doing in the states that I am representing, the states where I'm doing redistricting work in, is we are taking that ACS data which is based on a 2010 geography and based on 2010 data, and we're converting it into 2020 geography so that they're comparing apples to apples, right.

And what we're doing then is we're taking those data and we're saying, okay, where are there big similarities and where are there big differences from where the 2020 census data are?

And so what we're doing is building infrastructure to say, where are the places that they're totally in alignment; they completely agree? And we're saying we're not going to worry very much where those parts of the state where differential privacy doesn't seem to have made the data that much different than what the ACS would say; but where there are places -- and there's going to be many -- where the ACS says, hey, this whole area is Black and the

census data comes in being differentially privatized says, no, no, it's all white.

Those areas are the areas that in the process of redistricting what we're going to do is use the ACS data to inform our effort, to put guardrails up, and to help us know and understand the places where there are significant differences, and where we can expect legal challenges.

Because if the ACS data says this area is all Black or Hispanic or Asian, and it really is, that helps highlight where there's going to be an issue and where people are potentially going to raise legal issues.

So you can't replace -- to Dr. Duchin's point, you can't replace the census; it is what it is, and we have to use it, and in most regards I will agree it's a gold standard. But it would be foolhardy to go into the exercise with the changes differential privacy is giving us and just completely ignore another really big, powerful, important piece of information that can tell you at least where you need to look to see where the impact of differential privacy is likely being most severely felt.

DR. DUCHIN: I would just add a small comment. I'm sorry I had to move because I'm finally back at work and therefore can't control the noise level that's going on around me, so I hope you can hear me.

But what I would say is that the ACS has always

been a really useful sort of supplementary data, but courts have reminded us over and over again that it's -- for redistricting purposes it's not up to the level of decennial census partly because it's based on the survey data, and the decennial census is based on a -- on a complete enumeration.

So there are a few interesting cases, for instance, in Dallas County in Texas, where late in a cycle -- and this is kind of I think an interesting point, late in a decennial cycle a Voting Rights Act came forward, and the use of ACS data was compared to use of -- of now eight-year-old decennial data, and the court said: When in doubt, the gold standard is decennial.

And I see nothing in any of the experimental work that we've done to shake that faith in the decennial census as a valid and reliable source for you to do best practices in your districting.

CHAIRPERSON NEUBERG: I have a question about the different implications of this for congressional versus legislative boundaries.

It -- it sounds like the smaller, narrower group is going to be more at risk of the noise. Dr. Duchin, you mentioned I believe a 20,000 population range in which you felt was reasonably accurate.

I'm curious if the two of you, given an epsilon value of 19-plus, can agree on the size -- and I know that

it's different depending on -- on the, you know, the nature of the groups, but -- but is there a specific population number with an epsilon value of 19 that you would agree we can feel really very comfortable with the accuracy?

And that's to both of you.

MR. BRYAN: Yeah, I mean, I think that that's an excellent question.

I think that what -- this is going to get fought out repeatedly in courts; this is what I'm going to do for a living for the next 20 years now, is that, you know, so many times when you're doing redistricting and you're fighting over these cases, you know, you have to make a majority district. A majority can be 50.0001 percent, right?

And so now the question is going to be, you know, given the uncertainty introduced by differential privacy is how much higher does that number have to be before you're sure that it really is a majority-minority district, and it's going to withstand litigation?

So I am 100 percent sure that this will get litigated, and the question is going to be: How high does that bar need to be?

I don't have an answer yet because I haven't seen the -- the data. I can't wait to see it and to find out; but I believe -- it's my belief -- that the days of us crafting districts of 50.0001 percent majority are over, and

you're going to have to have a higher bar because of the uncertainty of differential privacy to withstand legal scrutiny and say you're really sure it is a legal majority district.

DR. DUCHIN: Tom -- let's just be, I just want to be sure we agree on this. You do not have to be over 50 percent minority to withstand VRA challenge, right? To be clear --

MR. BRYAN: Right.

DR. DUCHIN: -- that's widespread misunderstanding about the VRA. All you have to show to be a VRA plaintiff is that you could have drown a 50 percent district.

MR. BRYAN: Right.

DR. DUCHIN: You agree? Just checking. Okay. Good.

MR. BRYAN: Right. You can't be sure, so the certain what's the "can you drive or not" part is going to get fought over.

DR. DUCHIN: So for you as land drawers, please do not focus on the majority-minority thing, that is not -- it is, in fact, counterintuitive when you're trying to be safe in VRA terms, it's very not much what the VRA directs you. I think we agree to just be clear.

