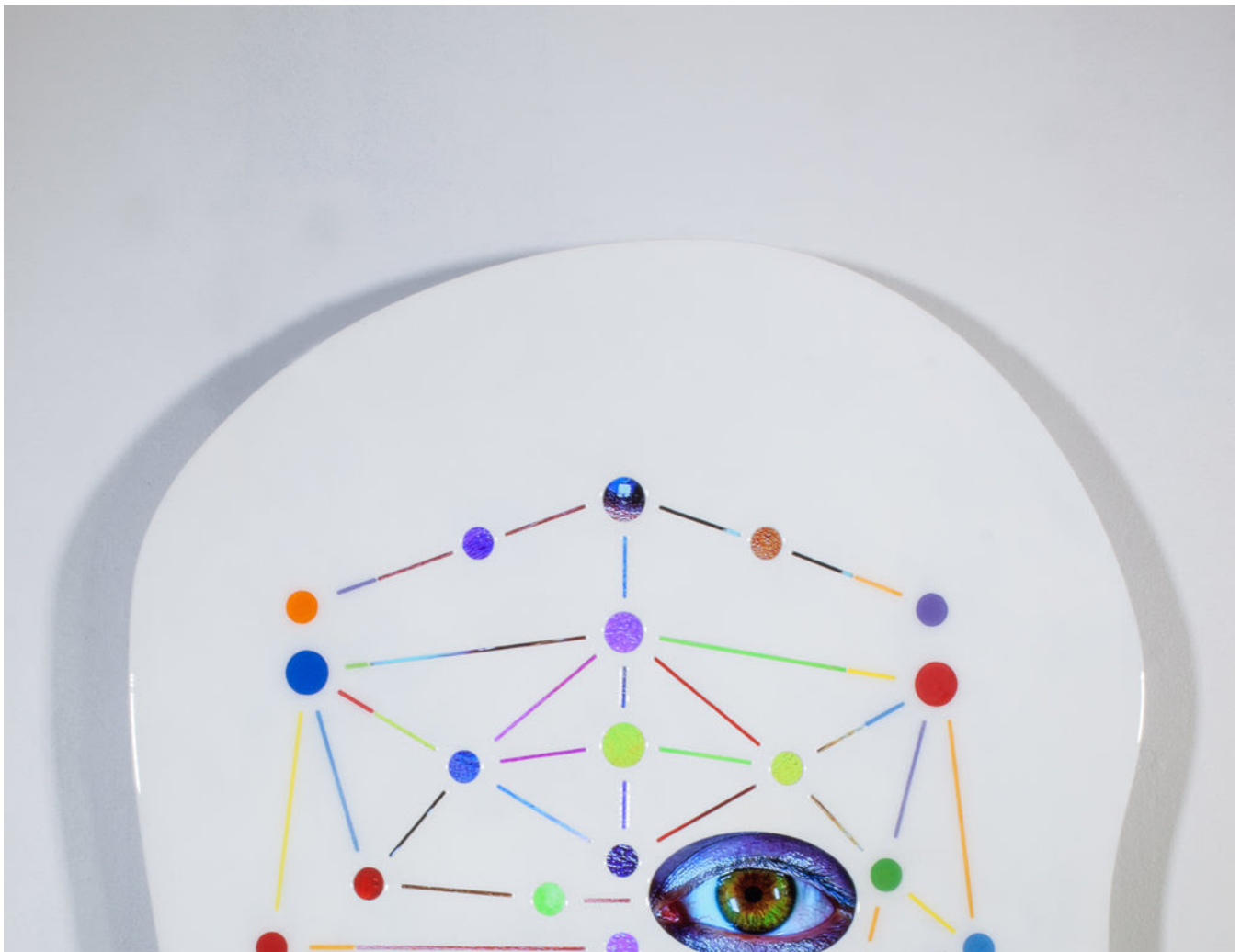# Counting the Countless

Why data science is a profound threat for queer people

Os Keyes
April 08, 2019
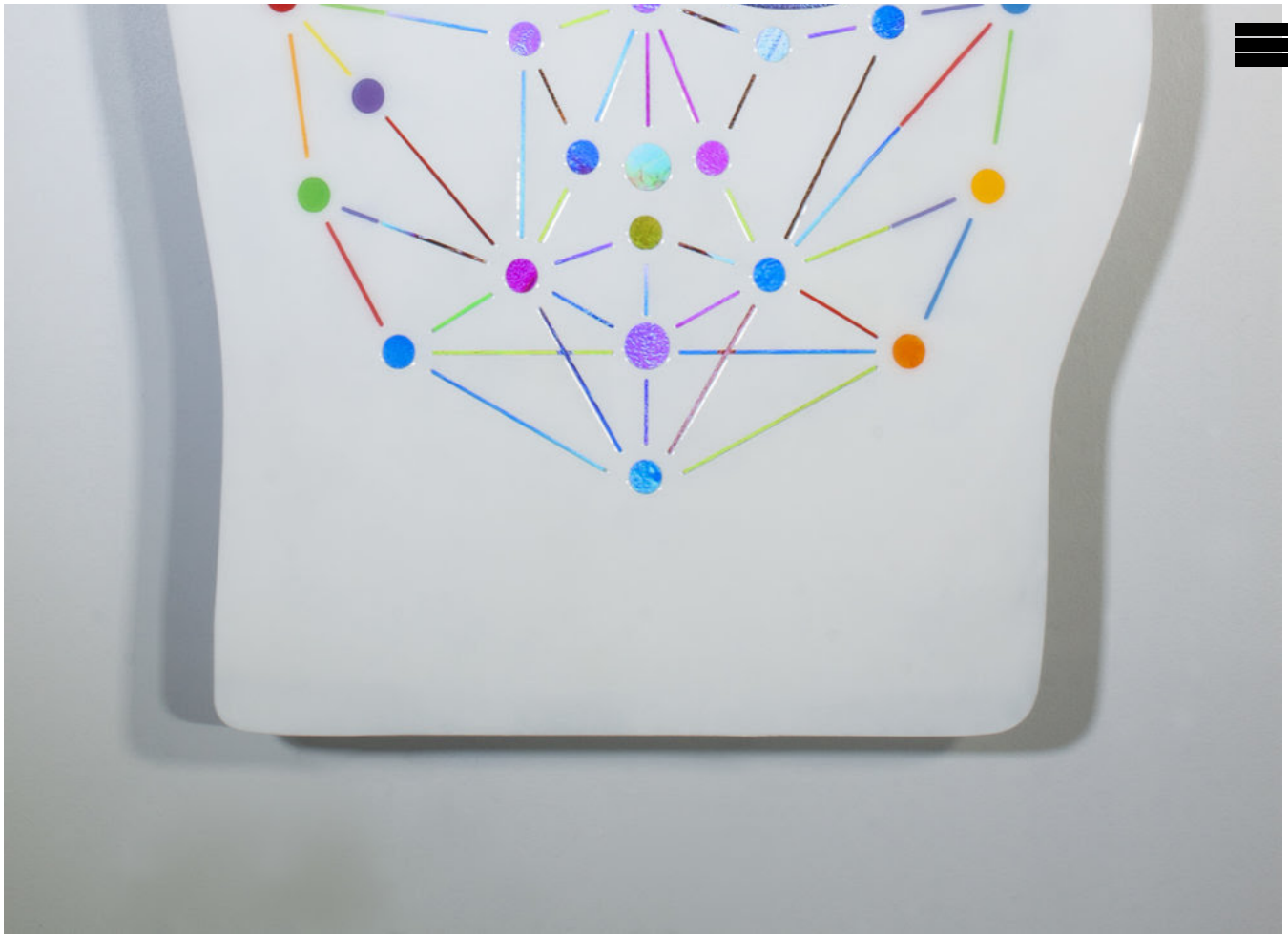
Image: From *’pU#*c’* by <u>Tony Oursler</u>. © Tony Oursler. Courtesy the artist and Lisson Gallery.

This piece <u>originated as a talk at Seattle University</u>. Well, really it originated as a question from the <u>Seattle Non-Binary Collective</u>, a group of Seattle residents and visitors with genders beyond the binary of "man" and "woman." They reached out and asked, "We hear you're a data scientist. Could you do a talk on how trans and/or nonbinary people can get involved in data science?"

I replied, "Well, to be perfectly honest, I think data science is a profound threat to queer existences." And then for some reason they stopped replying! Who can say why? But it prompted me to reflect on how unexpected my position probably is — how, in a world where data science is often talked of as the future, as a path to a utopian society of perfect understanding, saying that you think data science is a path to cultural genocide makes people doubletake. So I figured I'd take the opportunity to explain *why* that is my position.

