I’d like to begin by speaking about Spotify’s API output.

<What is API output?>

My work on Spotify began when I had the reaction that almost every musician would have confronted with this kind of musical analysis; it’s oversimplified.

And it’s also sometimes wrong, even where you’d expect it to be pretty good. <CLICK> For example, the system analyses Ginuwine’s Pony as being in Ab minor, at 142 bpm. These metrics – tempo, key – are the exactly the kinds of content-based tasks at which machine listening is supposed to excel.

<SLIDE>

But if we compare it with an analysis done by a human (me), showing the “true” key and tempo, we see that both are pretty off.

<SLIDE>

To say nothing of the other elements of this analysis, which, at a minimum, raise a number of alluring musicological questions. All of which to say, the motivation behind this work is, ultimately, the difference between human and machine representations of musical meaning. It is quite tempting for musicologists to point out the many ways in which, on its very face, Spotify seems to do violence to the kind of musical complexity musicians have given so much of our lives to cultivating.

<SLIDE>

Which leads me to what I’ve come to think of as the “paranoid hypothesis” about Spotify – namely, that, relying on such reductions as those I just hinted at, it will tend to have a homogenizing effect on musical listening and perhaps even musical production.

<SLIDE>

Here is Pelle Snickars, a professor of Media Studies at Umea University in Sweden, giving voice to that hypothesis.

<READ QUOTE AND DISCUSS>

<SLIDE>

Snickars is part of a research team that has written what I think is the best book on this subject, Spotify Teardown, published last year by the MIT Press. This is the only work I know of that interrogates Spotify not just as a force in the music industry, but as a case of platform capitalism, one that participates in broader trends toward corporate surveillance, financialization, and affect management. A company, in other words, not particularly different from Facebook, to which Spotify has historically had close ties.

<SLIDE>

The natural question, then, for an investigator interested in substantiating the paranoid hypothesis is, "how does Spotify make its recommendations?" It's a question that's worth asking, but it's difficult. First, the degree to which Spotify has become a music *discovery* company – the sheer number of different *kinds* of recommendations it makes to the user, from autocomplete in search to the Radio function to the Discover Weekly, makes it hard to know what exactly qualifies a "recommendation," and where the critique should focus. Second, although they do make some of it public, Spotify's algorithm is a carefully guarded industry secret protected by non disclosure agreements.

That "recommendation algorithm," moreover, is probably not even the *kind* of thing that would admit of transparent, straightforward understanding in the first place. It is too complex, too frequently updated, and too internally variegated to be "knowable" in any simple way. As Nick Seaver, an anthropologist who focuses on the music recommendation world, writes,

> There is not one playlisting algorithm, but five, and depending on how a user interacts with the system, her radio station is assigned to one of the five master algorithms, each of which uses a different logic to select music.

<SLIDE>

Nevertheless, it's possible to make an educated guess, and, if you're careful, to draw certain conclusions from those guesses. Here are three strategies I am reasonably confident operate in Spotify's recommendations today – or that might soon operate. The first phrase is taken from a group authored paper from a handful of Spotify employees about recommender systems that use reinforcement learning to incorporate "explanations" with their recommended content.

However, since the recommendation style described in that paper – termed "recsplanations" – seems not to have been implemented yet as far as I can see in Spotify, this strategy can be regarded as a future direction, and as a valuable indicator of just how much behavioral information is already logged by Spotify every day.

The second strategy is machine listening – tools that analyze actual audio content, producing the kind of analysis I showed at the beginning of this presentation. In 2014, Spotify acquired the Echo Nest, an important music information company whose software uses machine learning to join various kinds of "cultural metadata" (think of reviews, blog posts, lyrics, etc) to the audio signal. There's a lot to say about the philosophical orientation such a strategy implies, but I'm not going to get into that today.

