

Curation by code: Infomediaries and the data mining of taste

European Journal of Cultural Studies

2015, Vol. 18(4-5) 446–463

© The Author(s) 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1367549415577387

ecs.sagepub.com



Jeremy Wade Morris

University of Wisconsin–Madison, USA

Abstract

Automated recommendation systems now occupy a central position in the circulation of media and cultural products. Using music as a test case, this article examines the use of algorithms and data mining techniques for the presentation and representation of culture, and how these tools reconfigure the process of cultural intermediation. Expanding Bourdieu's notion of cultural intermediaries to include technologies like algorithms, I argue that an emerging layer of companies – call them infomediaries – are increasingly responsible for shaping how audiences encounter and experience cultural content. Through a critical analysis of The Echo Nest, a music infomediary whose databases underpin many digital music services, I trace the shift from intermediation to infomediation and explore what is at stake at the intersection of data mining, taste making and audience manufacture. The new infomediary logics at work are computational forms of power that shape popular culture and highlight the social implications of curation by code.

Keywords

Algorithms, big data, cultural industries, cultural intermediaries, curation, digitization, infomediaries, recommendation engines, The Echo Nest

Give the wreckommender an artist that you like and it will give you a playlist of tracks from artists that are very different from the seed artist. Some obvious use cases: Your 14-year old daughter's slumber party is getting too loud. Send the girls home by putting on the Hannah

Corresponding author:

Jeremy Wade Morris, Department of Communication Arts, University of Wisconsin–Madison, 6132 Vilas Hall, 821 University Ave, Madison, WI 53706, USA.

Email: jwmorris2@wisc.edu

Montana Wreckommender – which yields a playlist with tracks by Glenn Gould, Dream Theater and Al Hirt.

Lamere (2009)

The Wreckommender is a music recommendation program with a twist. Most recommendation systems build desirable playlists based on a seed artist or song (e.g. if you liked Weezer, you might also enjoy Tony Molina). The Wreckommender, on the other hand, provides anti-recommendations, or ‘wrecklists’. Software developer Paul Lamere created The Wreckommender using the algorithms and database from the company he works at called The Echo Nest. Normally, The Echo Nest’s database groups similar sounding artists together, but Lamere used the database against the grain to see what would come of an app that turned out the most ill-fitting recommendations possible.

Lamere created the program as a novelty for a music hack day event in 2009, but this article takes seriously three key insights The Wreckommender animates about digital goods and the information technologies with which they are imbricated. The first, which I explore in the first two sections of this article, concerns data mining and recommendations and how these techniques combine to serve in the intermediation of contemporary musical experiences. The Wreckommender is one of the countless examples of new forms of ‘curation’ possible now that the music commodity is highly digitized. As part of a broader ‘algorithmic turn’ (Uricchio, 2011), automated recommendation systems now occupy an increasingly central position in the circulation of media and cultural products. Whether through the recommendation engines that Netflix, Apple or Amazon employ to advertise and sell products or through the algorithmic analysis companies like Epagogix and Music X-Ray use to predict the success of future films or albums, these automated (sometimes autonomous) tools combine massive databases of digital goods with behavioural tracking technologies to create connections between purchases, listens and views that were previously difficult to make, or severely limited in scope. Far from neutral purveyors of predictions, recommendation systems measure and manufacture audiences to provide targeted suggestions for popular cultural goods and exert a logistical power that shapes the ways audiences discover, use and experience cultural content.

The second focus of this article, which I explore in the ‘From intermediation to information’ section of the article, is on the role companies like The Echo Nest play in the cultural industries. The information The Echo Nest collects and allows access to through its application programming interface (API) provides the functionality for hundreds of database-powered applications and recommendation services, from The Wreckommender to Spotify. It is part of an emerging layer of organizations that I refer to as *infomediaries*: organizations that monitor, mine and mediate the use of digital cultural products (e.g. e-Books, music files, video streams, etc.) as well as audience responses to those products via social and new media technologies. Building on Pierre Bourdieu’s (1984) notion of cultural intermediaries as an ‘occupational grouping’ (Negus, 2002) of meaning-making workers that mediate both consumers’ experiences of cultural goods and producers’ understandings of their target audiences, I argue that infomediaries – largely using automated and data-based technologies and methods to surveil taste – complicate practices that were traditionally the province of cultural intermediaries and highlight the

increasing amount of personal and socially aggregated data that now inform how users discover and experience cultural goods.

Finally, through a case study of The Echo Nest's platform that analyzes its business documents, developer tools and database capabilities in the section 'Curating the algorithmic tune', I explore how a specific infomediary builds and implements the informational infrastructure on which new musical experiences and technologies are currently being built. While Romantic discourses about music and art have turned the debate on algorithmic curation into a discussion over the quality or accuracy of machine recommenders, infomediarities like The Echo Nest trouble this reductive reading. The Wreckommender, for example, is less about how accurate or knowing it is about a particular user's tastes and more about the possibilities that come from music that can be segmented and sorted across countless attributes. The Wreckommender employs the affordances of The Echo Nest's database to present counter-intuitive recommendations. It confirms that the logics to which algorithms conform are editable and capable of being tuned to multiple ends.

Ultimately, I argue that a more significant analysis remains regarding the role infomediarities play in the surveillance of taste, and the political economic infrastructure that undergirds recommendation systems. As digitization blows cultural goods up into millions of data points, and big data collide with ubiquitous media, cultural studies scholars need to question how the increasing integration of personal and socially aggregated behavioural data shift notions of curation, discovery and the experience of digital cultural commodities. By looking at the transition from intermediation to infomediation, this article considers the new infomediary logics at work in the production of culture and explores what is at stake at the intersection of data mining, taste making and audience manufacture. In doing so, it attends to the computational forms of power shaping popular culture and interrogates the social implications of curation by code.

