Algorithmic Effects on the Diversity of Consumption on Spotify

Ashton Anderson

University of Toronto Spotify



Ashton Anderson



Lucas Maystre

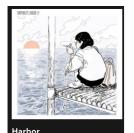


Ian Anderson









Tomppabeats















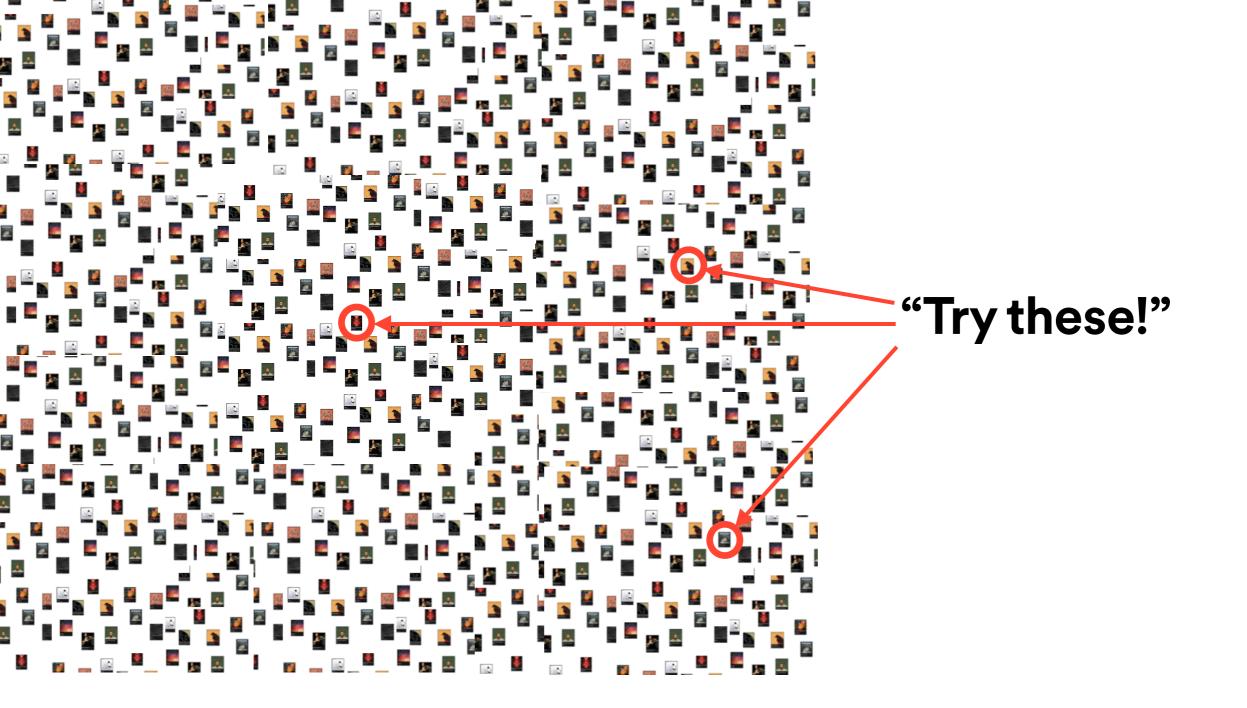




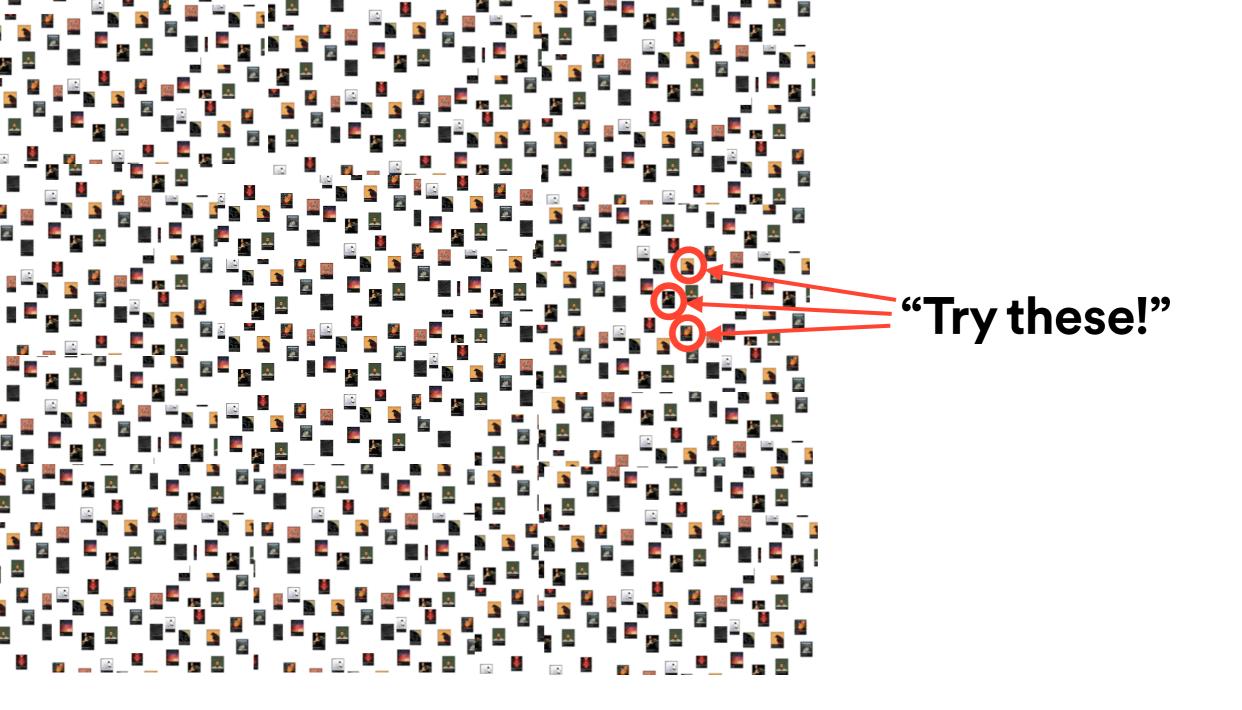
lent Shout (Audio/Vis

Our Endless Numbered Days Iron & Wine

Online platforms present users with a universe of content to choose from



Recommendation algorithms influence which items users consume



But there is concern that recommendation algorithms concentrate on overly narrow sets of content

[1] Eli Pariser. 2011. The filter bubble: What the Internet is hiding from you. Penguin, UK.

[2] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. Science 348, 6239 (2015), 1130–1132.
 [3] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. Public opinion guarterly 80, S1 (2016), 298–320.

Blue user: diverse consumption

Red user: narrow consumption

Consumption diversity is potentially beneficial for many reasons:

- → Exploring more of what the platform has to offer
- → Satisfying more needs on the platform

→ In many domains, diverse exposure is considered a virtue (news, information, etc.)

What is the association between algorithmic recommendation and the diversity of content users consume...

...and on the user experience in turn?

Data

We study consumption diversity on Spotify:



On Spotify, users can listen to over 50 million different songs on various digital devices.

