12-16-2022 1:00 PM

Folk Theories, Recommender Systems, and Human-Centered Explainable Artificial Intelligence (HCXAI)

Michael Ridley, *The University of Western Ontario*

Supervisor: Burkell, Jacquelyn, *The University of Western Ontario*
: Rothbauer, Paulette, *The University of Western Ontario*
: Taylor, Graham, *University of Guelph*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Library & Information Science

Abstract

This study uses folk theories to enhance human-centered "explainable AI" (HCXAI). The complexity and opacity of machine learning has compelled the need for explainability. Consumer services like Amazon, Facebook, TikTok, and Spotify have resulted in machine learning becoming ubiquitous in the everyday lives of the non-expert, lay public. In an age of "surveillance capitalism" (Zuboff, 2018), effective HCXAI is a critical component of consumer confidence and trust. The following research questions inform this study:

> What are the folk theories of users that explain how a recommender system works?
>
> Is there a relationship between the folk theories of users and the principles of HCXAI that would facilitate the development of more transparent and explainable recommender systems?

Using the Spotify music recommendation system as an example, 19 Spotify users were surveyed and interviewed to elicit their folk theories of how personalized recommendations work in a machine learning system. The results of the survey were statistically analyzed, and the interviews were probed using thematic analysis techniques. Seven folk theories emerged: complies, dialogues, decides, surveils, withholds and conceals, empathizes, and exploits. The folk theories describe beliefs about agency, power, process, intent, and relationships.

These folk theories support, challenge, and augment the principles of HCXAI. Taken collectively, the folk theories encourage HCXAI to take a broader view of XAI. The questions and concerns implicit in the folk theories indicate that users have explanatory issues that extend beyond model veracity and accountability.

The objective of HCXAI is to move towards a more user-centered, less technically focused XAI. The elicited folk theories indicate that this will require adopting principles that include policy implications, consumer protection issues, and concerns about intention and the possibility of manipulation. As a window into the complex user beliefs that inform their interactions with Spotify, the folk theories offer insights into how HCXAI systems can more effectively provide machine learning explainability to the non-expert, lay public.

Keywords

Folk theories, mental models, explainable artificial intelligence (XAI), human-centered explainable artificial intelligence (HCXAI), recommender systems, Spotify, explanations, machine learning.

Summary for Lay Audience

This study uses folk theories to enhance how artificial intelligence systems (AI) explain their behaviours. The results allow machines to better understand humans and humans to better understand machines. Folk theories are the beliefs people hold about how a system works. These beliefs aren't necessarily fully accurate, but they must be functional if they are to help people use a particular system effectively. This study gathered the folk theories of the Spotify music system, an advanced AI system that provides personalized music recommendations for its users.

Consumer services like Spotify, Amazon, Facebook, and TikTok are ubiquitous in our everyday lives. These services use machine learning techniques, the leading-edge of AI. Machine learning systems are powerful, complex, consequential, and opaque. They have difficulty explaining their actions and as a result users have difficulty fully trusting them. The field of "explainable AI" (XAI) is about enabling machine learning systems to tell us what they did, why they did it, and why they didn't do something else instead. The principles of human-centered explainable AI (HCXAI) place the non-expert, lay public at the center of the AI explainability challenge.

The folk theories of Spotify users describe beliefs about agency, power, process, intent, and relationships. These folk theories support, challenge, and augment the principles of HCXAI. Taken collectively, the folk theories encourage HCXAI to take a broader view of XAI. The questions and concerns implicit in the folk theories indicate that users have explanatory issues that extend beyond how and why the system works. The objective of HCXAI to move towards a more user-centered, less technically focused XAI means adopting principles that include policy implications, consumer protection issues, and concerns about intention and the possibility of manipulation. As a window into the complex user beliefs that inform their interactions with Spotify, the folk theories offer insights into how HCXAI systems can more effectively provide machine learning explainability to the non-expert, lay public.

Dedication

To my father, Wilfred John Ridley (1913-1993), and my mother, Peggy Suzanne Ridley (1919-2003), for the best and most enduring gift, a love of learning.

"We can only see a short distance ahead, but we can see plenty there that needs to be done."
Alan Turing, *Computing Machinery and Intelligence* (1950)

Acknowledgments

My thanks to Carol Shepstone and Fangmin Wang at the Toronto Metropolitan University Library where I was appointed a Visiting Scholar and to the Vector Institute where I was privileged to be a Postgraduate Affiliate for two years. Each of these experiences enriched my research and my life.