Thinking through intermediaries

The turn towards digital cultural commodities and services has expanded the possibilities for tracking the use of cultural goods and for putting those usage statistics in service of further commodification. Whether or not we have hit what Uricchio (2011) calls an 'algorithmic turn' is an open-ended question, but it is clear that algorithms for predicting tastes, preferences and behaviours are increasingly central to a variety of industries (e.g. health care, government services, finance, insurance, customer service, etc.), and they have found a particularly strong foothold in the realm of culture. From Amazon book recommendations to Netflix movie suggestions to Spotify song selections, there now exists a dizzying array of algorithms and recommender services monitoring our tastes and suggesting future content, each operating according to their own corporate, technical and cultural logics.¹ If we previously looked to cultural intermediaries like the bookstore retailer, the film festival programmer or the radio DJ to help curate culture, algorithms and recommendations services increasingly carry out these roles. My question for the next two sections, then, is whether or not we should consider the algorithms that underpin recommendation engines as cultural intermediaries. If so, how does this shift our conception of the work intermediaries do?

A special issue of this journal in 2013 begins with the claim that ‘the term cultural intermediaries has been good to think with’ (Smith Maguire and Matthews, 2012: 551). This is undoubtedly because the term turns us towards theoretical issues of taste, class and value as well as to more practical concerns, such as the everyday labour of creating cultural products and preparing them for circulation. Since Bourdieu called attention to the role of cultural intermediaries in *Distinction*, and despite the contested nature of the term’s use after Bourdieu, cultural industries scholars keep returning to it as a concept for making sense of various occupational roles that affect the circulation of cultural goods (Featherstone, 1987; Hennion, 1989; Hesmondhalgh, 2006; Miege, 1979; Nixon and Gay, 2002). Bourdieu’s original use was relatively specific (Hesmondhalgh, 2006). He situated cultural intermediaries as a subset of the new petite bourgeoisie² who were ‘the producers of cultural programmes on TV and radio or the critics of “quality” newspapers and magazines and all the writer-journalists and journalist-writers’ (Bourdieu, 1984: 325). By presenting and representing symbolic content, the new petite bourgeoisie generally, and cultural intermediaries more specifically, deployed a distinct set of skills to establish themselves as key agents involved in defining the symbolic characteristics of cultural goods.

Beyond Bourdieu, the term’s definition has stretched. Scholars have used the term to link graphic designers (Soar, 2002), television buyers (Kuipers, 2012), ratings companies (Childress, 2012), bartenders (Ocejo, 2012), record producers (Hennion, 1989), fashion designers (Skov, 2002) and even workers in roles not immediately thought of as creative or cultural, such as the accountants at record labels (Negus, 2002), through their shared occupational position between consumers and producers, as well as their shared influence in shaping the symbolic meanings and physical visibility of cultural products. The term proved ‘good to think with’, but its expanded definition meant that almost anyone in the chain of cultural production qualified as a cultural intermediary, making it ‘a very poor starting point for an enquiry into the relationships between media and cultural production and consumption’ (Hesmondhalgh, 2006: 227). The term’s flexibility has been exacerbated with the proliferation of amateur blogging, podcasting, citizen journalism and related practices that have extended the curatorial capabilities traditionally limited to cultural intermediaries to a much larger group of individuals.

This tension between the growing number of workers engaged in what might be considered cultural intermediary roles and the inability to define precisely what intermediaries do is what prompts Smith Maguire and Matthews (2012) to ask ‘Are we all cultural intermediaries now?’ (p. 551). Their answer is a qualified no. If someone is only transmitting or passing on a cultural object, for example, and not adding new meaning to it, she or he is not necessarily an intermediary. For Smith Maguire and Matthews (2012), ‘[Cultural intermediaries] construct value, by framing how others – end consumers, as well as other market actors including other cultural intermediaries – engage with goods, affecting and effecting others’ orientations towards those goods as legitimate’ (p. 552). Intermediaries, then, are contextually specific actors who are involved in framing the interactions between cultural goods and those who encounter them, and they do so by virtue of the cultural legitimacy they accrue (i.e. a legitimacy which comes from a demonstrated ability to perform and be recognized as an intermediary in a given field).

With these definitional challenges in mind, it might be more fruitful, as Powers (2015) suggests, to shift our analytic gaze from intermediaries to the *process* of intermediation:

I use intermediation not just as a grammatical slight of hand to get away from nitpicking about who is a veritable intermediary and who is not. Rather, I conceive of cultural intermediation as a dynamic process of circulation that involves people, symbolic forms, and objects as much as it involves the modes of transportation and transmission that allow these elements to be linked. (p. 122)

For Powers, the symbolic value and meaning emerges from the processes of circulation and exchange, rather than exclusively from the work of any given occupational actor(s).

Powers' intervention points beyond the term's imprecision to another, and perhaps more important, gap in the literature on cultural intermediaries: the fact that studies of intermediaries deal almost exclusively with people. The role of a cultural intermediary is a job (e.g. finding talent, conducting focus groups, etc.) and it is usually *someone's* job (e.g. Artists and Repertoire Representative or A&R reps, advertising executive, bartender), not *something's*. The majority of studies on cultural intermediaries focuses on the role of humans, rather than the kinds of intermediation that might take place by objects, technologies and devices (Moor, 2012; see also Latour, 1988). As Moor (2012) argues, 'non-human and/or material forms of agency can be just as significant contributors to "intermediary" or mediating activities as human ones, and [...] they should be acknowledged as such' (p. 565). Looking specifically at the role statistical and ratings firms and their measurement techniques³ play in the intermediation process, Moor reminds us that devices and non-human actors can play an equal if not more important role than traditional cultural intermediaries.