There are free and premium versions of Spotify, with the free version being ad-supported and the premium version being subscription-based.

Main dataset:

- → Listening history of >100 million premium users during July 1–28, 2019
- → This comprises 70 billion "streams" (an instance of a user listening to a song)
- → For each user, we calculate their total activity and consumption diversity

How should we measure consumption diversity?

What do we want from a measure of consumption diversity?

Captures breadth: diverse consumption means consuming from across the spectrum of content on the platform.

→ Diversity captures the extent to which consumed items are meaningfully different from each other.

Consistency: similarities and differences between items should be mutually consistent.

→ Measure of diversity should be comparable between users.

Scalability: should be efficiently computable for millions of items and millions of users.

→ Applicable to real-world online platforms.

Existing approaches:



Problem: both entropy and Gini fail to capture breadth in a way that is sensitive to how **similar** the consumed items are

Compare:10x "Let It Be" — The Beatles10x "Blackbird" — The Beatles10x "Imagine" — John Lennon

10x "Angel of Death" — Slayer
vs. 10x "Only Time" — Enya
10x "So What" — Miles Davis

Entropy and Gini would consider these two users equally "diverse"!

To consistently and scalably capture similarities between songs, we use song embeddings.



Treat user playlists as "documents" and songs as "words" and run word2vec



Songs are arranged in a high-dimensional space such that "similar" songs are close to each other in the space.

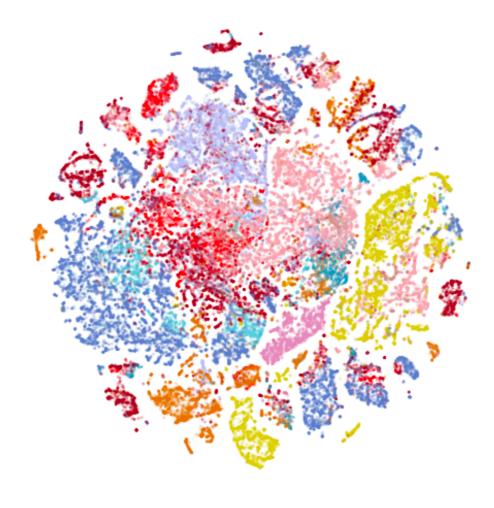
What does "similar" mean?

We define similarity empirically: the more often two songs appear together in user playlists, the more similar they are

Our song embeddings are:

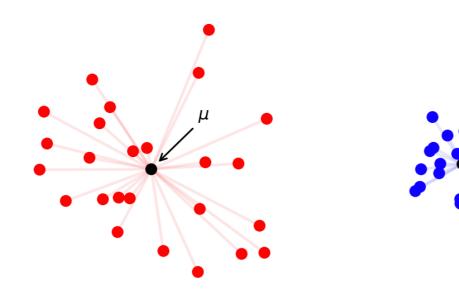
- \rightarrow 40-dimensional
- → trained on 850 million playlists
- → comprise millions of songs

Song embeddings give us a way to consistently and scalably compute similarities between any pair of songs



With a proper way of measuring arbitrary song similarities in place, we can now measure the diversity of a user's consumption.

Intuition: the more "spread out" a user's songs are in the embedding space, the more diverse their consumption





Diverse listening

Narrow listening

The GS-Score

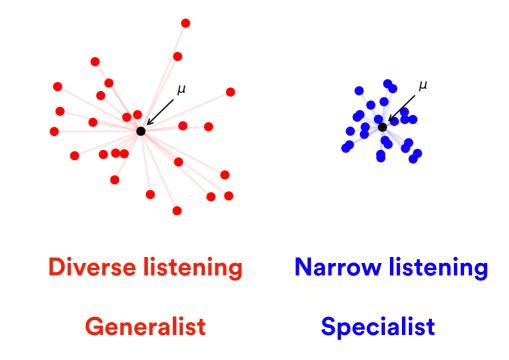
We use a recently proposed definition of diversity based on embedding spaces

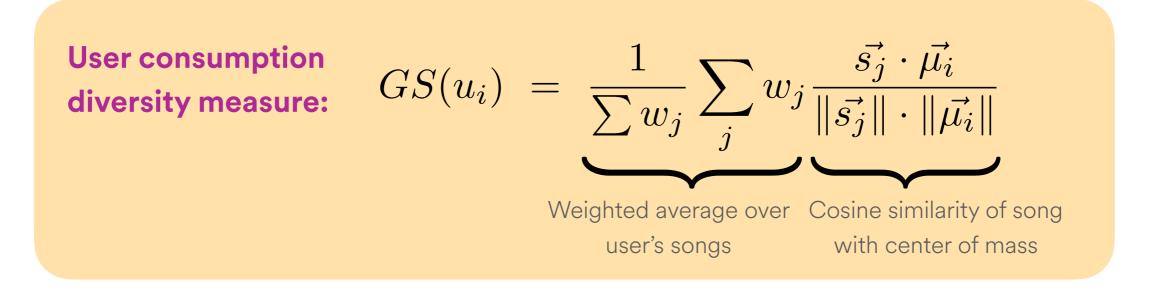
The **Generalist-Specialist (GS) score**: average cosine similarity between a consumed item and the user's center of mass

Say user i listens to song j w_j times, then:

Center of mass (simple mean):

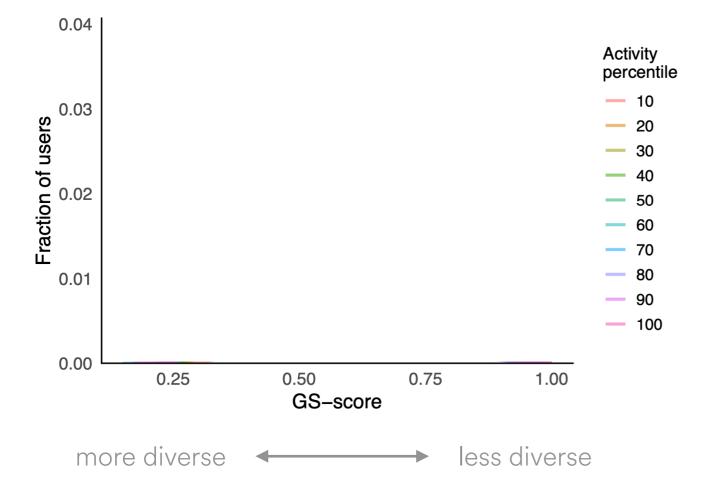
$$\vec{u_i} = \frac{1}{\sum w_j} \cdot \sum_j w_j \vec{s_j}$$



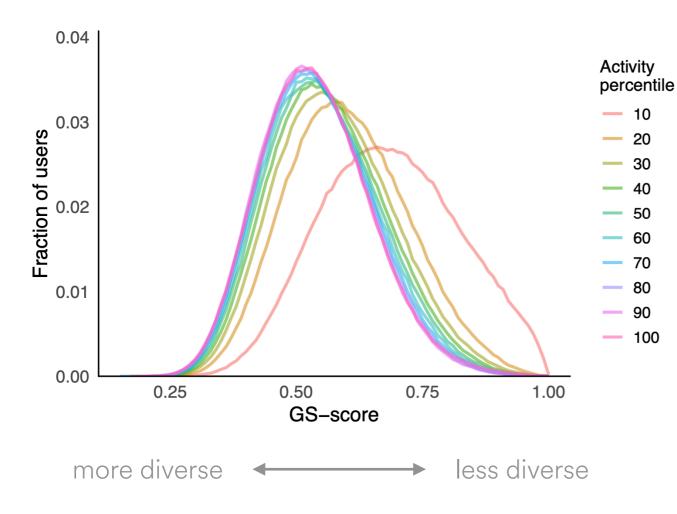


[1] Isaac Waller and Ashton Anderson. Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms. WWW 2019.