There are so many people who influenced me and to whom I owe so much. There are too many to name here, but it is also too difficult not to acknowledge a few:

> Margaret Beckman, John Black, Dorothy Fitzgerald, Sally Grande, Christine Harrison, Bill Kurmey, Katherine LeButt, Maureen Mancuso, Ron MacKinnon, Donald Rowan, Murray Shepherd, Alastair Summerlee, and Bruce Whiteman.

Many thanks to my PhD cohort (kirstyn seanor, Sarah Cornwell, Alex Mayhew, and Toluwase Asubiaro) for their support and inspiration, and to Danica Pawlick Potts for her assistance with thematic coding and for our many conversations and collaborations about artificial intelligence. Amazing people, each of them.

I want to acknowledge the support of the staff at FIMS who assisted me and smoothed the way. These are remarkable people dedicated to helping students and faculty achieve their goals.

To my supervisor and the other members of my Advisory Committee, Dr. Jacquelyn Burkell, Dr. Paulette Rothbauer, and Dr. Graham Taylor (University of Guelph):

Paulette was the first to encourage me to undertake this journey and was an engaged and supportive mentor throughout. Her caring approach to the joys, trials, and tribulations of PhD work sustained me and enriched my experience.

Graham is an extraordinary connector and polymath. His guidance and encouragement have introduced me to new colleagues and experiences, and led me to explore more deeply and critically. His generosity is boundless.

As my supervisor, Jacquie has been nothing short of exceptional. She has challenged, supported, questioned, and enthusiastically agreed. She has opened doors and, thankfully, kept others at bay. Jacquie's attention to detail, textually and conceptually, have done what a good PhD experience should accomplish: I think differently now. It has been an honour and a true privilege to work with Jacquie.

Most of all, my whole heart goes to the loves of my life, my wife Lynn Ridley and our daughters Jane and Sarah. Fajah couldn't have done this without your love and support.

Finally, a wish for our granddaughters, Charlotte, Reese, and Annie, that a love of learning be my gift to you.

Table of Contents

List of Tables

List of Figures

List of Appendices

1 Introduction

1.1 Purpose of the Study

Machine learning systems are ubiquitous, complex, and increasingly consequential in everyday life. For users to understand, trust, and manage them, these systems must be able to explain their decisions, recommendations, predictions, and processes. Derived from the larger field of explainable artificial intelligence (XAI), Human-centered XAI (HCXAI) emphasizes the explanatory needs of the non-expert public who are users of machine learning systems such as Spotify. In this qualitative study, the elicited folk theories (aka mental models) of users of the Spotify music recommender system (www.Spotify.com) are used to inform and enhance the principles of HCXAI.

1.2 Background

Folk theories of recommender systems tell us what users believe about how these machine learning systems work. These theories, accurate or not, contain subjective perspectives about that technology (Norman, 1983b; Payne, 2003). Folk theories are explanatory and "they must be functional" (Norman, 1983b, p. 7), allowing people to successfully use the system. Insights into how users, particularly the lay population, perceive recommender systems could be helpful in making transparent and explainable systems by informing the principles of HCXAI that guide explanatory system development. One of the challenges of explainability for public-facing machine learning systems is that the folk theories of users may present obstacles to explication, acceptance, and trust.

The field of explainable AI (XAI) consists of a set of techniques, processes, and strategies that facilitate explanations of how machine learning systems make decisions and recommendations (Adadi & Berrada, 2018; Biran & Cotton, 2017; Miller, 2019; Mohseni et al., 2021; Mueller et al., 2019). These explanations may be highly technical or broadly conceptual, and they may be designed for specific users such as technologists, professionals, and policy makers. However, the explanations that address the needs of the public are in many ways the most complex and the most urgent. Recently, as AI research has embraced the importance of user studies in XAI, it recognized that folk theories "can powerfully shape how machines are developed and how these machines are ultimately perceived by users" (Bonnefon & Rahwan, 2020, p. 1019). The focus of HCXAI is to understand the perspectives of these users and to devise and implement methods that facilitate explainability.