But -- so just to address the point about the size of the districts. So as you're drawing congressional and

legislative district, so Arizona's population decided by 60, which is the number of seats in the House, gives you an ideal size for your House districts of about 120,000. That's why I was showing you districts all the way down to 20,000 and saying even those are very, very safe.

So that means for your legislative, House, Senate, and congressional districts, you can feel rock solid that the numbers are going to be within a few people.

So, you know, that was the -- and I should have emphasized that earlier. All the way down to below the lowest level you'll be drawing, the numbers are -- are really in great shape for everything, every legitimate purpose.

COMMISSIONER MEHL: Dr. Duchin, it actually would be double that, because our legislative district are only 30 not 60 --

DR. DUCHIN: 60 --

COMMISSIONER MEHL: -- and we have two --

DR. DUCHIN: 60 people elected from 60 districts, yes. Thank you. Good point.

So -- so that means that means 240,000, you are in excellent shape and have -- that's far, far larger than the demonstration districts in my presentation, which are already sound at the size of just 20,000.

CHAIRPERSON NEUBERG: And, Thomas, I'm curious,

does the differential sizes of the districts at all influence your bottom line with -- with what your -- your perspective is?

MR. BRYAN: Yeah, I don't know yet. We'll -- we'll have to see it.

I mean, in theory it shouldn't if it's been, you know, applied the way that we've been told it's been applied, but I just -- I want to reserve judgment until I get to play with it.

CHAIRPERSON NEUBERG: Very fair.

MR. BRYAN: Thank you.

Other questions?

Well, I think we all found these presentations incredibly informative, so well-done, really constructive for us on a practical level, you know, to learn and -- and apply the information which we will have further discussion.

But unless there are any other questions or comments from my Commissioners, we can thank you very much.

And also thank our legal counsel for doing such an excellent job of -- of vetting and looking for expert witnesses to come and be part of our process.

DR. DUCHIN: Thank you so much for your invitation.

Nice to see you, Thomas.

MR. BRYAN: Yeah, it's nice to see you, too.

And always, it's a -- it's just such a pleasure. I

love these environments where I can learn, you know, from colleagues like Dr. Duchin always has excellent points to help me be a better expert and from you, the people I inevitably end up in court with, it's nice to see you and I'll probably see you again.

CHAIRPERSON NEUBERG: And if there's any new data that, you know, come up, please feel free to, you know, send our way.

MR. BRYAN: Yeah.

CHAIRPERSON NEUBERG: We have some interested parties here.

DR. DUCHIN: And to that point, if you have any requests for, you know, investigations that are more particular than -- that what you saw, we'd be really happy to be called on to support your important work.

MR. BRYAN: Yeah, for sure. My pleasure.

DR. DUCHIN: Take care.

CHAIRPERSON NEUBERG: Okay. Thank you. Bye-bye.

MR. BRYAN: Thank you, guys. See ya. Bye.

CHAIRPERSON NEUBERG: Thank you.

I think with that, I imagining maybe everybody participating would like a brief break before we resume our agenda, which is Agenda Item No. VI, update from the mapping consultants which I also anticipate to be a robust discussion.

So how about we take a five-minute break. Does that work for everybody?

Okay. We'll reconvene. It's 10:13. We'll reconvene, you know, about 10:18, -19, so at 10:20 we're in the midst.

(Recess is taken at 10:13 a.m. to 10:21 a.m.)

CHAIRPERSON NEUBERG: Okay. Just a heads-up to everybody -- thank you for reconvening so quickly -- I just heard there's a power outage at our office, and it's not clear that our staff is online, not to mention that if there's a power outage that's air conditioning and a lot of other, just, practical challenges.

So let's all stay tuned. We cannot reconvene until we have our Executive Director and -- and our staff online.

So -- so please, everybody -- and I don't even know if the audio and is everything working, but thank you for everybody's patience.

(Recess taken from 10:22 a.m. to 10:27 a.m.)

CHAIRPERSON NEUBERG: Okay. If I could have everybody's attention, our staff is moving buildings into a location where power is up and running. We believe that in 15 minutes we will be up and ready to go.

So I'm going to suggest that we take a 15-minute

recess and hope that the power and everything will be up and running.

So at 10:45 we will reconvene. You're welcome to exit and reentry [sic], you know, this link or -- or stay online, you know, with your mic and video off.

So we will reconvene at 10:45. Thank you for everybody's patience.

(Recess taken from 10:28 a.m. to 10:45 a.m.)