How is consumption diversity distributed?



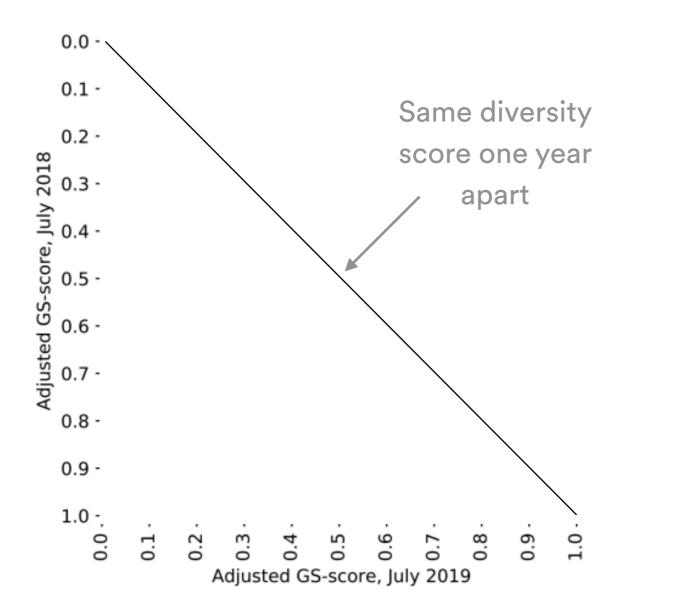
How is consumption diversity distributed?



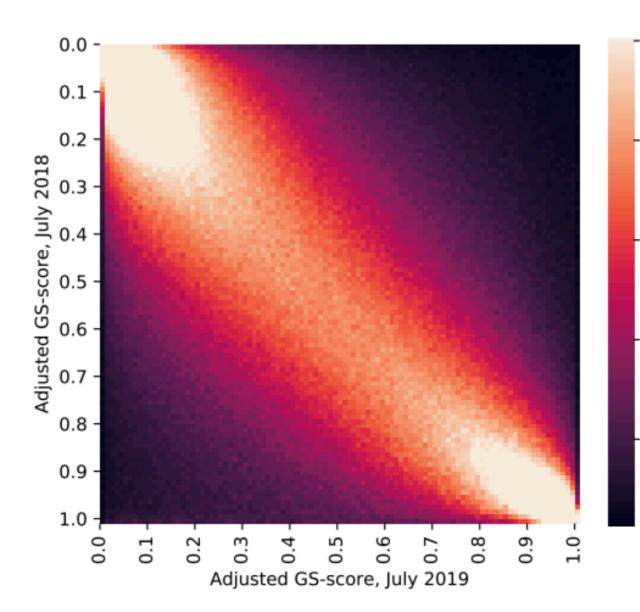
Wide range of consumption styles, from extremely narrow to extremely broad

Lowest activity users are more narrow, but activity and diversity are mostly uncorrelated

How stable is the diversity of a user's consumption over time?



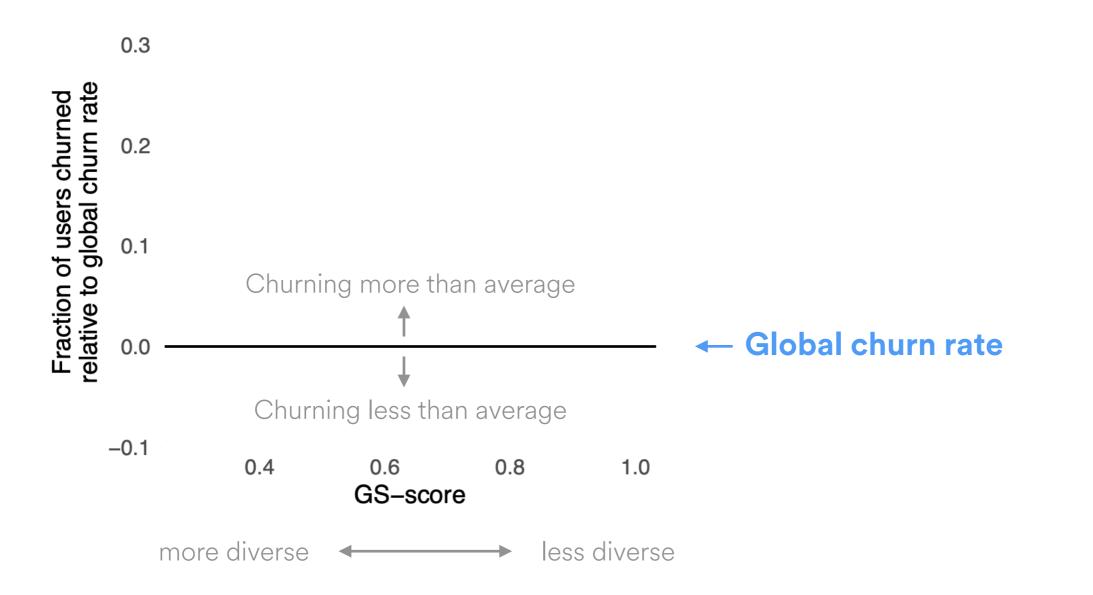
How stable is the diversity of a user's consumption over time?



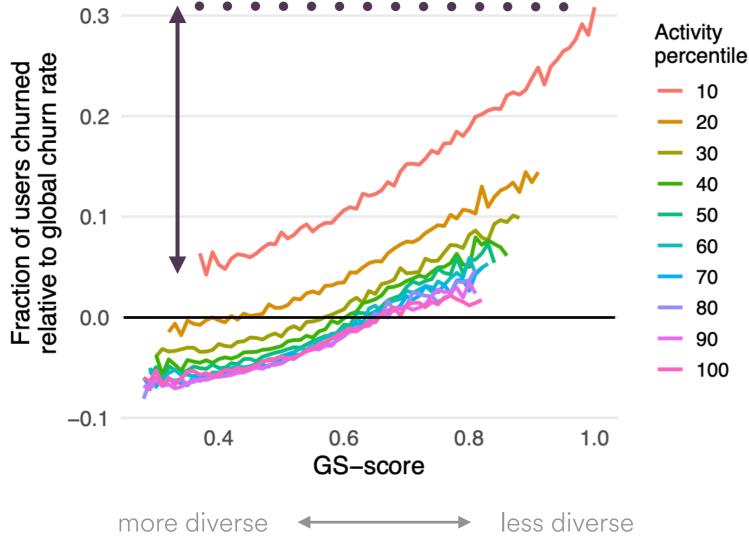
- 0.020 - 0.016 One year apart, user consumption diversity scores are similar, especially at the extremes - 0.008 User consumption diversity is typically stable over time

What are the relationships between consumption diversity and important user outcomes?