Recommender systems, such as those used by Amazon, Spotify, Netflix, Facebook, and Google, employ machine learning techniques to provide personalized "suggestions for items to be of use to a user" and to allow "users to cope with information overload and help them making better choices." (Ricci et al., 2015, p. vii). These systems are the "public face" of artificial intelligence because "our entanglement with algorithmic personalization is non-negotiable: it is a market driven pre-condition of the digital everyday" (Kant, 2020, p. 214). As complex and opaque systems, recommender systems are a widely used example of machine learning that require explanations. The non-expert, lay public have extensive experience with these systems and likely have developed folk theories that reflect their beliefs about how they work. Uncovering these folk theories allow for an exploration of how they might inform explanations.

Library and information science (LIS) has a longstanding and continuing interest in both intelligent information systems (Aluri & Riggs, 1988; Cox et al., 2019; Griffey, 2019; Hsieh & Hall, 1989; G. Liu, 2011; L. C. Smith, 1976, 1989) and folk theories (Cho, 2018; Cole et al., 2007; Han et al., 2020; He et al., 2008; Lin et al., 2012; Makri et al., 2007; Michell & Dewdney, 1998; Seadle, 2003) as aspects of information retrieval, information behaviour, user experience, and system design. However, XAI has received limited attention in the LIS literature with most work primarily identifying the importance of XAI and providing examples of the roles for the LIS community (Bunn, 2020; Cordell, 2020; Cox, 2021; Gasparini & Kautonen, 2022; Johnson, 2020; Lippincott, 2020; Østerlund et al., 2021; Padilla, 2019; Ridley, 2019, 2022). Uniquely, this study brings together machine learning systems, folk theories, and HCXAI to explore the influence that user folk theories of recommender systems can have on machine learning explainability.

1.3 Significance of the Study

As greater portions of our lives are "algorithmically mediated" (J. Anderson, 2020), "the danger is not so much in delegating cognitive tasks, but in distancing ourselves from–or in not knowing about–the nature and precise mechanisms of that delegation" (de Mul & van den Berg, 2011, p. 59). Applying the insights from folk theories to the principles of HCXAI can help machine learning developers create better systems, educators address algorithmic literacy, policy makers devise consumer protection, and the public navigate the complexities of using these systems.

1.4 Research Questions

The results of this research can inform HCXAI in the specific context of recommender systems, but they cannot be generalized more broadly to all machine learning systems. However, it is anticipated that general principles or approaches will emerge that can be adapted to other contexts (e.g., different systems, domains, and user groups) making these results valuable if not directly transferable.

The following research questions inform this study:

What are the folk theories of users that explain how a recommender system works?

Is there a relationship between the folk theories of users and the principles of HCXAI that would facilitate the development of more transparent and explainable recommender systems?

1.5 Structure of the Thesis

In Chapter 2, a literature review provides a general review of folk theories followed by a specific discussion of folk theories in the context of AI and explanation. A general review of recommender systems is followed by a specific discussion of folk theories in the context of recommender systems. The section on explainable AI (XAI) provides an overview of this field with a detailed exploration of human-centered XAI (HCXAI). The principles are discussed and specific issues regarding HCXAI are explored. The section following reviews the characteristics of a good explanation in the context of HCXAI. The last section in this chapter brings together a discussion of folk theories, machine learning systems, and XAI. It describes three key papers regarding folk theories and algorithmic systems. and four papers regarding folk theories and HCXAI. The purpose of this chapter is to describe and position three key areas of focus in this study: folk theories, recommender systems, and HCXAI.

Chapter 3 provides a detailed discussion of the Spotify music streaming and recommendation system. This includes an historical context, objectives of the system, data collected and used by the system, and technical details about how the recommendation process works. The purpose of this chapter is to provide a description and analysis of the Spotify recommender system as context for the elicited folk theories of users.

Chapter 4 discusses the difficulties of eliciting folk theories. The challenges of conducting research on folk theories are reviewed, followed by the mitigating strategies employed in this study.

Chapter 5 outlines the research methodology used in the study. An overview of the Spotify user survey and interviews is provided. The recruitment strategy for participants is described. Specific details on each method are described in subsequent chapters: user survey (Chapter 6) and user interviews (Chapter 7). The limitations of the study are discussed, and the ethical considerations reviewed.