CHAIRPERSON NEUBERG: To those who are already live and listening, our staff is actually going to join our legal team at their official offices, and that is going to be about another maybe five to ten minutes.

So, you know, obviously we apologize for the disruption, but it's essential that our staff be involved and have full access, so stay tuned.

But bottom line is new estimated time of reconvening at 10:55.

(Recess taken from 10:45 a.m. to 10:57 a.m.)

CHAIRPERSON NEUBERG: Just an update for anybody who's listening to our feed live, our staff has moved to our legal counsel offices; they are setting up. Obviously it takes a bit of time to relocate, get their equipment up and running; I do believe it will be a matter of minutes, but

please stay tuned.

So it will be a bit longer while they are relocating.

(Recess taken from 10:58 a.m. to 11:04 a.m.).

(Whereupon Commissioner York is not present.)

CHAIRPERSON NEUBERG: Okay. Everybody, if everybody is online, we have some updates, and I'll wait until everybody -- I can see the tiles all open, and everybody is here and ready to move forward.

I see our Commissioners, our counsel.

Director Schmitt or staff will not be able to join given the power outage, and so I believe and I would just like to confirm with legal counsel that we have the group necessary to be able to decide next steps.

We have a challenge, and so our YouTube feed is dead and we cannot recreate that; and because of that we are not able to continue the meeting as is with providing the notice to the public.

And so given these legal constraints, we really cannot move forward with our business today and so, unfortunately, again, given these unusual circumstances of a power outage and observing the legal requirements of the Commission, we will need to reconvene next week and continue our business.

We are scheduled for our regular, you know, meeting on Tuesday the 20th at 8:00 a.m. We actually have quite a bit of business to do, and it's possible that we will not be able to do our full business on that one day on the 20th. Before I continue, I would like to turn it over to our counsel because this is an unusual emergency situation with scheduling, and I don't want to make any errors with scheduling or doing business that's outside of -- of -- of the norm and what's required.

So before we solidify next steps, I'd like to ask our legal counsel to chime in with what's available to us and any recommendations.

MR. JOHNSON: Yes, thank you, Chairwoman, and you -- you've explained it very well.

Due to the emergency, the YouTube link if it goes down for a certain period of time, it's defunct. That was the YouTube, that's the open meeting venue for lack of better terms that's required. Because it's defunct, we cannot send another link to the public within a shorter than 48-hour period.

This discussion which we're having right now is pursuant to emergency exception for the open meeting laws for purposes of discussion and determination as to when we can have a follow-up meeting.

We do have the Tuesday meeting that's scheduled and

then the concept is we will try to set aside some other time that week or to either reconvene the meeting that is going to be taking place on -- on -- on Tuesday, so things will be tabled to the next meeting or a whole separate meeting agenda.

But the reality right now is trying the figure out availability for the Commissioners.

Very unfortunate.

COMMISSIONER MEHL: Chairwoman Neuberg, I'm traveling next week and I had blocked out Tuesday to make sure I was available Tuesday, but it will be difficult for me to be available on another day; but I could go as long as we would need on Tuesday, if -- if that would be possible.

CHAIRPERSON NEUBERG: Other Commissioners?

I believe it would be a little longer than we, you know, normally would expect. But, you know, I mean it's a robust conversation, we have our timelines; we have a lot of follow-up conversation.

But it is what it is.

COMMISSIONER LERNER: This is Commissioner Lerner. I'm available on Wednesday or Thursday of -- of that week, but would not want to meet unless Commissioner Mehl can also be participating.

CHAIRPERSON NEUBERG: And can you, Commissioner Lerner, do a long day on the 20th?

COMMISSIONER LERNER: Sure. I can do whatever we need to on the 20th, I always set aside the whole day just 'cause I never know.

MR. JOHNSON: And just -- just for -- for planning purposes, just understand what are the consequences are of what's going on here, the mapping consultant was planning on doing a presentation on competitiveness today; as part of that presentation, they were going to present the various options for testing competitiveness that are -- that are available, and a lot of them are new from the last redistricting process.

Because we're going to have to wait to have that presentation until next week, the idea was to bring on presenters kind of similar to what we did today that was going to start next week as to those competitive folks. We can talk offline about whether or not we can basically stack those as part of next week's meeting, but there is a -- there is a chance that we'd be moving the schedule off a whole week due to what the circumstances.

So I just raise that as a -- as an issue, and we can try to figure it out.