How is diversity associated with churn?



How is diversity associated with churn?



Activity

10

20

30

40

50

60

70

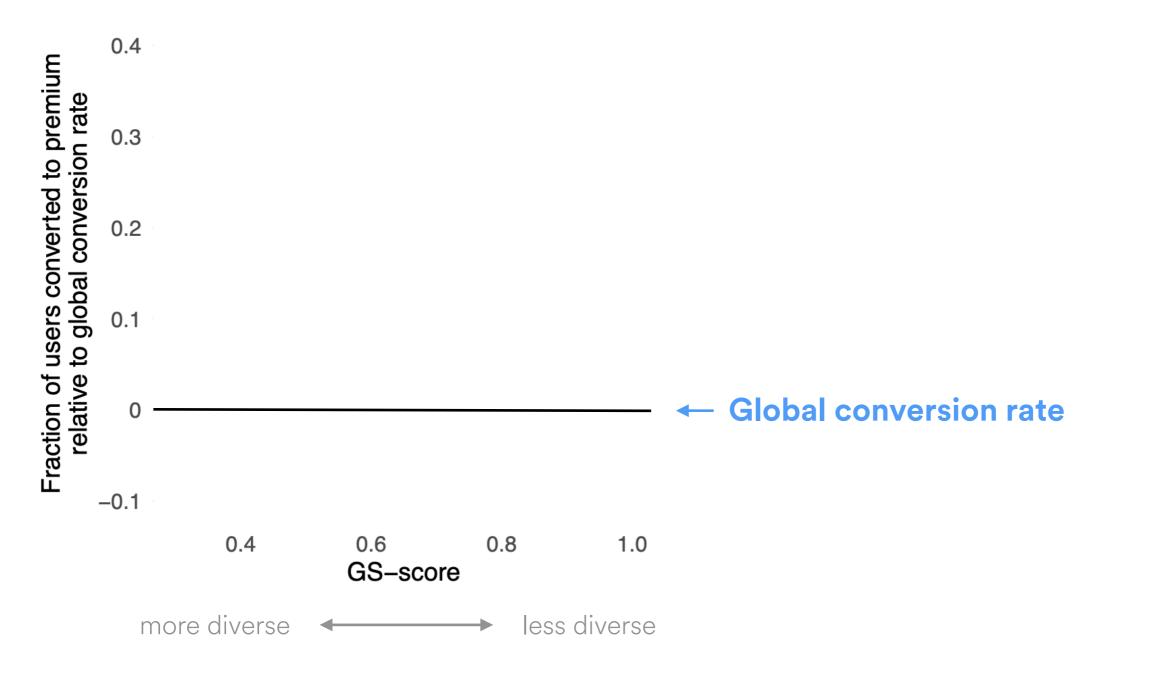
80

90

100

- **25 percentage points** decrease in churn
- Even controlling for activity,
- more diverse users are far
- less likely to churn

How is diversity associated with conversion?



How is diversity associated with conversion?

10

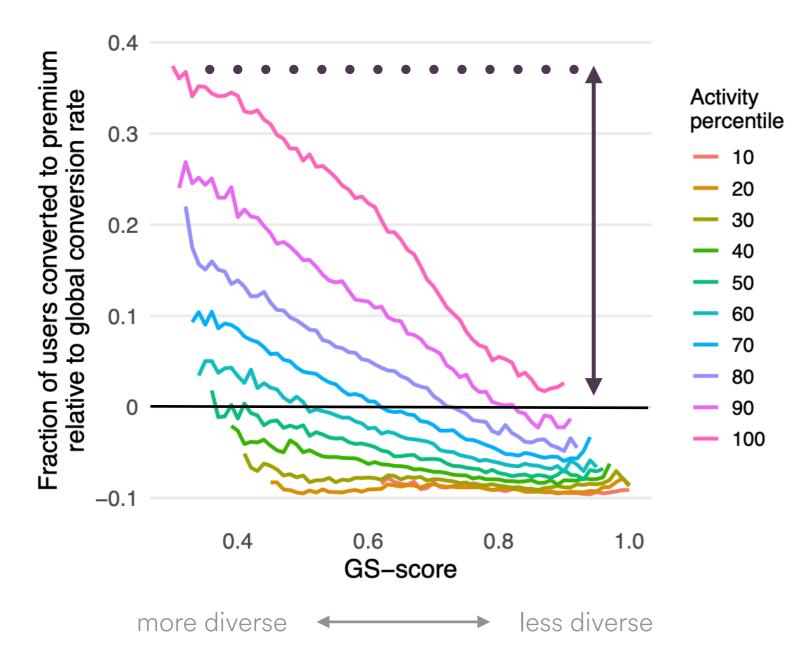
20

30

40

50

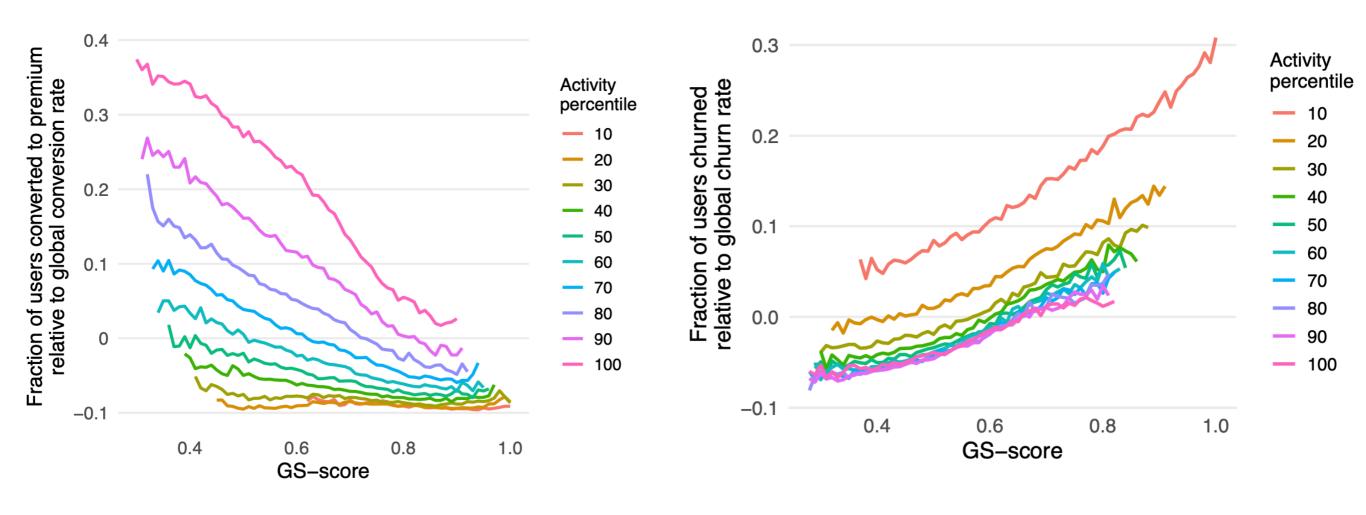
90



30 percentage points increase in conversion

Controlling for activity, more diverse users are far more likely to become premium members

Diverse listeners...