Chapter 6 provides a detailed discussion of the Spotify user survey. The nature and content of the survey is described. The data are analyzed, including a quantitative analysis and a factor analysis using Q methodology, and the overall findings are presented. The purpose of this chapter is to describe the results of the user survey and to present the preliminary folk theories arising from the survey.

Chapter 7 provides a detailed discussion of the Spotify user interviews. The nature and context of the interviews are discussed. The data are analyzed using content and thematic analysis, and the findings are presented. Elements of folk theories are synthesized from the analysis. The purpose of this chapter is to describe the results of the user interviews and to present user observations and beliefs augmenting and extending the folk theories from the survey.

Chapter 8 provides a synthesis of the folk theories of the Spotify music recommender system that were elicited from the user survey and interviews. Given the resulting folk theories, a section discusses whether explanations are important to Spotify and Spotify users. The purpose of this chapter is to provide a detailed description of the folk theories of users of the Spotify music recommender system that will be used to inform and enhance the principles of HCXAI.

In Chapter 9, the folk theories of Spotify users are discussed in relation to the principles, challenges, and issues of human-centered explainable AI (HCXAI). Folk theories are shown to support, challenge, and augment the HCXAI principles. The purpose of this chapter is to document how folk theories can be used to address the objectives of HCXAI by enhancing the principles that guide the developers of XAI systems.

Chapter 10 revisits the research questions guiding this study and addresses the study’s contributions, implications, and limitations. Areas for future research are considered.

2 Literature Review

2.1 Folk Theories

The terms folk theories and mental models have generally been used interchangeably to refer to “the mental representations that humans use to structure experience” (Gelman & Legare, 2011, p. 380). They allow people to “systematically investigate what [they] believe to be true about particular domains” and provide “a mental structure of possible states of the world that the user can search in order to plan their behavior” (Payne, 2003, p. 152). Importantly, they are “not neutral or passive snapshots of experience; they embody cognitive biases that influence thought and action” (Gelman & Legare, 2011, p. 380).

Mental models and folk theories are “surprisingly meager, imprecisely specified, and full of inconsistencies, gaps, and idiosyncratic quirks” (Norman, 1983b, p. 8) and yet they are centrally “causal and explanatory” (Gelman & Legare, 2011, p. 380) and “must be functional” (Norman, 1983b, p. 7). They are important because of their “utility for the user, rather than their verisimilitude” (Hamilton et al., 2014, p. 638).

However, they have been viewed as “a tantalizing—rather than fulfilling—theoretical concept” (Payne, 2003, p. 135) in part as a result of differing definitions of their nature and scope (Allen, 1997; DeVito et al., 2017; Gelman & Legare, 2011; Moray, 1996; Norman, 1983b; Payne, 2003; Staggers & Norcio, 1993). Compounding this definitional breadth is the difficulty in eliciting the substance and nuances of folk theories because of the limitations of research methods (Doherty & Doherty, 2018; He et al., 2008; Norman, 1983b; Payne, 2003).

Despite this, mental models and folk theories have remained a key means for understanding and improving human-computer interaction (HCI), user experience, and systems design (Doherty & Doherty, 2018). The recommendations from a 1987 report about folk theories and software development established the research agenda for this field which remains pertinent (National Research Council. Committee on Human Factors, 1987). Key among these was the recognition that folk theories are not merely mechanistic or procedural. Folk theories have been used to align system specifications with user needs, to develop interfaces that conform to user expectations, and to identify gaps in user understanding of system behaviours and processes (Young, 2008). Folk theories surface belief systems (acquired through observation, instruction or inference), observability (a correspondence between operation of the system and user observation), and predictive power (understanding and anticipating the behaviour of the system) in accordance with the "antecedents" (information, beliefs, motivation) of attribution theory (Doherty & Doherty, 2018; Kelley & Michela, 1980; Norman, 1983b).