If cultural intermediaries such as critics, talents scouts and the like were once responsible for the 'presentation and representation' (Bourdieu, 1984) of culture, or for the construction of value and framing of encounters with cultural goods (Smith Maguire and Matthews, 2012), it is hard to deny that recommendation services, and the algorithms that constitute them, are increasingly part of the intermediation process. As a variety of cultural content moves towards streaming or subscription models, where a relatively unlimited library of content is made available for a fixed monthly price or for free with advertisements, recommendation services and algorithms become increasingly important vectors on which companies seek to differentiate themselves. Although the above debates might seem like purely academic semantics about whether or not a particular role, profession or process fits an analytical term's criteria, Bourdieu's main purpose for deploying the term was to identify an emerging social class, and the cultural capital this group both held and produced through their distinct tastes and preferences. Lacking some of the more formal training and education of those in higher social classes, taste and the ability to curate information, styles and fashions became the source of new capital, and eventually, power and influence over the production and reception of cultural content. Recognizing algorithms and recommendation engines as part of the intermediation process turns our attention not only to how these new technologies take part in an old process of shaping curation and discovery, but also to the ways algorithms are

discursively and technically deployed to justify the legitimacy and quality of the services they underpin. In other words, the question at hand is less about whether or not algorithms and recommendation services are involved in intermediation; rather, it is about how they reconfigure the work traditionally carried out by cultural intermediaries.

Recommendation as intermediation

Although it is difficult to understand exactly how an algorithm works without a background in computer science, most users familiar with computers or the Internet have a basic idea of what an algorithm does. From Google searches to computer-crafted playlists to iTunes' Radio, most users understand that if they provide (or have provided) input data, then an algorithm will, through some combination of math and computation, analyze, sort and organize that information to return some kind of desired output. However, while algorithms are on the one hand 'encoded procedures for transforming input data into a desired output', they are also 'code with consequences' (Gillespie, 2014: 167, 192). They not only present information, but represent it in particular ways that shape what information users discover and how they do so. Algorithms, like intermediaries, exert a kind of power and control over the 'the social construction of knowledge' (Napoli, 2013: 23–24; see also Bucher, 2012; Barocas et al., 2013). They are like 'a distinct form of institution' (Napoli, 2013), a 'decision system' (Striphas, 2012) that combines a wide variety of factors – economic, technical, sociological – to reflect what the service providing them considers 'important' or 'accurate' or 'popular' and so on. As Beer notes (2009), 'This is undoubtedly an expression of power, not of someone having power over someone else, but of the software making choices and connections in complex and unpredictable ways in order to shape the everyday experiences of the user' (p. 997).

Algorithms and recommendation systems enact power through organization, through what they include and exclude and through how their results are organized and passed off as computational objectivity (Beer, 2009). From cases of companies dropping hundreds of pages down in search results due to modifications by Google to its algorithm (e.g. see the case of Kinderstart.com) to adult content being filtered from the first page of YouTube search results to maintain a particular corporate image (Gillespie, 2010), there is no shortage of examples where the power of algorithms is on display. This kind of power is, in John Durham Peters' (2008) words, logistical. Unlike media that rely heavily on the effects of their content (e.g. TV, film, music), the job of logistical media is 'to organize and orient, to arrange people and property in time and space' (Peters, 2008: 6). Calendars, maps, clocks and newer technologies like Google are all logistical because 'they have the power to set the terms in which everyone operates' (Peters, 2008: 6). Logistical technologies do not produce content like traditional media, but they organize, sort and repackage existing information to shape the ways we experience a variety of cultural content. They are, in a sense, organizational intermediaries (Andrejevic, 2013). Instead of offering exclusive content or services, organizational intermediaries like Google or Facebook offer free services and then use the data generated from the activities to commodify audiences. Organizational intermediaries impose new forms of control through data monitoring and presentation:

If, once upon a time, mass mediation imposed scarcity through the limitations of content and distribution, in the digital era, it imposes scarcity through the activity of organizing access to information – that is, determining which content will be prioritized for which users. (Andrejevic, 2013:199)

The organization of information, rather than its content, becomes the free lunch in the digital era (Smythe, 1981).

That said, today's algorithms and recommendation systems are not entirely autonomous systems. They depend highly on the people who design them and use them. Rather than neutral or objective technologies of presentation, algorithms and recommendation services are the result of the interactions of human actors (within a market context) and code (acting in both intended and unintended ways) seeking to frame meaning and curate experiences using a given set of raw materials (a database) and towards a distinct end (increasing listening time, pointing users to other related commodities, etc.). Arguably, DJs, A&R executives and marketing managers all exerted forms of logistical power, but where algorithms and recommenders differ from previous cultural intermediaries is in the amplification and automation of behavioural data collection and the deployment of that data in the very process of recommendation and curation. This makes the work of algorithms and recommenders far more embedded and integral to the intermediation processes for digital goods; consumption is the first step in further curation. By breaking digital commodities up into countless tags, segments and connections, algorithmic intermediation involves the fragmentation of individual pieces of cultural content (e.g. the song, the film, the book, etc.) and allows for the repackaging of cultural content in ways that not only seem highly personalized (as might have been the case with the radio DJ), but discursively claim to be drawing from troves of individualized and aggregated data.

From intermediation to infomediaation

Given that culture, as data, can be manipulated, sorted and related in ways that far extend human capacities, a new class of services has emerged that truly treats cultural goods as software, as code upon which other code can be written or built. Extending the discussion in the previous section, I want to call these services *infomediaries* – organizational entities that monitor, collect, process and repackage cultural and technical usage data into an informational infrastructure that shapes the presentation and representation of cultural goods. In the late 1990s, business scholars and practitioners coined the term infomediary (see for example Hagel and Singer, 1999) to name companies involved in helping consumers or businesses take control over the information advertisers and marketers were collecting on them, and for companies involved in aggregating data on behalf of consumers to package and present to vendors. Although my use of the term here refers to similar practices (i.e. information collection, aggregation), I want to expand its use by focusing on how infomediaries are at once highly technical and highly cultural. They primarily use automated and data-based technologies and techniques, though they also involve significant contributions from software developers, algorithm designers and database engineers. The term also shares similarities with Andrejevic's

(2013) organizational intermediaries or Alison Hearn's (2010) feeling-intermediaries, but it stresses the crucial role information – its collection, deployment and discursive power – plays in the intermediation process. Infomediaries expand the capabilities of digital commodities by helping users discover new cultural goods or providing novel experiences of familiar cultural content, and they do so largely through persistent and pervasive data collection.