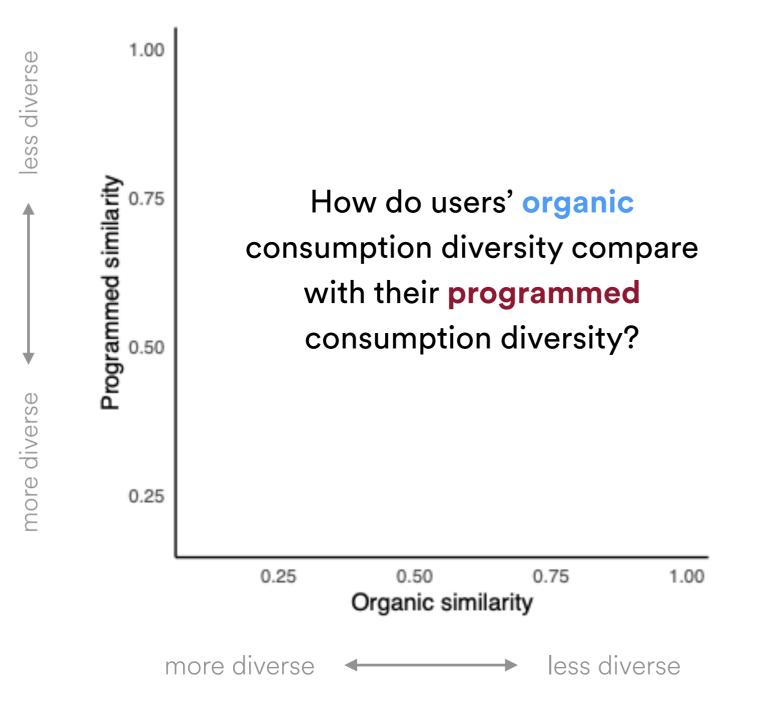


Convert more

and churn less

(controlling for activity!)

How are algorithmic recommendations related to the diversity of consumption?



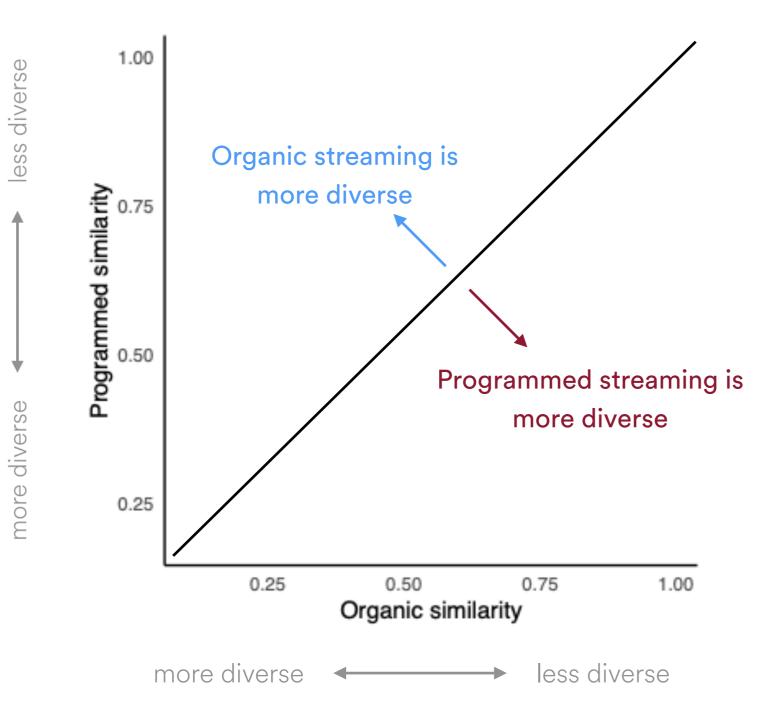
1.00 e.g. a user with **organic** GS-score = 0.40 and programmed GS-score = 0.80 Programmed similarity 0.25 0.25 0.50 0.75 1.00 Organic similarity more diverse less diverse

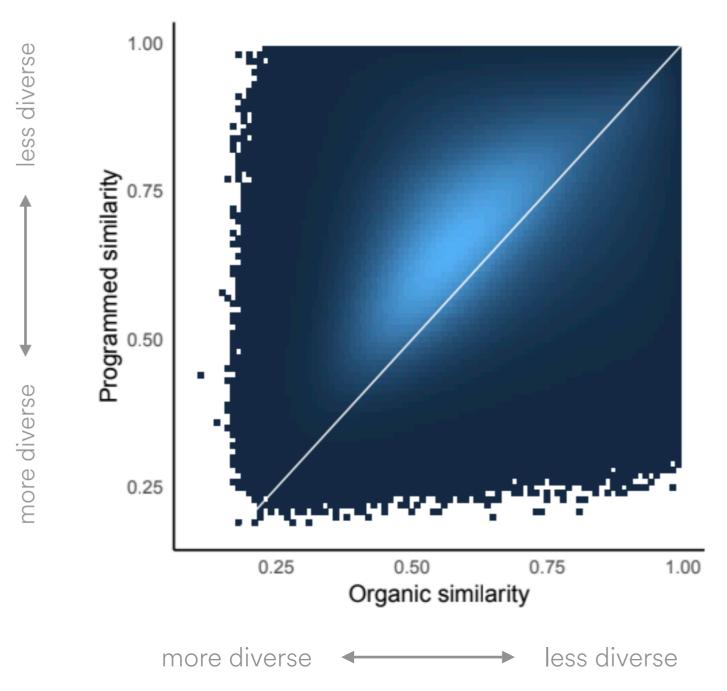
less diverse

more diverse

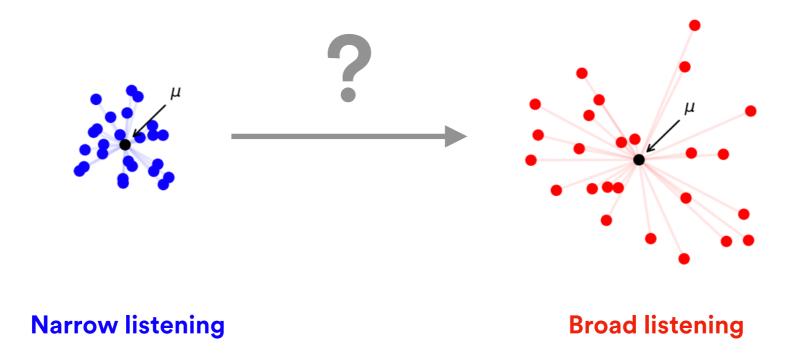
For each user, calculate **separate** diversity scores for their **organic** streaming only and their **programmed** streaming only

Examine joint distribution over all users to observe whether organic streaming tends to be more or less diverse than programmed streaming