This study uses the term folk theories to refer to both folk theories and mental models. While some see mental models as "functional blueprints" and view folk theories as "less formal, less mechanistic, and more expansive, representing guiding beliefs", folk theories and mental models are viewed as "close cousins" (DeVito et al., 2018, p. 120). For example, French & Hancock's (2017) use of the term folk theories with respect to cyber-social systems echoes descriptions and characteristics identified by others describing mental models (Carroll & Olson, 1988; Norman, 1983b). Viewed exclusively in the context of algorithmic systems, Bucher calls these characteristics the "algorithmic imaginary" (Bucher, 2017, 2018): "the algorithmic imaginary is not to be understood as a false belief or fetish of sorts but, rather, as the way in which people imagine, perceive and experience algorithms and what these imaginations make possible" (Bucher, 2017, p. 31). Bucher distinguishes the algorithmic imaginary from folk theories by noting their "productive and affective power" which enables people to act on and influence algorithms in addition to being passive recipients (Bucher, 2017, p. 41). While Gelman & Legare are clear that folk theories are both passive and active, and can "influence thought and action" (Gelman & Legare, 2011, p. 380), Bucher adds a new dimension by identifying the affective aspect of folk theories.

While it is possible to differentiate these terms (Ytre-Arne & Moe, 2021), within the context of recommender systems and machine learning, folk theories, mental models, and algorithmic imaginaries are effectively synonyms. To align with the focus on recommender systems, this study uses the technologically specific definition of folk theory as “intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems” (DeVito et al., 2017, p. 3165).

2.2 Explainable AI (XAI)

Lacking “a theory of explainable AI, with a formal and universally agreed definition of what explanations are” (Samek & Muller, 2019, p. 17), the fundamentals of this field are still being explored, often from different disciplinary perspectives (Mueller et al., 2019). The lack of an accepted XAI definition (Palacio et al., 2021; Samek & Muller, 2019; Verma et al., 2021; Vilone & Longo, 2020) is confounded by the related concepts of interpretability, transparency, and traceability (Lipton, 2016; Mohseni et al., 2021; Walmsley, 2021). This has resulted in what Lipton calls “a surfeit of hammers, and no agreed-upon nails” (Lipton, 2017).

However, according to the widely referenced US Defence Advanced Research Projects Agency (DARPA) description, the purpose of XAI is for AI systems to have “the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future” (DARPA, 2016) and to “enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners” (Turek, 2016). XAI is a set of strategies, techniques, and processes that include testable and unambiguous proofs, various verification and validation methods that assess influence and veracity, and authorizations that define requirements or mandate auditing (Abdul et al., 2018; Adadi & Berrada, 2018; Das & Rad, 2020; Molnar, 2018; Mueller et al., 2019; Vermeire et al., 2021).

While XAI has received limited attention in the LIS literature, AI researchers have acknowledged the narrow disciplinary perspectives that have guided much AI work and are actively encouraging other fields to apply their methods and insights to machine learning and XAI in particular (Mueller et al., 2019; Saeed & Omlin, 2021). A systematic meta-survey of the challenges and research directions of XAI identified thirty-nine key issues including the need for multidisciplinary collaboration, understanding and enhancing the user experience, matching XAI to user expertise, and explaining the competencies of AI systems to users (Saeed & Omlin, 2021).

2.2.1 Purpose and Consequences of XAI

2.2.1.1 The Right to Explanation

While explainability has been an issue for AI since the earliest days of expert systems in the late 20th century (Clancey, 1983; Swartout, 1983; Swartout et al., 1991), the European Union's 2018 General Data Protection Regulation (GDPR), first tabled in 2016, transformed explainability and XAI from a technical issue to a public policy priority (European Union, 2016). However, the existence of a "right to explanation" in the GDPR is ambiguous. For some it is clear (Goodman & Flaxman, 2017; Kaminski, 2019), while for others is it not only absent from the regulation, it is not the preferred strategy for algorithmic accountability (Edwards & Veale, 2017, 2018; Wachter et al., 2017).

Opposition to the right to an explanation came from pro-business groups like the Centre for Data Innovation who were concerned about the negative effect on innovation and competitive advantage for the EU and North America (Wallace, 2017; Wallace & Castro, 2018), and from the AI research community who felt an explainability requirement would impact system performance (Gunning & Aha, 2019; Kozyrkov, 2018; Lipton, 2016; Sokol & Flach, 2020). Despite this, IBM and other leading AI research and development companies and organizations could see the rising imperative and initiated

XAI programs (Rossi, 2016). Since then XAI research and development has exploded (Adadi & Berrada, 2018; Mueller et al., 2019; Vilone & Longo, 2020) and many jurisdictions beyond the EU have explored the idea of a right to explanation, including Canada through consultations with the federal (Canada. Office of the Privacy Commissioner of, 2020) and provincial (Ontario, 2021) privacy commissioners.