Although infomediaries are surfacing across various sectors in the cultural industries, I want to consider the case of a particular infomediary, The Echo Nest, and to use music as a prophetic example for changes taking place across a wider spectrum of digital cultural commodities.⁴ That the music industries are relying on logistical informational technologies is hardly a new phenomenon; organizational intermediaries like Nielsen's Soundscan (McCourt and Rothenbuhler, 1997) or Gracenote's CD Database (Morris, 2012) have provided crucial informational infrastructures for the last two decades. However, the emerging infomediaries differ in the kinds of services they provide and the ways in which they integrate themselves into the practices of digital music consumption. I use the term infomediary to mark this shift; it is meant to denote the influence database and algorithmic technologies have not just for the organization of digital goods, but for the curation of culture through the mining, repackaging and provision of data. Infomediaries include the technologies themselves as partially independent actors (i.e. developers can program an algorithm to work in certain ways, but frequently other unintended consequences can occur) as well as the actions of the designers and developers who are most directly responsible for creating them.

The Echo Nest was founded in 2005 by two MIT Media Lab PhD students with interests in music and computing. Launched officially in 2008, The Echo Nest's (2014d) bio page refers to itself as a music intelligence company that provides the largest repository of dynamic music data in the world – over a trillion data points on over 35 million songs and 2.7 million artists to music and technology services such as Clear Channel, Vevo, MOG, MTV, Nokia, Rdio, Spotify, Coca Cola, Microsoft, Twitter and over 18,000 independent app developers. It offers musical services (i.e. data attributes about various songs and artists), measurement services (i.e. fan and trend tracking) and technical services (i.e. a platform that underpins other applications). If you have used a digital music service online other than Pandora or iTunes, or if you have used any of the 400 Echo Nest-powered music apps for iOS and other platforms, you have made some use of the company's platform, likely without knowing it. Recently acquired by Spotify for US\$100 million (Flanaga, 2014), The Echo Nest is gaining more widespread prominence. However, as a business-to-business affair, it remains barely known despite being what some commentators call the 'Most Important Music Company on Earth' (Faraone, 2011).

The Echo Nest generates the bulk of its data through a technique called machine listening. While some recommendation services, like Pandora, use trained musicologists to code songs by hand and ear (e.g. classifying songs by their key, their genre, the instruments they features and other such variables), The Echo Nest relies largely on computer extraction of data. Their software parses an entire song in a few seconds and processes the signal into thousands of unique segments, including timbre, beat, frequency, amplitude, vocal syllables, notes and other computer-measurable characteristics. Their

techniques are proprietary, but they claim to simulate ‘how people perceive music [by incorporating] principles of psychoacoustics, music perception, and adaptive learning to model both the physical and cognitive processes of human listening’ (Echo Nest, 2014b; Jehan et al., 2010). However, since acoustic similarity is a fairly limited way to classify music – think of the acoustic similarities but cultural differences between Christian Rock and mainstream rock – The Echo Nest also crawls the web for contextual information, or what they call ‘cultural’ data. Using natural language processing techniques, their technology crawls the web for text about songs or artists from blogs, music reviews, user forums and social media (about 10 million documents a day). For example, if terms like ‘dreamy’ or ‘ethereal’ are frequently used to describe a Beach House album, The Echo Nest assumes there is a connection between these words and the band’s sound. They combine the resulting machine and cultural listening data points into a database called the ‘musical brain’ (Faraone, 2011; Jehan et al., 2010).

Interestingly, The Echo Nest makes much of its database available for use through its developer APIs. Although the specific technologies it uses to crawl the web for discourse or to segment acoustic data are proprietary, significant portions of the resulting database are free to use for independent developers and can be licensed at variable rates by commercial services depending on the level of access to the database they require. Independent website designers and software developers can use the API to request data for use in any number of novel music-related apps, like The Wreckommender. The Echo Nest’s Dynamic Music Data solution, for example, aggregates the most current and up-to-date information on artists, including bios, reviews, videos, high-resolution photos from Getty Images and other paratextual content from and about artists. Using The Echo Nest API, developers can design music applications – say an app that suggests a playlist based on the weather – confident that their service will serve up current artist and song information without having to gather that data themselves. The service positions itself as a platform: a base on top of which a multiplicity of services will be built. It enables small-scale apps like Roadtrip Mixtape or The Music Maze to more professional endeavours, like Jog.fm, Deezer and Eventful (Echo Nest, 2014a).⁵

The music industries have long been information management industries as much as ones that create and distribute sound recordings. The acquisition, management and licensing of rights have been a driving logic of the sound recording industry since at least the introduction of sheet music. Infomediaries, however, create an extra layer of informational commodities based on paratextual and microtextual data that can be extracted through machine analysis. Rather than exploiting intellectual property rights, infomediaries exploit the algorithmic potential of digital commodities by collecting and compiling data that can be put in the service of creating novel ways to recommend, curate and experience cultural goods. If early digital music services struggled to gain access to large enough catalogues of music to be successful, infomediaries bypass the issue by providing a catalogue not of content, but of connections and relationships. End-point providers (Spotify, Rdio, etc.) manage the complexities of musical rights, while infomediaries mine and manage data rights for a spectrum of stakeholders (i.e. digital music service providers, music publications, labels, broadcasters, etc.) using a variety of commercial applications (e.g. Song Science, Trend Analysis, Strategic Planning, etc.). Many digital music services act as both end-point providers and infomediaries. Spotify, for example,

took part in infomediation before it began its relationship with The Echo Nest, though it has gradually shifted all these responsibilities to The Echo Nest since the acquisition. However, The Echo Nest still maintains some relationships with other business customers as well, suggesting infomediation is a complex interrelation of content and information about that content, and the distinction between end-point providers and infomediaries is not always clear. Despite this blending of roles, the products of infomediation – the databases, connections, statistics and correlations – still exist as a separate layer of informational commodities: a layer that parses songs into its component pieces for reassembly through various services and that provides value in a way that is distinct from traditional licensing rights.