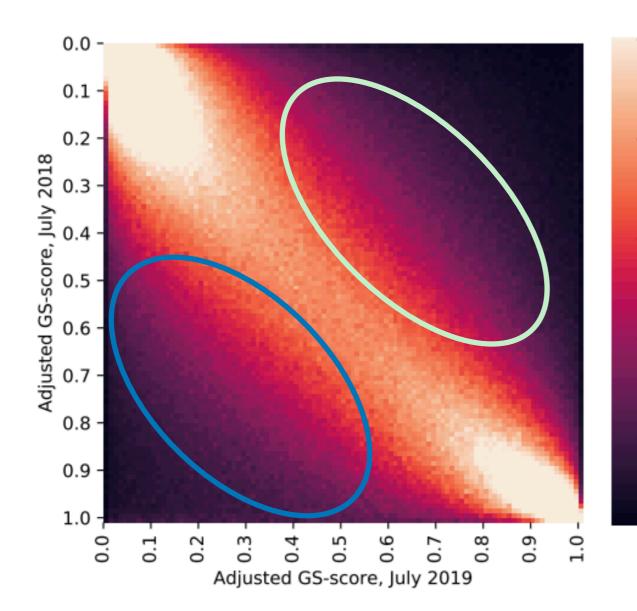




For most users, their **organic** consumption is more diverse than their **programmed** consumption



Dynamic view: when users change the diversity of their consumption over time, how do they do this?

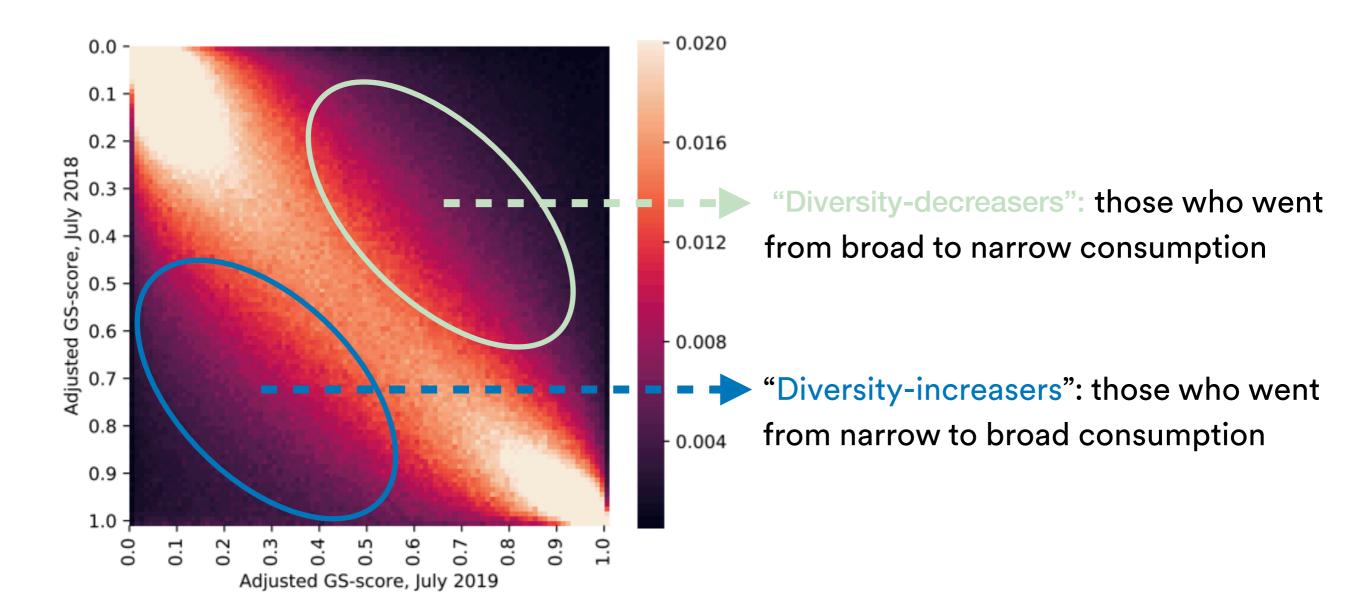


0.020

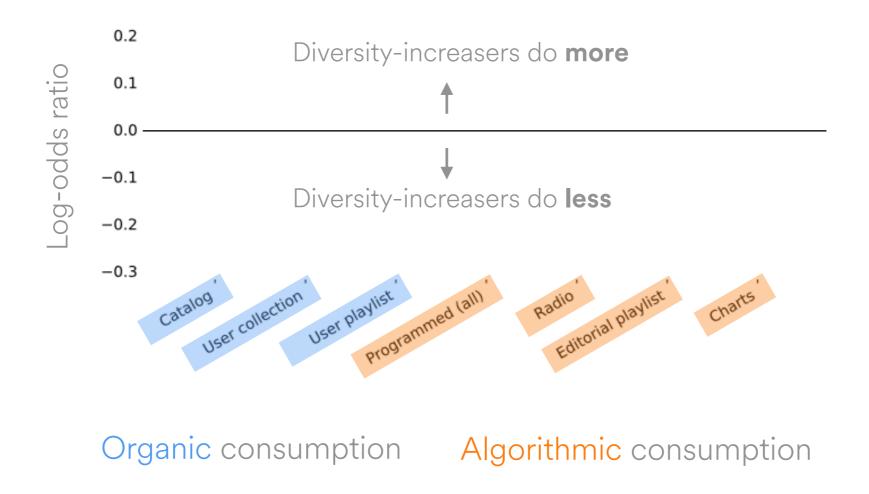
Although consumption diversity is typically stable, many users still change the diversity of their listening substantially from one year to the next

0.008

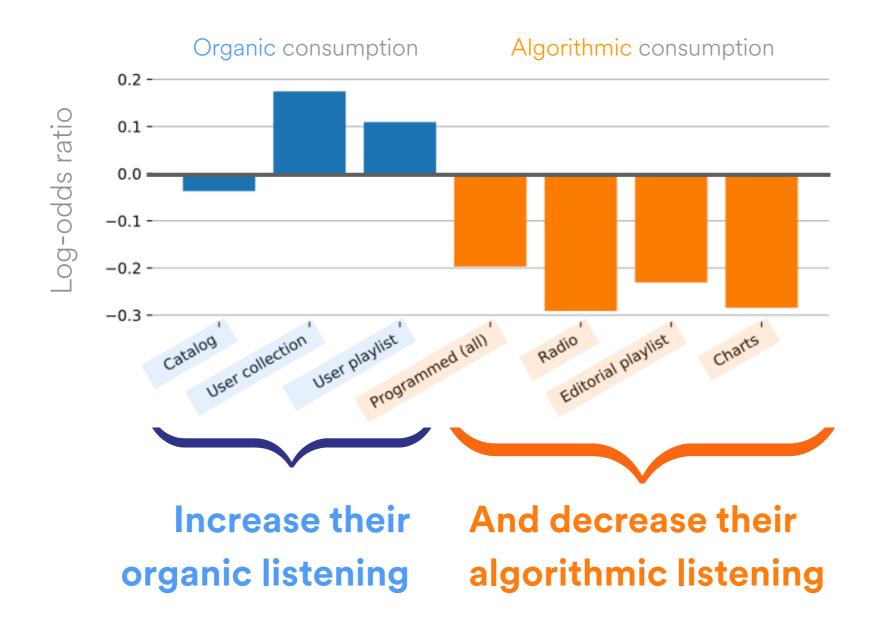
What changes in consumption patterns 0.004 drive these changes in diversity?