Trans lives are ultimately (to a certain degree) about autonomy: about the freedom to set one's own path. Society isn't a tremendous fan of this

Before we get to data science, let's start with some definitions. What does it mean to say someone is queer? What does it mean to say someone is trans? In both cases, there really *isn't* a fixed definition that holds everywhere. Trans identity is contextual and fluid; it is also autonomous. There's no test that you give someone to determine they're "actually" trans, unless you're a doctor, or a neuroscience researcher, or a bigot (but I repeat myself).

This lack of clear definition is important to trans existences and experiences, because trans lives are ultimately (to a certain degree) about autonomy: about the freedom to set one's own path. Society isn't a tremendous fan of this, resulting in the *one* almost-constant of trans lives: They are often miserable, most visibly evidenced by the drastically higher rate of suicide and self-harm that trans people experience. Not because we're trans, but because we exist under what bell hooks refers to as a "white supremacist capitalist patriarchy." That is, we live in an environment that is fundamentally racist, fundamentally built around capitalism, fundamentally based on rigid and oppressive gender roles. And as trans people, we suffer under all these facets, both collectively and individually. Trans people of color experience racism and transphobia, the latter induced by rigid patriarchal norms. Poor trans people — which is a lot of us, given how poverty correlates with social ostracization — suffer under both transphobia and capitalism. Those of us who are disabled and so don't fit norms of a "productive" worker experience that poverty tenfold.

These norms and forms of harm do not exist "just because": They exist as a self-reinforcing system in which we are coerced to fit the mold of what people "should be." Those who can are pressured to; those who can't or refuse are

punished. Dean Spade, an amazing thinker on trans issues and the law, has coined the term "administrative violence" to refer to the way that administrative systems such as the law — run by the state, that white supremacist capitalist patriarchy — "create narrow categories of gender and force people into them in order to get their basic needs met."

Let's look at a common example of what "administrative violence" looks like and what it does. Suppose you want to update the name and gender associated with your mobile phone: You go in, and the company says that you need a legal ID that matches the new name and gender. So you go off to the government and say, Hey, can I have a new ID? And they say, Well, only if you're *officially trans*. So you go off to a doctor and say, Hey, can I have a letter confirming I'm trans? And the doctor says, Well, you need symptoms X, Y, and Z. And then when you do this, and jump through all that gatekeeping, *everything breaks* because suddenly the name your bank account is associated with no longer exists. You attempted to conform, and you still got screwed.

This example demonstrates how administrative violence reinforces the gender binary (good luck getting an ID that doesn't have "male" or "female" on it) and the medicalized model of trans lives — the idea that being trans is a thing that is *diagnosed,* should one meet certain, medically decided criteria. It communicates that gender is *not* contextual, that you can only be one thing, everywhere. It enables control and surveillance, because now, even aside from all the rigid gatekeeping, a load of people have a note somewhere *in their official records* that you're trans.

So could we reform this? Spade would say no — the rigid maintenance of hierarchy and norms is really what the state is *for*. Moreover, attempts to reform this often leave the most marginalized among us out and those of us who are multiply marginalized the least listened to. Efforts to expand the two-category gender system in IDs to contain more options are great, *if you can afford a new ID,* and if you can afford to come to the attention to the state — and in some ways they're counterproductive even then, legitimizing the idea of

gender as a fixed, universal attribute that the state ultimately gets to determine. We are attempting to negotiate with a system that is fundamentally out to constrain us.

○ ○ ○

Defining Data Science

So: trans existences are built around fluidity, contextuality, and autonomy, and administrative systems are fundamentally opposed to that. Attempts to negotiate and compromise with those systems (and the state that oversees them) tend to just legitimize the state, while leaving the most vulnerable among us out in the cold. This is important to keep in mind as we veer toward data science, because in many respects data science can be seen as an extension of those administrative logics: It's gussied-up statistics, after all — the "science of the state."

We are attempting to negotiate with a system that is fundamentally out to constrain us

But before we get too far ahead of ourselves: What is data science, anyway? There are a lot of definitions, but the one I quite like is: *The quantitative analysis of large amounts of data for the purpose of decision-making.*

There's a lot that's packed into that, so let's break it down. First: *quantitative analysis*. A field based on quantitative analysis — on *numbers* — raises a lot of questions. For example: What can be counted? Who can be counted? If we've decided to take a quantitative approach to the universe, then by definition we have to exclude any factors or variables that can't be neatly tidied into numbers, and we have to constrain and standardize those that can to make sure

tidying them is convenient.