The requirements for explainability, arising from public advocacy (Broussard, 2018; Campolo et al., 2017), consumer protection (Consumers International, 2019), regulation (Canada, 2019; Casey et al., 2019; European Union, 2016; Goodman & Flaxman, 2017), and professional codes of conduct (Association for Computing Machinery, 2017; FAT/ML, n.d.; IEEE, 2019) have further accelerated a focus on explainable AI generally and XAI specifically.

Despite the acceptance and importance of XAI, some within the machine learning community believe it holds the field to "an unrealistically high standard" (Zerilli et al., 2019, p. 661) because "the sorts of explanations we cannot obtain from AI we cannot obtain from humans either" (Zerilli et al., 2019, p. 680). Geoffrey Hinton argues that requiring an explanation from an AI system would be "a complete disaster" and that trust should be based on the system's performance not its explainability (Simonite, 2018).

2.2.1.2 Manipulation and Deception

Assuming the use of XAI is benign ignores situations where explanations can "suppress curiosity and reinforce flawed mental models ... overwhelm people with details ... not allow questions ... make people feel reticent, [and] include too many open variables and loose ends" (Hoffman et al., 2019 p.17). It is a reminder that one of the goals of an explanation can be persuasion (Mueller et al., 2019). Some view XAI as a deceptive means to further "surveillance capitalism" (Zuboff, 2018):

> "Explainable AI is guaranteed to fail as a means of empowering individuals to effectuate their value preferences in their interactions with AI, but that in failing it will succeed in legitimizing new regimes of control, including the expansion of punitive and surveillance-based AI." (Knowles, 2022)

Post-hoc explanations, including counterfactuals and feature selections, are susceptible to manipulation (Lipton, 2016; Slack et al., 2021; Wachter et al., 2018) and can be used to promote over-reliance on algorithmic decisions or recommendations (Smith-Renner et al., 2020). Changes in data distribution, inadvertent or deliberate, can have a similar distorting effect (Woźnica et al., 2021).

Provocatively, Kim et al. ask whether AI should be able to lie (i.e., deliberately manipulate users) noting that deception is often part of negotiations and not always viewed as unethical (T. W. Kim et al., 2021). The authors conclude with a framework to guide when an AI may be permitted to lie. However, XAI manipulation can also be subtle. Prasad et al. assume that we want "models that behave in ways that people expect them to. We want models that are aligned to human expectations" (Prasad et al., 2020). While generally desirable, these objectives could rely on confirmation bias to reassure users of potentially manipulative outcomes.

Research by Schneider et al. confirmed that intentionally deceptive explanations embedded in machine learning can fool humans, but they also note that,

> one can deploy machine learning methods to detect seemingly minor deception attempts with accuracy exceeding 80% given sufficient domain knowledge. Without domain knowledge, one can still infer inconsistencies in the explanations in an unsupervised manner, given basic knowledge of the predictive model under scrutiny (Schneider et al., 2020).

XAI can be both a source of and a solution to manipulation and deception in explanations. Interactive, prospective (e.g., ante-hoc) exploratory systems are proposed to "enable users to investigate how their choices affect the outcome" (Shneiderman, 2022, p. 169) and providing a tool to assist in assessing persuasive or manipulative intent.

2.2.1.3 Consumer Protection

The objectives and requirements of XAI can be viewed through the lens of consumer protection. Users of recommender systems are consumers of a commercial product and might reasonably expect protection in its use as with any other product. Consumer protection, often covered in different but interrelated regulations and legislation, focuses on privacy, risk, and harm, and includes complaint and mitigation processes.

At present, “consumers are largely incapable of exercising meaningful choice with regard to the rapidly expanding array of points within the matrix of networked interactions in which their interests will be placed at risk” (Gandy, Jr., 2011, p. 185). While this suggests the need for consumer protection beyond what currently exists, others think “technological innovation need not necessarily disrupt the extant legal regime of consumer protection” (Howells, 2020, p. 146). While “appropriate modification” may be necessary, “consumer protection seems able to retain its traditional values and in many instances its traditional form” (Howells, 2020, p. 171).