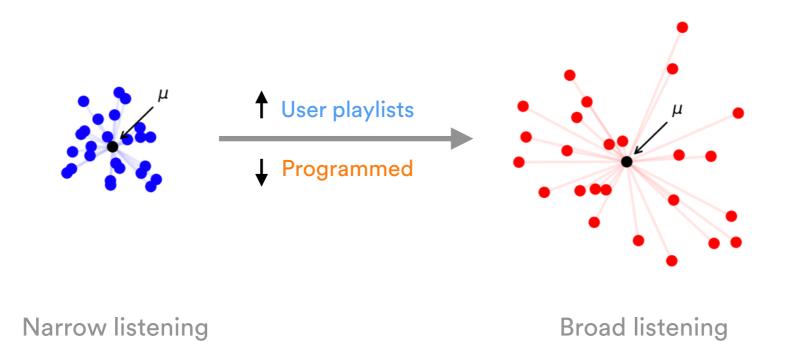


Analysis strategy: compare diversity-increasers to diversitydecreasers. How does listening between these two groups differ?

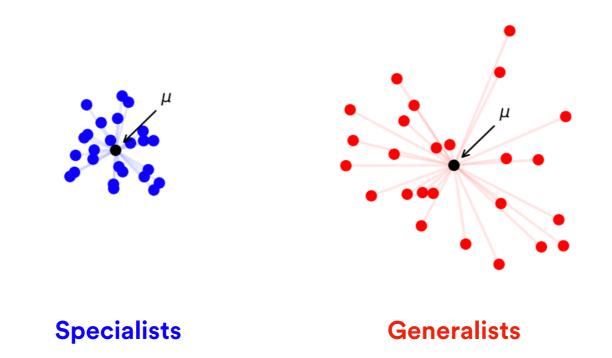


When users diversify their consumption over time, they:





When users diversify their consumption, they choose more for themselves and rely less on algorithmic recommendations How do users respond to algorithmic recommendations based on their consumption diversity?



How do generalists and specialists respond to algorithmic recommendations?

We ran a randomized experiment comparing how recommendation algorithms affect users based on their consumption diversity

Experimental context:

- Free users engaging with 7 popular playlists
- When they stream from one of them, a song is chosen at random from the top k songs
- Algorithms rank the songs to try to improve the user's experience

We randomly test 3 simple algorithms and measure their performance on generalists vs. specialists

•										
All Out 9	All Ou	All Out 90s								
16		Boy bands, girl groups, britpop, grunge & a golden era for hip hop. Created by Spotify - 89 songs, 6 hr 24 min								
	PLAY (•				LOWERS 664,850				
Q Filter										
TITLE										
♡ Fade Ir	nto You	Mazzy Star	So Tonight That I							
♡ On & C	Dn EXPLICIT	Erykah Badu	Baduizm							
♡ Got Ti	I It's Gone BPLICIT	Janet Jackson, Q	The Velvet Rope							
♡ Can I F	Gick It?	A Tribe Called Qu	People's Instinctiv							
♡ High A	nd Dry	Radiohead	The Bends							
	sion (Don't Ever Wonder)	Maxwell	Maxwell's Urban							
🗢 You An	id Me Song	The Wannadies	Be A Girl							
🗢 Doo W	op (That Thing)	Ms. Lauryn Hill	The Miseducation							
♡ Unfinis	hed Sympathy - 2012 Mix/Master	Massive Attack	Blue Lines (2012	2019-12-11		5:08				

A	Il Out 90s	Boy bands, girl gro	ut 90s pups, britpop, grunge & a 89 songs, [•] hr 24 mir [•]	a golden era for hip hop.			LOWERS ,664,850
Q Filter							
	TITLE		ARTIST	ALBUM	Ē		<u>(</u>)
\heartsuit	Fade Into You		Mazzy Star	So Tonight That I	2019-12-11	•••	4:56
\heartsuit	On & On		Erykah Badu	Baduizm	2019-12-11		3:46
\heartsuit	Got 'Til It's Gone		Janet Jackson, Q	The Velvet Rope	2019-12-11		4:02
\heartsuit	Can I Kick It?		A Tribe Called Qu	People's Instinctiv	2019-12-11		4:12
\heartsuit	High And Dry		Radiohead	The Bends	2019-12-11		4:17
\heartsuit	Ascension (Don't Eve	er Wonder)	Maxwell	Maxwell's Urban	2019-12-11		5:46
\heartsuit	You And Me Song		The Wannadies	Be A Girl	2019-12-11		2:50
\heartsuit	Doo Wop (That Thing	g)	Ms. Lauryn Hill	The Miseducation	2019-12-11		5:20
\heartsuit	Unfinished Sympathy	y - 2012 Mix/Master	Massive Attack	Blue Lines (2012	2019-12-11		5:08

Rank the songs in the playlist

		TITLE	ARTIST	ALBUM	1	ŀ
	\bigcirc	Fade Into You	Mazzy Star	So Tonight That I	2019-12-11	4:56
	\heartsuit	On & On EXPLICIT	Erykah Badu	Baduizm	2019-12-11	3:46
	\heartsuit	Got 'Til It's Gone	Janet Jackson, Q	The Velvet Rope	2019-12-11	4:02
	\heartsuit	Can I Kick It?	A Tribe Called Qu	People's Instinctiv	2019-12-11	4:12
	\heartsuit	High And Dry	Radiohead	The Bends	2019-12-11	4:17
	\heartsuit	Ascension (Don't Ever Wonder)	Maxwell	Maxwell's Urban	2019-12-11	5:46
	\heartsuit	You And Me Song	The Wannadies	Be A Girl	2019-12-11	2:50
	\heartsuit	Doo Wop (That Thing)	Ms. Lauryn Hill	The Miseducation	2019-12-11 …	5:20
	\heartsuit	Unfinished Sympathy - 2012 Mix/Master	Massive Attack	Blue Lines (2012	2019-12-11	5:08
ıg) ♡		5:16	C K	5:20	<u>는</u> [3 d) 	

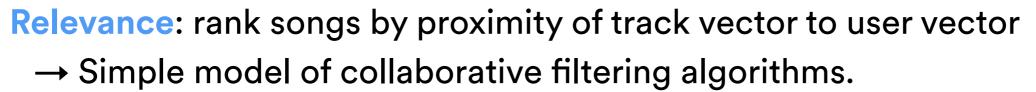
One of the top k songs plays at random

Measure number of plays (listening to the song) and skips (skipping the song)

3 ranking approaches:

Popularity: rank songs by number of streams.

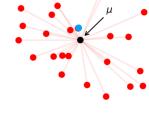
 \rightarrow Un-personalized baseline.