Suppose you decide to do a survey in which you ask participants to provide their gender, and to be inclusive you have "gender" represented by a free-text box in which people can put whatever identity or definition they want. Great! Except: If you're taking a quantitative approach to analysis, wanting to look at the distribution between answers or how answers correlate with other variables, you're going to have to standardize or normalize them in some way. You're going to have to determine that "bigender" and "nonbinary" should be in the same bucket (or shouldn't). You're going to have to work out whether you treat trans women and trans men separately, whether you clump them with "women" and "men" respectively, what values mean "trans woman" in the first place — you are going to lose definition. You are going to have to make judgment calls on where the similarities are and what is (and is not) equivalent. You have to, because quantitative methods work on the assumption that you'll have neatly distinguishable buckets of values.

Second: *large amounts of data* — that "big data" you've likely heard about. A data-science approach encourages the collection of as much data as possible (all the better to measure you with). This data is vast across time: We should have as much of your history as possible. It's vast across space: We should be able to measure as much of the world as possible, in the tidied, standardized way described above. It's also vast across subjects: We should be able to measure as many *people* as possible, in this tidied, standardized way. Our data should be collected ubiquitously, it should be collected consistently and perpetually, and any variation that complicates the data collection should be eliminated.

The ideal data-science system is one that's optimized to capture and consume as much of the world — and as much of your life — as possible. Why? Because of the final part of that data science definition: *for the purpose of decision-making.* This is the bit that proponents of data science really seem to drool over. We can use datalogical systems for efficiency gains and for

consistency gains; we can remove that fallible, inconsistent "human factor" in how we make decisions, working more consistently and a million times faster. Which is, you know, fine, sort of. But by definition, a removal of humanity makes a system, well, inhumane!

So perhaps a more accurate definition of data science would be: *The inhumane reduction of humanity down to what can be counted*.

This redefinition resonates with Spade's work, and with that of Anna Lauren Hoffmann, who has coined the term "data violence." Just as "administrative violence" refers to violence perpetuated through administrative systems, "data violence" refers to the perpetuation of violence through datalogical systems: everything from YouTube's recommender algorithm to facial recognition to online advertising.

But data violence differs from administrative violence in a few ways: the ubiquity at which it operates (the state is not the only entity that can perpetuate data violence, as the examples suggest), the scale at which it operates, and the fluidity of it. Ubiquitous data systems, routinely treated by business executives, governments, and war criminals as the future, can capture a lot more of your life than the DMV. This is not accidental; the entire appeal and *point* of data science is ubiquitous collection.

Data science is premised on using its own view to control and standardize the paths our lives can take, and it responds to critique only by expanding the degree to which it surveils

Returning to the example of name changes: the elaborate process of jumping through hoops with the phone company, the government, the doctor, the bank,

and so on means that when you resolve all that, you out yourself and mark yourself forever. Everyone and their pet dog has a record that you're trans; every data system you have to interact with is oriented to keep that known for as long as possible in case it becomes relevant to their model.

And even if it somehow doesn't, a background-check website can take that old data — the phone number, your deadname — and make it available, ensuring that you're outed every time someone googles your number. Your administrative transition may be a boon to you, but it's also a boon to the vast number of systems dedicated to tracking the course of your life, for their own not necessarily benevolent purposes. And by the standardizing logic of data science, this tracking is going to be reductive: It will track only the things that matter to those building the systems, in the ways that are acceptable to them, punishing you when you deviate.

To take one example: Some health insurance companies are now tracking the food you buy. If you eat healthily, you get lower premiums; if you eat unhealthily, you get higher premiums. Let's set aside for a second the obvious problems with quantifying "food healthiness." You're still tracking food purchases through phone apps and store purchases. Who gets tracked, and who doesn't? Presumably if I go to Trader Joe's, the system is all neatly and nicely integrated, and my health insurance company gets all my data. But I don't go to Trader Joe's: I live in the Central District of Seattle, and there's precisely fuck-all there except an Ezel's fried chicken, where I pig out once a month, and a bodega, where I do my shopping. If that bodega isn't integrated with my insurance company's new fancy data science system for determining premiums, then according to their systems, I subsist off fried chicken approximately once a month, and nothing else. I'm guessing my premiums are going to be pretty high!