COMMISSIONER LERNER: Chair Neuberg, I've been thinking that -- well, this may be difficult to do this for next week because everybody already had things scheduled, but we may want to be starting to think about having

meetings a couple of times a week anyway as we're starting to get into the mapping phase and all of this, so this is a good discussion.

Commissioner Mehl, are you free on Monday by any chance? Or is that whole week out besides Tuesday?

COMMISSIONER MEHL: The whole week is pretty messed up other than -- other than Tuesday.

I can certainly try to figure something out if there was just a -- you know, a few hours on a -- on a Monday or Wednesday, but -- I would do my best to try to figure it out.

CHAIRPERSON NEUBERG: Other Commissioners, what is your schedule next week Monday, Wednesday, Thursday?

VICE CHAIR WATCHMAN: Madam Chair, this is Vice Chair Watchman. Monday a few hours is good for me. Like Commissioner Mehl, the whole week is -- is scheduled for me, but Tuesday most the day works for me and part of Monday. But the rest of the week is not good for -- for me.

Wednesday, Thursday are tight. Very tight.

MR. JOHNSON: Commissioner Watchman, morning or afternoon on Monday?

VICE CHAIR WATCHMAN: Morning.

COMMISSIONER MEHL: If I need to figure out a way to do a couple hours on Monday morning, I will figure it out.

CHAIRPERSON NEUBERG: I -- is Commissioner York on? I don't believe he's live. I know he's having challenges.

Commissioner Lerner, Monday works for you as well?

COMMISSIONER LERNER: Yeah, Monday morning works fine as long as it works for the others who have already previously commitments and all that, it's fine for me.

MR. JOHNSON: And -- okay, so here is going to be our suggestion, if we could figure out the Monday morning, it will be great that we hopefully get Timmons, I know Ivy, I know Doug, if you want to come off mute, can do their presentation on Monday morning if that's going to work out; and then we'll work -- the legal counsel with work with the mapping consultants to try to identify a few of the more traditional competitive mechanisms, and see if we can't get some presenters for Tuesday on that.

Does that make sense?

MS. SAKANSKY: Yes, it makes sense. Doug is the one who is scheduling the presenters.

MR. JOHNSON: Right. So we can circle back on that and see what we can figure out.

COMMISSIONER LERNER: So the Monday meeting, because both -- for both Commissioner Mehl and Commissioner Watchman in particular for their schedules, what do you anticipate in terms of time?

MR. FLAHAN: Doug is available to 1:30 on Monday.

Sorry I was on mute when I said that.

Doug, do you know how long your competitiveness presentation would be?

MR. JOHNSON: The competitive -- can you hear me?

CHAIRPERSON NEUBERG: Yes.

MS. SAKANSKY: Yes.

MR. JOHNSON: Okay. It's probably 35 minutes or so or presentation and a discussion. But I -- the competitiveness presentation and training is probably the least of our priorities in terms of the things we have to cover, so that -- and -- and it's also the most flexible.

MR. FLAHAN: We do want to talk about schedule the next available meeting.

CHAIRPERSON NEUBERG: Yeah.

MR. FLAHAN: We see that as a critical path, so that might be more of a lengthy discussion if you guys have questions.

CHAIRPERSON NEUBERG: Maybe that could be -- if the Commissioners can reconvene on Monday morning, the 19th, that competitive presentation would -- I think would be helpful; and the timeline, I mean, most importantly for us to be able to dig in on -- on what you've presented in the attachment on the agenda, which we were prepared to discuss. I think that's of paramount importance.

MR. FLAHAN: And with extra time, if Brett is okay

with it, I can provide you the PowerPoint that I presented for this meeting, and I will give you guys a heads-up I created, an order of operations slide, that Commissioner Lerner would like, that we she can see what tasks needs to be done, what is waiting upon it, and what is the outcome.

CHAIRPERSON NEUBERG: Okay. So that -- so without Commissioner York here, it sounds like what may work for the majority of us is to plan for a meeting early Monday morning as well as our traditional meeting on Tuesday the 20th, and hopefully we can cover, you know, the material that we expected to up until the 20th; and we can work out the specific details, the -- the revised agenda, et cetera, we'll -- we'll have the time to -- to provide the public, you know, notice with what the -- the work agenda will be.

COMMISSIONER LERNER: And, Commissioner Mehl, do you have a time limit then on that Monday? Because I know it wasn't scheduled.

COMMISSIONER MEHL: No, I'll -- I'll make whatever we need to make work. I'll just -- I'll juggle some things around.

MR. JOHNSON: So planning to start at 8 o'clock on Monday?

VICE CHAIR WATCHMAN: Okay.