Infomediaries are organizational, but they are also iterative. They offer a set of raw materials upon which other services are built. Many infomediaries create databases of information as well as a set of technologies for accessing and working with those data. While many companies might treat such a database as proprietary, The Echo Nest's technology gathers more data the more it propagates. Users who use a service that employs The Echo Nest (2014c) also send data back to the musical brain: 'The Echo Nest anonymously captures each user's digital musical activity in a Taste Profile – a persistent, real-time record of music play data (artists and songs) and behavior (favorites, ratings, skips and bans)'. The company's database is not just a static collection of machine-extracted variables and discourse about music, it is an ongoing taste monitoring project that brings in data from its network of Echo Nest-powered platforms and applications. With such a rich source of information on users – what The Echo Nest calls its 'Audience Understanding Solution' – the company has been actively putting these data to use. For example, they recently partnered with TargetSpot, an online digital audio advertising placement network whose ads reach over 64 percent of all online listeners. The deal gives TargetSpot access to The Echo Nest's audience insights to create 'more in-depth, actionable insights around the connection between a brand's target audience and music behavior' (Anonymous, 2013). The Echo Nest (2014c) has also used its Audience Understanding data to make connections beyond music, trying to make statistically meaningful connections between musical taste and a series of non-musical lifestyle categories, such as political affiliation, or taste in movies.

While the company is quite explicit about its ability to provide highly segmented and targeted advertising opportunities through results like 'fans of console sports video games are more likely to listen to rap artists (Ice Cube, Nas, Lil Wayne)' while 'fans of online "casual" social online games are more likely to prefer country artists (Blake Shelton, Sara Evans, Reba)' (Echo Nest, 2014c), these seemingly mundane connections belie how highly surveilled listening practices are when they take place on digital platforms. The increased ability to segment musical tastes and to use the data gleaned from musical practices makes each listening instance an economic opportunity for a host of unseen actors. The new digital traces created by skipping, blocking or rating a track serve as indicators that get rolled back into a much larger data profile for further targeting and refining. Although any consumer using a recommendation service is aware his or her tastes are being monitored and profiled, the detailed level of behavioural analysis is obscured from view. In an era of streaming and digital platforms, every single micro-cultural practice and usage statistic can be parcelled up as another data point to influence the process of intermediation.

Curating the algorithmic tune

Infomediaries shape tastes and derive legitimacy in a different manner than cultural intermediaries. Their ways of framing cultural goods are more organizational and embedded into everyday use than cultural intermediaries. Rather than the presentation of advertisements, or the placement of a song in a television show, infomediaries work behind the scenes/screens to affect the very interfaces of programs like Spotify or Netflix. Infomediaries collect past usage data and combine them with a much larger database of tastes and preferences and offer customized suggestions. They also provide the platforms on which new musical applications (and thus encounters with music) are built. The legitimacy of infomediaries, in the rhetoric of those who create and employ them, is based both on the cultural knowledge of those creating the databases and algorithms, but also on the size and scope of the databases and the efficacy of the algorithms themselves. If a traditional cultural intermediary was someone with intimate knowledge of what was good and could gauge the quality of cultural content before presenting it to you, infomediaries rely on the efficacy of the algorithms and databases to know what is essential about you and your tastes. The cultural content towards which infomediaries point users has less to do with quality and more to do with a supposed fit with quality than with a supposed fit with a user's individual preferences.

Just as the term cultural intermediaries has referred almost exclusively to people, we tend to think of curation as a distinctly human capability as well.⁶ Romantic ideals around the nature of music and the intimacy of personal taste heavily structure the discourse around recommendation services and their various differences (Steiner, 2012). Pandora, for example, employs a team of dozens of music analysts, under the assumption that despite how good computers may be at processing music, they can never understand it like humans can (Platoni, 2006). The recently announced music trumpets the expertise of their curators – authoritative voices from ‘Rolling Stone to Grand Ole Opry to Pitchfork to Hot 97’ (Beats, 2014). Both services still make heavy use of algorithms, but the claims they make about the legitimacy of their recommendation process revolve around distinguishing human responsibilities from machine ones. Music's digital format has made it infinitely analyzable and quantifiable in ways that make it increasingly amenable to algorithmic exploitation, but for almost all current music recommendation services, humans and machines work together during intermediation; services like Spotify, Beats and Pandora simply play up the ratio differently in their quests for distinction and differentiation.

A simple dichotomy of humans versus machines, then, misses the mark. While we might chafe at the idea that technology might be able to predict our taste in music or other cultural content, curation has long been a ‘symbiosis’ between human and machine inputs and outputs (Razlagova, 2013: 63). Tracing the rise of the DJ in the 1930s and 1940s to their cultural peak in the 1960s and 1970s, Razlagova mines a rich and quirky history of cybernetics and music to argue that curation has always been a hybrid process between people and the technologies at their disposal (Razlagova, 2013). The reductive dichotomy also ignores the fact that algorithms and recommendation engines are fundamentally human in their construction, if not their execution. Most recommendation services adapt to or depend entirely on how, and how many, users are using them. Algorithms employ individual and collective user profiles, acoustic data and other

metadata to collect, organize, filter and play songs. Recommendations are not based solely on what we do, but on what the user base does and what the algorithms that interpret that base do (Beer, 2009: 996).

Moreover, the very design of the algorithms and the decisions on which data to collect and parse are fundamentally human affairs. Music and other data, as Gillespie (2014) suggests, have to be made ready to be processed by algorithms (p. 171). Despite the nods to machine objectivity, the 3000 plus segments that The Echo Nest's software parses are not naturally occurring or neutral ones (Razlagova makes a similar argument, noting that Pandora's 'exotic' variable depends heavily on a Western/Anglo-Saxon-centred sonic perspective). While many musical features The Echo Nest selects are inherent to a song, the decisions on which segments to analyze and the relative weight to accord them in the overall database are, for the most part, the result of human choices. The text documents The Echo Nest collects, for example, come with a certain amount of noise – like sorting out what articles relate to bands like Boston or Pink or Fun whose names are also commonly occurring English words – that need to be cleaned up before being included in the database and distilled into more useable categories such as 'familiarity', 'hotttness' or 'most talked about'. This process is a significant challenge for computers and one that involves modifying and tweaking the database and its algorithms on a trial-by-trial basis (Jacobson, 2013).