The third strategy is the one I am focused on today: collaborative filtering, in particular the technique known as matrix factorization. This is a technique for using observed user behavior – like numbers of streams per song for a given user – to model users' taste, and, thereby, to make recommendations. The easiest way to conceptualize it is: if I am a recommender system, and I know that your friend liked a given track, and I know that you are somehow similar to your friend, then I'm going to recommend that track to you. It's a standard part of the recommendation systems toolkit, especially since the $1million Netflix Prize made it famous in 2009.

<SLIDE>

Collaborative filtering doesn't incorporate information from the audio signal, and in that regard it's a little musically naive. Nevertheless, collaborative filtering still does seem to be a major part of how

Spotify works, as this slide from a 2015 presentation by Spotify Recommender Systems engineer Andy Sloane, shows.

<SLIDE>

Here are some standard images representing how matrix factorization works, taken from Andy Sloane again.

<EXPLAIN MATRIX FACTORIZATION, both sides>

The real meat of the machine learning discipline is how best to arrive at that factorization. There are many techniques for finding it, each tailored to a given task. But I'm not interested in the technical details of different recommendation systems, except where they bear upon musicological and philosophical concerns. And the part of this system that is most philosophically provocative is the use of latent factors.

<SLIDE>

In this large matrix, there will be statistical patterns – that's because you are likely to enjoy things that people like you enjoy. This system aims to capture those statistical correlations. To do that, it postulates these latent factors – variables that, while they are not directly observed the way user plays are, are nevertheless postulated as explanations for the statistical correlations in the actual observed user behavior.

The factorization performed in this slide relies not only on ingenious mathematical transformations, but on the supposition of a numerically measurable, though unobservable cause (or, "factor") for the statistical correlations under consideration.

There are two crucial features of latent factors that I want to note here:

1) The number of latent factors in a given factorization is a matter of choice for the analyst. Could be 2, could be 40. For Spotify, it seems to be about 40.

2) They don't need to be interpreted in order to be useful. That is, factor one might be something like "conservativeness," and every track and every user would then have some rating along that axis – but there is nothing in the math that suggests that interpretation, or any other. The interpretation of latent factors, in other words, is more of an art than a science, and, since it has no bearing on the quality of recommendations, it is not one that is often practiced in this world.

These latent factors beg ontological questions: in what sense, if any, are we committed to the notion that they are "real?" What I'm arguing in this talk is that, while this question is familiar territory from the philosophy of science, it goes largely un-asked in recommender systems research. I also think that, while the answers to these epistemological questions are surely far from simple, they ought to matter to us, to the extent that musical meaning matters to us at all. I also believe that, in the case of recommender systems, there are straightforward economic reasons why such questions are not even attempted.

<SLIDE>

It is useful to point out that factor analysis has its roots in psychology and psychometrics. Charles Spearman, an English psychologist born in 1863, was one of the first to use the concept of a latent, unobserved factor to reduce the dimensionality of his data. He did this in two seminal papers of 1904: one charts the statistical techniques themselves, while the other applies them, in an effort to characterize human intelligence. Spearman arrives at a 2-factor explanation, with the “g” factor standing in for general intelligence. G-theory, and derivations, still have a lot of currency today, in popular and academic circles – in a sense *g* is what one attempts to measure with a general IQ test. Whether g theory is good or bad, it is clear that for Spearman, the factor is a truly “contributing” force, not a mere statistical expedient.

> Our true correlation in no way deserves the reproach of being a theoretical abstraction, for it only represents the limit to which the observed correlation itself will continually approach.

<SLIDE>

Factor analysis was also applied to personality, notably in the work of the Louis Leon Thurstone. His choice to use five factors to characterize human personalities eventually gave us, toward the end of the 20th century, the so-called “big five” personality traits that are popular in corporate organizational strategy, marketing, and popular psychology.

<SLIDE>

The big five have, in turn, become a basic part of how advertisers, corporate structures, and political operations comprehend and control us. This slide is taken from the site of the late Cambridge Analytica. Here, they are explaining that they use the science of personality to maximize the power of whatever campaign they’re involved with.