CHAIRPERSON NEUBERG: I think that sounds like a

plan.

In terms of legal and, you know, notifying the public, I presume that will be done in our agenda which we will work and get out to the public; but in terms of a tentative schedule, it sounds like we're looking at Monday at 8:00 a.m. and Tuesday at 8:00 a.m.

MR. JOHNSON: That sounds good.

CHAIRPERSON NEUBERG: And hopefully by the end of Tuesday, the 20th, we will be at the same point that we would have otherwise been, you know, had this unfortunate outage not happened.

Okay. So, with that, how do I legally and what do I need to do in terms of public comments and adjournment?

MR. JOHNSON: Yeah, you're going to adjourn -- adjourn the meeting and order a reconvening of this meeting for Monday morning at 8:00 a.m.

So it's going to reconvening -- it's technically a reconvening of this meeting.

CHAIRPERSON NEUBERG: And so what does that mean for public comments, because they're still open?

MR. JOHNSON: Yeah, I mean -- I mean --

MR. HERRERA: Here they just close it out.

MR. JOHNSON: Yeah, and reopen it.

MR. HERRERA: Reopen it and just have to issue a statement or public.

MR. JOHNSON: We'll figure that out. But close out public comment here and to be reopened at 8:00 a.m. on Monday morning.

CHAIRPERSON NEUBERG: Okay. And we're -- so that's okay to close public comments?

MR. JOHNSON: That's right. Because technically you're out of session and you're on -- you're on a long break, let's put it that way.

CHAIRPERSON NEUBERG: Okay. All right. Any other conversation?

Okay. With that team, public, you know, our apologies, it's frustrating to all of us, but we are going to close public comments given the upcoming adjournment of this meeting.

Please note members of the Commission may not discuss items that are not specifically identified on the agenda. Therefore, pursuant to A.R.S. 38-431.01(H), action taken as a result of public comment will be limited to directing staff to study the matter, responding to criticism, or scheduling the matter for further consideration and decision at a later date.

With that, I'm going to recommend that we adjourn our meeting for today and reconvene Monday morning the 19th at 8:00 a.m.

I will entertain a motion if I need one.

COMMISSIONER MEHL: So moved.

VICE CHAIR WATCHMAN: So second. I'll second.

COMMISSIONER MEHL: Commissioner Mehl makes a motion.

CHAIRPERSON NEUBERG: Commissioner Mehl has a motion to adjourn; Vice Chair Watchman seconds. Any further discussion?

With that, I will take a vote to adjourn.

Vice Chair Watchman.

VICE CHAIR WATCHMAN: Aye.

CHAIRPERSON NEUBERG: Commissioner Mehl.

COMMISSIONER MEHL: Aye.

CHAIRPERSON NEUBERG: Commissioner Lerner.

COMMISSIONER LERNER: Aye.

CHAIRPERSON NEUBERG: Commissioner York I believe is not on there.

MR. JOHNSON: He's absent.

CHAIRPERSON NEUBERG: Yes.

And Commissioner Neuberg an aye.

With that 4-0, we adjourn, and we look forward to reconvening Monday morning at 8:00 a.m.

And, again, our sincere apologies to the team and to the public. It was absolutely something completely out of our control. So we will see you next week.

MR. JOHNSON: Thank you.

(Whereupon the proceeding concludes at 11:20 a.m.)

* * *

C E R T I F I C A T E

STATE OF ARIZONA)
) ss.
COUNTY OF MARICOPA)

BE IT KNOWN that the foregoing proceedings were taken before me, Angela Furniss Miller, Certified Reporter No. 50127, all done to the best of my skill and ability; that the proceedings were taken down by me in shorthand and thereafter reduced to print under my direction.

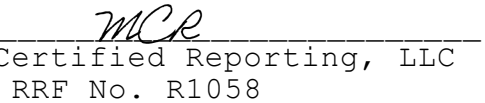
I CERTIFY that I am in no way related to any of the parties hereto nor am I in any way interested in the outcome thereof.

I FURTHER CERTIFY that I have complied with the requirements set forth in ACJA 7-206. Dated at Litchfield Park, Arizona, this 28th of July, 2021.

Angela Furniss Miller, RPR, CR
CERTIFIED REPORTER (AZ50127)

* * *

I CERTIFY that Miller Certified Reporting, LLC, has complied with the requirements set forth in ACJA 7-201 and 7-206. Dated at LITCHFIELD PARK, Arizona, this 28th of July, 2021.



Miller Certified Reporting, LLC
Arizona RRF No. R1058