As with administrative violence, we have to ask, Who does this harm the most? The people caught in this trap are disproportionately likely to be already marginalized, already marked — the poor, immigrants, people of color. The

system's integration with things like apps for discounts and coupons renders people without smartphones invisible. Faced with such critiques, those running such programs are likely to respond, *You're right! We should make sure there's a sensor in the bodega too!*

So the long and the short of it is that, as currently constituted, data science is fundamentally premised on taking a reductive view of humanity and using that view to control and standardize the paths our lives can take, and it responds to critique only by *expanding the degree to which it surveils us.*

Now, I don't know about you, but my idea of a solution to being othered by ubiquitous tracking is not "track me better."

◑ ◑ ◑

Reforming Data Science?

So can we reform these systems? Tinker with the variables, the accountability mechanisms, make them humane? I'd argue no. With administrative violence, Spade notes how "reform" often benefits only the least marginalized while legitimizing the system and giving cover for it to continue its violence. The same is true of datalogical violence, as can be seen in attempts to reform facial recognition systems.

In 2018, Joy Buolamwini, a researcher at MIT, published an intersectional analysis of gender-recognition algorithms: facial-recognition systems that identify gender and use them for decisions from the small (demographic analytics) to the vast (who can access bathrooms). Buolamwini found that these systems — which, she notes, are premised on an essentialist, biological view of gender — are biased against dark-skinned, female-coded people. Her recommendation, developed in this subsequent paper, was that the people building facial-recognition systems should use more diverse datasets. In other words, a fundamentally reformist approach.

There are two problems with this. The first is that gender-recognition systems are fundamentally controlling and dangerous to trans people and simply cannot be reformed to not be violent against us. Normalizing, reductive views of gender are *what these systems are for*. Curiously, the second paper — the one about how great expanding the diversity of data is — did not mention the constrained view of gender these systems take. The other problem is that incorporating more people into facial-recognition systems isn't really a good thing even for them. As Zoé Samudzi notes, facial-recognition systems are designed for control, primarily by law enforcement. That they do not recognize black people too well is not the problem with them, and efforts to improve them by making them more efficient at recognizing black people really just increases the efficacy of a dragnet aimed at people already targeted by law enforcement for harassment, violence, and other harms.

So this reformist approach to facial recognition — making the system more "inclusive" — will not really reduce harm for the people actually harmed. This control and normalization is part of the point of data science. It is required for data science's logics to work. All reform-based approaches do is make violent systems *more efficiently violent*, under the guise of ethics and inclusion.

◖ ◖ ◖

Radical Data Science

To summarize: Data science as currently constituted

- provides new tools for state and corporate control and surveillance
- discursively (and recursively) demands more participation in those tools when it fails
- communicates through its control universalized views of what humans can be and locks us into those views.

Those don't sound compatible with queerness to me. Quite the opposite: They sound like a framework that fundamentally results in the *elimination* of

queerness — the destruction of autonomy, contextuality, and fluidity, all of which make us what we are and are often necessary to keep us safe.

If you're a trans person or otherwise queer person interested in data science, I'm not saying, "Don't become a data scientist under any circumstances." I'm not your mother, and I get that people need to eat to survive. I'm just explaining why I refuse to teach or train people to be data scientists and why I think reformist approaches to data science are insufficient and that co-option *into* data science, even to fix the system, is fundamentally inimical unless your primary question is asking who is left out of those fixes. You need to make the decision that is right for your ethics of care.

For me, my ethics of care says that we should be working for a radical data science: a data science that is *not* controlling, eliminationist, assimilatory. A data science premised on *enabling autonomous control of data*, on enabling *plural ways of being*. A data science that *preserves context* and *does not punish* those who do not participate in the system.

How we get there is a thing I'm still working out. But what you can do right now is build counterpower: alternate ways of being, living, and knowing. You can refuse participation in these systems whenever possible to undercut their legitimacy. And you can remember that you are not the consumer but the consumed. You can choose to never forget that the harm these systems do is *part of the point*. ◑

Os Keyes is a PhD student at the University of Washington and an inaugural Ada Lovelace Fellow who studies gender, data, technology, and control.