There are also the various biases and assumptions that appear in the results list of recommendations. A search for female duo Tegan and Sara on my version of Spotify in 2012, for example, returned a range of recommendations, all of which sounded, to varying degrees, like Tegan and Sara, but shared the common denominator of being female (or having lead female vocalists). In this instance, the weight of 'gender' as a defining variable overrode sonic similarities. Similar queries with both male and female songwriters returned over 90 percent of the playlist of the same gender. Admittedly, this kind of trivial testing is more anecdotal than scientific. In fact, repeating this test on a different machine, on a different day or from a different user account would likely yield different results. The experiment vaguely hints at how the signals and cues that constitute recommendations are always culturally inflected and can rely on overly general classifications, like gender. When these assumptions get fed back into the database, they perpetuate further suppositions between practices and tastes, such as the simplistic results above regarding casual versus console games and musical preferences. But it also points to the difficulties of how we study and know things about infomediaries. Knowing what cues matter, and are incorporated into various recommendations is, at the moment, an impossible task for users; finding biases and omissions is obscured by the complexity of the systems making the recommendations and the variety of data sources that may be drawn on in the process.

We might also think of Mark Katz's (2004) notion of phonographic effects – where musicians, during the advent of early recording technology, altered their style of play to be better captured by microphones – to trace whether or not there are any discernable algorithmic effects taking place in these services. As these systems become primary means of distribution, musicians may attempt to record certain styles or sounds or resort to particular information generation techniques because they are easier or more likely to be picked up by algorithms. This may be partly what is at work in the growing amount

of artists known as musical spammers (D'onfro, 2014). These artists use a variety of tactics to game the algorithms, naming themselves after popular search terms (such as the name of a popular artist or song), cloning popular songs, releasing the same song under a thousand different titles and other such techniques. While some of these represent legitimate cover acts, others seem dedicated, like search engine optimization companies, towards making their material more visible among millions of other artists by manipulating search results. In either case, they represent noise that services like The Echo Nest need to filter or downplay in their algorithmic calculations. Whereas the effects Katz noted were immediately audible – musicians changed their styles of play in response to recording technologies – algorithmic effects are logistical and organizational. Users, for example, can rarely be sure that the results they are shown are the result of the actual functioning of the algorithms, or the result of entities (like musical spammers) trying to work those algorithms to their advantage. Either way, the search results will then contribute to user understandings of how algorithms function and what to expect from future searches or recommendations.

If we dispense with the binaries of human versus machine curation and we resist privileging one form over the other, we are instead able to focus on how thoroughly infused with culture decisions by infomediaries like The Echo Nest are, even if cloaked in the technical language of APIs and databases. Despite the shift from intermediaries to infomediaries, curation will continue to be a hybrid process and the use of algorithms as curators of popular culture will remain part of a larger and longer debate about discovery, taste, and aesthetic experiences (Powers, 2014). But the presence of infomediaries in this intermediation process has amplified the variables by which music can be sorted and recombined, and the avenues through which targeted advertisements and other messages can flow. By turning our attention to the practices of companies like The Echo Nest, to what their infomediary logics include and exclude, promote and hide, and to the ways cultural content itself is being readied for algorithmic analysis, media scholars can contribute to the pressing need to understand the materiality or political economy of algorithms (Anderson, 2012).

Conclusion

In an article about several digital music recommendation services, the *New Yorker* music critic Sasha Frere-Jones flashed back to the influence of the Roland TR-808 drum machine on the sound of hip hop and dance music, arguing that whoever programmed the sounds and the timing of the machine likely had more of an impact on the shape of these genres than the hundreds of famous artists who ended up using it on their albums. He goes on to say,

Similarly, the anonymous programmers who write the algorithms that control the series of songs in these streaming services may end up having a huge effect on the way that people think of musical narrative – what follows what, and who sounds best with whom. (Frere-Jones, 2010)

Frere-Jones puts his finger precisely on the key logistical role infomediaries play in the presentation, discovery and interconnectedness of cultural content. The drum machine programmers provided the raw materials on which a series of genres and styles were built; they did not make the songs but the sounds with which songs could be made. Infomediaries take this organizational quality one step further. They provide the databases upon which new services can be built, new data can be mined and new audiences can be manufactured.

Ultimately, thinking about recommendation services as infomediaries offers several contributions to existing literature on both algorithms and cultural intermediaries. First, it helps us place algorithms and recommendation systems in their industrial context, removing them from some of the mystery shrouded in technical language and placing them squarely as actors in the cultural industries that deserve equal attention as other cultural intermediaries or actors involved in intermediation. This entails not only understanding the various cultures of those who create the algorithms (see, for example, Nick Seaver's (2013) ongoing ethnographic work on music intelligence companies), but also trying to probe the algorithms and recommendation systems themselves to reverse-engineer their logics (Madrigal, 2014). Second, it broadens our understanding of cultural intermediaries to see them as a combination of human actors and technical devices. This has long been the case, though not often recognized in the literature on cultural intermediaries. The legitimacy of the work of cultural intermediaries is often dependent on a particular knowledge and use of certain technologies. Infomediaries foreground this hybridity. Third, infomediaries also point us to the increasing power information and the organization of information play in the infrastructures of the cultural industries. Data collection about uses and preferences is now embedded directly into the processes of presentation and consumption of cultural content. Owning the content still offers a substantial set of benefits and power to content holders, but in an era where content is not actually scarce, but falsely made to act scarce, organization is king.

Finally, algorithmic effects and the ways recommendation systems are biased and flawed and open to gaming should not be passed off as accidental or rare uses, but as a new kind of methodology for testing and probing the ways in which these systems work, are coded and expected to work, and perceived to be working. The relative complexity of algorithms makes them difficult objects (subjects?) of study since the exact equation and the cues that any given algorithm uses are usually private, proprietary information (Levy, 2010). Even if there was greater transparency about how algorithms and recommendation systems work (Beer, 2009), the fact that algorithms are highly fluid and dynamic entities – constantly tweaked and adjusted by their designers and regularly refined as a result of the data that get collected each time users employ the algorithm – means they are bound to remain at least somewhat always unknowable (Gillespie, 2014: 178).