A Consumers International study of consumers from India, Australia, and Japan regarding their understanding and use of AI enabled consumer services revealed their “enthusiasm and appreciation” for these technologies but concerns about who is “behind things and how data is used” (Consumers International, 2019, p. 3). The study noted,

> There is widespread confusion about recourse concerning AI enabled services. The notion that one could complain or even obtain an explanation for what is considered to be a problem often seems far from our participants' minds. Part of the problem seems to be the sense of ‘inevitability’ of technology having its own trajectory and as such there is a limit to how it can be changed (Consumers International, 2019, p. 22).

Calling this “learned helplessness” (Consumers International, 2019, p. 27), the organization called for “a duty to explain” on the part of AI enabled services (Consumers International, 2019, p. 31).

Few studies have examined XAI specifically from a consumer perspective. A study of consumer preferences regarding the nature of XAI explanations identified three findings: the importance rankings over counterfactuals, a tolerance for only a moderate level of complexity regardless of context, and a desire of explanatory specificity dependent on context (Ramon et al., 2021). Notable here is the preference of rankings over counterfactuals as an XAI technique. Multiple explanations are a common XAI outcome since they can vary in terms of scope, detail, and perspective (e.g., global or local). The ranking of explanations, rather than “what if” alternatives, provides the user with degrees of confidence or applicability that are more directly relevant.

Regulatory efforts to link XAI to consumer protection are rare. In a study of consumer attitudes towards sustainable and transparent, explainable AI, the authors conclude that “it seems doubtful that simply placing the burden on ‘the informed consumer’ will lead to a demand for transparent and sustainable AI” (König et al., 2022, p. 10). A statement preferencing competition ahead of consumer protection from the Federal Trade Commission, which oversees such regulation in the US, suggests that governments are a doubtful source of protection as well (Howells, 2020). Progressive legislation proposed in the US Senate (*Algorithmic Accountability Act*, 2022) appeals frequently to consumer protection but restricts its oversight to a select number of applications where “critical decisions” are at stake. Regulations derived from this bill could easily exclude consumer services such as Spotify from this protection.

A weakness in XAI is the lack of a consumer perspective and the resulting need to link XAI to the broader issues of consumer protection.

2.3 What Makes a Good Explanation?

A prerequisite to effective XAI is understanding what makes a good explanation. Explanations "are more than a human preoccupation – they are central to our sense of understanding, and the currency in which we exchange beliefs" (Lombrozo, 2006, p. 464).

There are five reasons for an explanation: 1) to predict similar events in the future, 2) to diagnose, 3) to assess blame or guilt, 4) to justify or rationalize an action, and 5) for aesthetic pleasure (Keil, 2006). These different objectives highlight that good explanations are recipient dependent and recipient sensitive: "Every theorist of explanation can admit that the idea of a good explanation is audience-variant" (Ruben, 2012, p. 19) and that "a good explanation is one that meets the interests, and assumes what it should assume about the beliefs, of the audience" (Ruben, 2012, p. 26). There is, between the explainer and the explainee, a shared level of general knowledge and previous, shared experiences (e.g., previous interactions or explanations) upon which to build an explanatory dialogue. This puts the explainee and their interests, motivations, and capabilities at the forefront of a good, and satisfactory, explanation. As Mayes notes: "All questions are interest driven, but explanation-seeking questions seem especially so. For not only do individual interests generate the question, but to a surprising extent they seem to determine what counts as an acceptable answer" (Mayes, 2000, p. 368).

Miller, in reviewing the literature on human explanations, identified four major findings that are important for XAI: "(1) why-questions are contrastive; (2) explanations are selected (in a biased manner); (3) explanations are social; and (4) probabilities are not as important as causal links" (Miller, 2019, p. 34). When people want to know the "why" of something "people do not ask why event P happened, but rather why event P happened *instead* of some event Q" (Miller, 2019, p. 3). The selective, social, and causal findings reinforce that "explanations are not just the presentation of associations and causes (*causal attribution*), they are *contextual*" (Miller, 2019, p. 3). However, the recommendation that explanations are contrastive (i.e., counterfactual) may not be as applicable in a low or non-consequential service such as Spotify (Ramon et al., 2021).