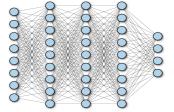


Learned: neural regression model learned from user-level, song-level, and interaction-level features.

 \rightarrow Learn from user's historical preferences.

User-level: country, affinity for various genres, user vector in embedding Song-level: popularity, genres, song vector in embedding Interaction-level: cosine similarity between user and song vectors (relevance), user's affinity for song's genres







Experimental design:

Randomly assign users to popularity, relevance, and learned conditions

540,000 users during 1-week period

7 playlists: All Out 50s, 60s, 70s, 80s, 90s, 00s, 10s

Outcomes: streams and skips

Compare outcomes for generalists (high-diversity users) versus specialists (low-diversity users)

PLAYLIST All Out 50 Case a time of charge Cost area a time of charge		ol.	Autor to Autor			AllOut AllOut AllOut AllOut AllOut All AllOut All All All Cut Cut All Cut Cut All Cut Cut All Cut Cut		thing for	The second secon			Boy bunch, girl	ut 90s reas. Integs., grunge & gislem en ter her her de Stange, der ziel min 1 ⓒ —	0. 600.04116 1.04.110	MOLEON PLAUET MOLEON All Out of the second	lecade that catapulted electronic r instream.	nusia, hip hop	Norsky All Out Transky Spatial Transky Spatial Transk		huir nask.
Q Filter			Q Filter			Q. Fitter			Q. Filter		S. 10				Q, Filter			Q. Filter		
TITLE			TITLE		œ	TITLE		ca	TITLE ART					2019-12-11 4:56	TITLE		m	TITLE		
I Fall In Love Too Easily - Vocal Version	Chet Baker		♡ lafahan	Duke Ellington	2019-11-29	Dreams - 2004 Remoster	Fleetwood Mag	2020-03-06	This Must Be the Place (Naive Melody) - 20 Talki	ing Heads 2019-12-11		2 0n140n	- Erykah Bedu - Badulen	2019-12-11 3.46	Alwara Like Thia	Bombey Bicycle Club	2019-10-11		James Blake	
Easy Living	Bilis Holdsy		Covery Sky Bast	Alice Coltrane	2019-11-29	ON Sweet Nuthin' - 2015 Remastered	The Velvet Undergro	2020-03-06	This Charming Man - 2011 Remaster The	Smiths 2019-12-11				2019-12-11 4:02	Weind Fishes/ Arpensi	Reficheed	2019-10-11	Georgia		
O Mr. Bue	The Fleetwoods		Ø Myself When I Am Real	Charles Mingus	2019-11-29	Feel Like Makin' Lave		5050-02-00	Running Up That Hill (A Deal With God) Kate	Bush 2019-12-11				2019-12-11 4:12	Mvisnos	Fleet Foxes		Something Good Can Work	Two Door Ginema Cl	
I Only Have Eyes for You	The Flamingos		Black Christ of the Andes (St. Martin de I	Porr Mary Lou Williams	2019-11-29				Everywhere - 2017 Remaster Fleet	twood Mac 2019-12-11				2019-12-11 4:17			2019-10-11	My Number	Foals	
Blue Skies	Ella Fitzgerald, Harry		This Will Be Our Year	The Zombies	2019-11-29	A Song for You		2020-03-06	Everybody Wants To Rule The World Team	s For Fears 2019-12-11				2019-12-11 5:46	Cittle Bit	Lykko Li	2019-10-11	The Mother We Share	CHVRCHES	
						⇔ Sugar	Stevie Wonder	2020-03-06						2019-12-11 2:50	♥ Burning	The Whitest Boy Alive	2019-10-11			
I Get Along Without You Very Well (Exce	pt S Chet Baker		On The Sunny Side Of The Ocean		2019-11-29	🗇 It's A Shame		2020-03-06	Smalltown Boy Bron	ski Beat 2019-12-11				2012-12-11 5:20	O New Slang		2019-10-11		London Grammar	
♡ All of Me	Bilie Holday		○ Today	Jefferson Airplane	2019-11-29	C A Case of You	Jani Mitchell	2020-03-06	There is a Light That Never Goes Out - 2011 The	Smiths 2019-12-11		2 Unfinished Sympathy - 2012 Mix/Maste		2019-12-11 E-DR	Sea of Love	Cat Power	2019-10-11	Do You Remamber		

Comparison	User Type	Song Streams	Song Skips
Dalawanaa awar Danularitu	Generalists	+10.03%	+4.71%
Relevance over Popularity	Specialists	+25.66%	+2.89%
Learned over Relevance	Generalists	+1.82%	+0.90%
	Specialists	+1.30%	-9.76%

1. Personalized algorithms deliver big short-term wins for all users, but **especially for specialists**.

2. Incorporating more information (learned) also benefits specialists more than generalists.

3. Algorithmic recommendations don't work as well for users with diverse consumption (generalists).

Diversity-aware recommendation

There is a clear need for diversity-aware recommendation algorithms.

→ Recommending content for specialists is different than recommending content for generalists.

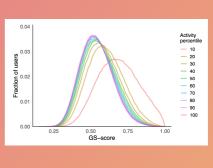
Recommendation algorithms may be over-optimizing for short-term goals at the expense of long-term goals.

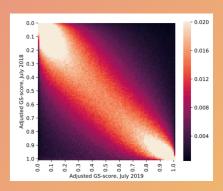
→ A "grand challenge" for recommendation: How do you satisfy pressing user needs and keep the big picture in mind at the same time?

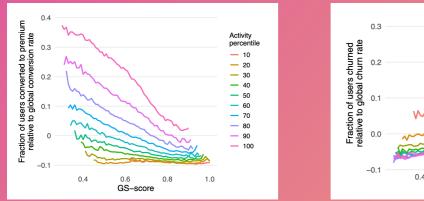
The causal effects of recommendation algorithms on consumption diversity and user outcomes are still unclear.

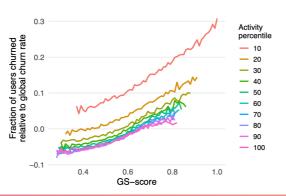
- \rightarrow Our analyses are correlational.
- → A recent field experiment on Spotify suggests that personalized recommendation algorithms causally decrease diversity.

Consumption diversity varies a lot between people but is typically stable within people.



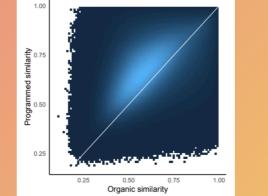


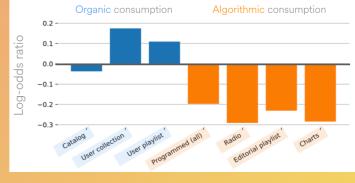




Diverse listeners convert more and churn less.

Recommendation algorithms are associated with reduced consumption diversity.





Comparison	User Type	Song Streams	Song Skips
Relevance over Popularity	Generalists	+10.03%	+4.71%
	Specialists	+25.66%	+2.89%
Learned over Relevance	Generalists	+1.82%	+0.90%
	Specialists	+1.30%	-9.76%

There is a need to develop diversityaware recommendation algorithms.

Thank you!

ashtona@spotify.com