Referring to the study they released about the correlation between political affiliations and musical preferences, one of The Echo Nest's founders noted, 'It started as an office joke, but after running the numbers, I can't escape the data' (Whitman, 2012). The comment – possibly sincere, possibly ironic – speaks to the company's desire to parse as much of the sonic, discursive and behavioural musical ecosystem as possible. It is also a nod to broader discussions taking place around 'data fundamentalism' (Crawford, 2013)

and the ability many infomediaries claim to have when using data on usage and preferences to predict taste, lifestyles and beliefs. As infomediaries like The Echo Nest quietly build the infrastructure on which many of our experiences with digital cultural goods rest, the very acts of interacting with cultural products fuel a recursive loop of future cultural recommendations. When every skip, rewind and pause feeds into a process of intermediation that curates what we view, hear and read next, we may find that we too cannot escape the data, even if we were not entirely sure what it tells us.

Funding

This research was funded in part by the University of Wisconsin-Madison's Graduate Studies Fall Research Competition.

Notes

1. Recommendation services are technically made up of multiple algorithms, which in turn draw on a variety of databases and information. Users rarely see algorithms at work, but see the results in things like recommendation systems. A company like The Echo Nest, for example, can provide the algorithms that underpin a third-party's recommendation apparatus, or it can provide the data from which other companies can build their own algorithms and recommendation systems. Companies invested in recommendation systems might draw data from a number of infomediaries to provide their recommendations.
2. Bourdieu described the new petite bourgeoisie as having less formal education than upper classes, but still possessing distinctive tastes and practices that proved useful in then-emerging professions like advertising, fashion and public relations.
3. For more on the role of new media measurement firms, see Morris (2015).
4. On account of its small size as a digital file, its ubiquity and its low bandwidth requirements, music felt the impact of digitization most intensely and immediately. As such, the case of music holds lessons for cultural commodities of all kinds.
5. Roadtrip Mixtape uses Google Maps and The Echo Nest's location data to craft a travel playlist using artists from cities users drive through. Jog.fm uses data on a song's beats-per-minute to match the pace of a user's run, The Music Maze is a labyrinth game with walls made of album covers drawn from The Echo Nest's image database. It is unclear how many other corporate services it will support in light of the Spotify acquisition.
6. Eric Harvey (2014) notes the increasing prevalence and force of the term curator for the developers behind services like Beats and The Echo Nest:

At a time when digitized cultural objects, from high art to gifs, circulate widely for anyone to sort through, recombine and display, a title formerly reserved for the organizers of art galleries and museum exhibitions now applies to the omnivorous activities native to the Stream.

References

- Anderson CW (2012) The materiality of algorithms. *Culture Digitally*. Available at: <http://culture-digitally.org/2012/11/the-materiality-of-algorithms/>
- Andrejevic M (2013) 'Free Lunch' in the digital era: Organization is the new content. In: Manzerolle V and McGuigan L (eds) *The Audience Commodity in a Digital Age: Revisiting a Critical Theory of Commercial Media*. New York: Peter Lang, pp.25–30.

- Anonymous (2013) The Echo Nest and TargetSpot partner to bring smarter advertising to online radio. *PR Newswire*, 21 November. Available at: <http://www.prnewswire.com/news-releases/the-echo-nest-and-targetspot-partner-to-bring-smarter-advertising-to-online-radio-232825671.html>
- Barocas S, Hood S and Ziewitz M (2013) Governing algorithms: A provocation piece. *Governing Algorithms*. Available at: <http://governingalgorithms.org/>
- Beats (2014) We are Beats music. Available at: <http://www.beatsmusic.com/about>
- Beer D (2009) Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society* 11: 985–1002.
- Bourdieu P (1984) *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.
- Bucher T (2012) Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society* 14: 1164–1180.
- Childress CC (2012) Decision-making, market logic and the rating mindset: Negotiating BookScan in the field of US trade publishing. *European Journal of Cultural Studies* 15: 604–620.
- Crawford K (2013) The hidden biases in Big Data. *Harvard Business Review*, 1 April. Available at: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>
- D'onfro J (2014) This man makes \$23,000 posting music spam on Spotify and iTunes. *Business Insider*. Available at: <http://www.businessinsider.com/matt-farley-makes-23k-posting-music-spam-on-spotify-2014-1>
- Echo Nest (2014a) Applications built on our platform. Available at: <http://the.echonest.com/show-case/>
- Echo Nest (2014b) Dynamic music data. Available at: <http://the.echonest.com/solutions/music-data/>
- Echo Nest (2014c) Music audience understanding. Available at: http://corpcdn.echonest.com.s3.amazonaws.com/filer_private/2014/01/27/audience_understanding_finalx.pdf
- Echo Nest (2014d) Our company. Available at: <http://the.echonest.com/company/>
- Faraone C (2011) The most important music company on earth. *The Boston Phoenix*, 22 April, 18–19. Available at: <http://thephoenix.com/boston/music/119592-echo-nest-is-ready-to-fly/>
- Featherstone M (1987) Lifestyle and consumer culture. *Theory, Culture & Society* 4: 55–70.
- Flanaga A (2014) Spotify's Ken Parks on an IPO, the Echo Nest Purchase (Q&A). *Billboard*. Available at: <http://www.billboard.com/biz/articles/news/digital-and-mobile/5930388/spotify-ken-parks-on-an-ipo-the-echo-nest-purchase-qa>
- Frere-Jones S (2010) You, the D.J.: Online music moves to the cloud. *The New Yorker*, 14 June. Available at: <http://www.newyorker.com/magazine/2010/06/14/you-the-d-j>
- Gillespie T (2010) The politics of 'Platforms'. *New Media & Society* 12: 347–364.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot K (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp.167–193.
- Hagel J and Singer M (1999) Private lives. *McKinsey Quarterly* 1: 7–15.
- Harvey E (2014) Station to station: The past, present and future of streaming music. *Pitchfork*. Available at: <http://pitchfork.com/features/cover-story/9383-station-to-station-the-past-present-and-future-of-streaming-music/6/>
- Hearn A (2010) Structuring feeling: Web 2.0, online ranking and rating, and the digital 'Reputation' economy. *Ephemera* 10(3/4): 421–438. Available at: <http://www.ephemera-journal.org/contribution/structuring-feeling-web-20-online-ranking-and-rating-and-digital-%E2%80%98reputation%E2%80%99-economy>
- Hennion A (1989) An intermediary between production and consumption: The producer of popular music. *Science, Technology & Human Value* 14: 400–424.