So there is a sinister side to latent factor theory in pyschometrics and personality theory. To return to the paranoid hypothesis, we might ask whether we would say the same for its cousin in music recommendation. In other words, is Spotify a sort of musical Cambridge Analytica, as indifferent to epistemology and as crassly mercenary? By way of an answer, I’ll list a handful of things I think the Spotify system seems to presuppose.

<SLIDE>

1. That the “ground truth” data is reliable – in particular, that there is no “algorithmic confounding” effect. That is, Spotify assumes that the information in that first sparse matrix is not already too saturated by constant recommendations, in and outside of the Spotify universe, to be a true indicator of user preferences. <CLICK>
2. Second. In subjecting aesthetic experience to the same predictive logic that derives from intelligence tests and market research, Spotify assumes that there is nothing special about it; that as a consumer of music you’re in some sense no different from a consumer of toilet paper, or a voter waiting to be swayed.<CLICK>
3. That it is OK to make predictions without a serious epistemological and ontological consideration of the latent factors involved. In psychology, thinkers are careful about what kind of ontology they are committed to. Today’s recommender systems research, because it is so mathematically specialized, and because it is so baldly market oriented, manages to avoid such self-reflexive critique. But what, if anything, does it really mean, musically speaking, for a user to be represented as a vector in 40-dimensional latent musical factor space? And should we trust a system that doesn’t feel the need to answer that epistemological question?<CLICK>

4. Fourth, that good measures of success exist at all – and that they're probably just low root mean square errors scores and high user retention – especially high user retention.

This last point is probably the most instructive. It seems safe to say that Spotify, in the end, will care about one thing only, the same thing any corporation cares about. We should, then, probably regard Spotify's professed altruism the same way that we regard Google's mission to "organize the world's information" or Facebook's to "bring the world closer together." Slogans are usually mendacious, of course, but the enduring emotional power of music, combined with the novelty of the tech branding and the cultural prestige attaching to machine learning, seem to combine to make that easy to forget in the case of Spotify.

As a way of re-posing the paranoid hypothesis, I'll conclude by briefly comparing these two companies.

<SLIDE>

Both rely on unobservable variables, and neither one needs to concern itself very much with the question of what those factors mean, if they mean anything. It's enough that they seem to work in some way.

<SLIDE>

Both rely upon extensive user surveillance, which in recommender systems is euphemized as "implicit feedback."

<SLIDE>

Both have integrated with Facebook in ways that are pretty objectionable, from a privacy standpoint.

<SLIDE>

This is a slide from a 2018 article about how that data sharing went beyond what was normal, necessary, or moral.

<SLIDE>

Both are tools, ultimately, that influence our behavior. In the case of Cambridge Analytica, to maximize the power of political propaganda, in the case of Spotify, to keep us actively engaged for as long as possible.

<SLIDE>

And both, of course, are there in the end simply to make money. If that seems like an obvious point, it's one that I think is worth making – ultimately, the most compelling version of the paranoid hypothesis may be simply to remind us that Spotify, like so many other tech platforms, offers an addictive product, without any coherent way of demonstrating "success" or any critical attention to its scientific underpinnings. It logs lots of data about us and integrates with other similar platforms, all in the name of scaling an operation that will, hopefully, one day, become profitable.

These facts are are easy to miss in the “entrpreneurial haze” surrounding machine learning, to borrow a phrase from the early critic of AI Hubert Dreyfus. Dreyfus was writing in the early days of AI research at the Rand Corporation, when decent chess playing and simple language processing tasks were impressive enough to convince many that an automated form of general reasoning was within reach. Musical meaning is, possibly, a different kind of problem, one that first wave AI researchers probably wouldn’t have even recognized as a legitimate application. But in both cases – the 1970s euphoria around AI and today’s strikingly similar vogue for machine learning – it is important to make the distinction between a product and an insight. It’s possible that Spotify has some of both – but it will never be in its interests to help us make the distinction.

<SLIDE>

Thank you.