- Hesmondhalgh D (2006) Bourdieu, the media and cultural production. *Media, Culture & Society* 28: 211–231.
- Jacobson K (2013) How we resolve artists on the Internet. Available at: <http://blog.echonest.com/post/52159005051/how-we-resolve-artists-on-the-internet>
- Jehan T, Lamere P and Whitman B (2010) Music retrieval from everything. In: *Proceedings of the international conference on multimedia information retrieval (MIR'10)*, 29–31 March 2010, Philadelphia, PA, USA. New York, NY: ACM.
- Katz M (2004) *Capturing Sound: How Technology Has Changed Music*. Berkeley, CA; London: University of California Press.
- Kuipers G (2012) The cosmopolitan tribe of television buyers: Professional ethos, personal taste and cosmopolitan capital in transnational cultural mediation. *European Journal of Cultural Studies* 15: 581–603.
- Lamere P (2009) Paul's music Wreckommender. *Personal Blog*. Available at: <http://musicmachinery.com/2009/11/22/pauls-music-wreckommender/>
- Latour B (1988) Mixing humans and nonhumans together: The sociology of a door-closer. *Social Problems* 35: 298–310.
- Levy S (2010) How Google's algorithms rule the web. *Wired*. Available at: <http://www.wired.com/2010/02/how-googles-algorithm-rules-the-web/>
- McCourt T and Rothenbuhler EW (1997) Soundscan and the consolidation of control. *Media, Culture & Society* 19: 201–218.
- Madrigal AC (2014) How Netflix reverse engineered Hollywood. *The Atlantic*. Available at: <http://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/>
- Miege B (1979) The cultural commodity. *Media, Culture & Society* 1: 297–311.
- Moor L (2012) Beyond cultural intermediaries? A socio-technical perspective on the market for social interventions. *European Journal of Cultural Studies* 15: 563–580.
- Morris JW (2012) Making music behave: Metadata and the digital music commodity. *New Media & Society* 14: 850–866.
- Morris JW (2015) Anti-market research: Piracy, new media metrics, and commodity communities. *Popular Communication* 13: 32–44.
- Napoli PM (2013) The algorithm as institution: Toward a theoretical framework for automated media production and consumption. In: *Media in transition 8 conference*, Massachusetts Institute of Technology, Cambridge, MA, 3–5 May.
- Negus K (2002) The work of cultural intermediaries and the enduring distance between production and consumption. *Cultural Studies* 16: 501–515.
- Nixon S and Gay PD (2002) Who needs cultural intermediaries? *Cultural Studies* 16: 495–500.
- Ocejo RE (2012) At your service: The meanings and practices of contemporary bartenders. *European Journal of Cultural Studies* 15: 642–658.
- Peters JD (2008) The oldness of new media. In: *22nd Annual B: Aubrey Fisher Memorial Lecture*, Department of Communication, University of Utah, Salt Lake City, UT, 23 October.
- Platoni K (2006) Pandora's Box: Can a company's musicological data mining breathe new life into the music industry? *East Bay Express*, 11 January. Available at: <http://www.eastbayexpress.com/oakland/pandoras-box/Content?oid=1080116>
- Powers D (2014) Lost in the shuffle: Technology, history, and the idea of musical randomness. *Critical Studies in Media Communication* 31: 244–264.
- Powers D (2015) Intermediaries and intermediation. In: Bennett A and Waksman S (eds) *The Sage Handbook of Popular Music*. Thousand Oaks, California: Sage Publications Ltd, pp.120–134.

- Razlagova E (2013) The past and future of music listening: between freeform DJs and recommendation algorithms. In: Loviglio J and Hilmes M (eds) *Radio's New Wave: Global Sound in the Digital Era*. New York: Routledge, pp.62–76.
- Seaver N (2013) Knowing algorithms. In: *Media in transition 8 conference*, Massachusetts Institute of Technology, Cambridge, MA, 3–5 May.
- Skov L (2002) Hong Kong fashion designers as cultural intermediaries: Out of global garment production. *Cultural Studies* 16: 553–569.
- Smith Maguire J and Matthews J (2012) Are we all cultural intermediaries now? An introduction to cultural intermediaries in context. *European Journal of Cultural Studies* 15: 551–562.
- Smythe DW (1981) On the audience commodity and its work. In: Durham MG and Kellner D (eds) *Media and Cultural Studies: Keyworks*. Oxford: Blackwell, 2001 ed., pp.253–279.
- Soar M (2002) The first things first manifesto and the politics of culture jamming: towards a cultural economy of graphic design and advertising. *Cultural Studies* 16: 570–592.
- Steiner C (2012) *Automate This: How Algorithms Came to Rule Our World*. New York: Penguin.
- Striphas T (2012) What is an Algorithm? *Culture Digitally*. Available at: <http://culturedigitally.org/2012/02/what-is-an-algorithm/>
- Uricchio W (2011) The algorithmic turn: Photosynth, augmented reality and the changing implications of the image. *Visual Studies* 26: 25–35.
- Whitman B (2012) How well does music predict your politics? *Personal Blog*. Available at: <http://notes.variogr.am/post/26869688460/how-well-does-music-predict-your-politics>

Biographical note

Jeremy Wade Morris is an Assistant Professor of Media and Cultural Studies in the Communication Arts Department at the University of Wisconsin–Madison. He holds a PhD in Communication Studies from the Art History and Communication Studies department at McGill University. His research interests include the digitization of cultural goods and commodities, the protocols of digital culture, and the state of the popular music industry.