When an explanation is provided is also significant. Explanations can be *ex post*, *ex ante*, or interactive. In the first case, one receives an explanation following a decision. This is a justification of a decision. In the second case, an explanation precedes a decision. This occurs when the decision process is transparent. In the third case, a dialogue is established whereby an explanation is achieved simultaneously with the decision. This is effective for complex questions or in consequential settings. When an explanation is provided is contingent on the type of explanation being requested and the nature of the explanation process.

Good explanations can be abstract and generalized (Keil, 2006; Strevens, 2008) or concrete and detailed (Bechlivanidis et al., 2017). Garfinkel cautions that "explanations that are too concrete are not merely 'too good to be true' (i.e., impractical) but rather 'too true to be good'" (Garfinkel, 1981, p. 58). However, Tesler's Law of the Conservation of Complexity states that "there is a point beyond which you can't simplify the process any further; you can only move the inherent complexity from one place to another" (Saffer, 2010, p. 139). While contextually dependent, it may be that an explanation can "satisfice" when it "meets a variety of criteria but maximizes none" (Simon, 1992, p. 157).

Abstract explanations have been shown to overcome the inherent complexity of AI reducing barriers for users (Kizilcec, 2016; Kuleza et al., 2013; Tullio et al., 2007). Other research demonstrated that "rich explanations of the recommender system increased participants' understanding of the recommendation process, and this in turn improved their beliefs about the quality of the recommender system's performance" (Yeomans et al., 2019, p. 411). To overcome the reluctance of users to engage in complex explanations, some researchers have proposed "cognitive forcing" techniques to oblige user attention (Buçinca et al., 2021). Regardless of the approach and level of detail, explaining how a system works has been shown to improve (i.e., make more accurate) user understanding (Cramer, 2010; Kulesza et al., 2012).

How explanations are presented to the user is an important element of XAI and a highly contentious area of research. Multimedia visualizations have been proposed to exploit their information rich contexts (Abdul et al., 2018; Diederich, 2018), although visual explanations of deep learning have been found to be poorly understood by end users (Hohman et al., 2019). A 2017 survey of explanations for recommender systems found that textual explanations were overwhelmingly favoured over visualizations, lists, arguments, traces or other techniques (Nunes & Jannach, 2017). Another study found that simple Venn diagrams were preferred over other visual and textual explanations (Kouki et al., 2017). Others propose interactive methods (Sokol & Flach, 2020) and example-based explanations (Jeyakumar et al., 2020).

Also relevant is whether a global or local explanation is required or requested. A global explanation responds to the question, "how does the system work?" A local explanation responds to the question, "why did the system make that particular recommendation?" The former helps to "understand and evaluate the model" and "render more confidence in understanding the model" while the latter "scrutinize[s] individual cases" and exposes "discrepancies between different cases" (Dodge et al., 2019, p. 283). Some have argued that from an explainee perspective the categorization of explanations based as local or global is a difference without a distinction since both approaches have problematic implications for generalizability (Jacovi et al., 2022).

Given these requirements for good explanations, utilizing folk theories for effective XAI may require the type and extent of user personalization already enabled through recommender systems. This acknowledges that "different explanations are needed for different people [and] different mental models will require different explanations" (Kühl et al., 2020).

2.4 Human-Centered Explainable AI (HCXAI)

Human-centered explainable AI (HCXAI) is a specific area of XAI that responds to the DARPA definition with a focus on explainable AI for the lay, non-expert public. HCXAI has been widely discussed (Chari et al., 2020; Ehsan et al., 2022; Ehsan & Riedl, 2020; Liao & Varshney, 2022; Shen & Huang, 2021; Vaughan & Wallach, 2021; D. Wang et al., 2019) including the importance of folk theories (Gentile et al., 2021; Ngo & Krämer, 2021; Villareale & Zhu, 2021; D. Wang et al., 2019).

HCXAI, as a human focused view of XAI, arose in consort with a human-centered approach to AI more generally (HCAI) (Aragon et al., 2022; Shneiderman, 2022). The emergence of HCXAI was motivated by the lack of user studies, a focus on researchers and developers rather than the lay public, an almost exclusive emphasis on the technical aspects and techniques of XAI, the lack of pedagogical methods, the importance of actionable explanations, and the need to reduce the complexity of explanations. Machine learning systems are "often not tested to determine whether the algorithm helps users accomplish any goals" (Mueller et al., 2021). As a result, there have been numerous calls for more user studies of XAI (Burkart & Huber, 2021; Miller, 2019; Ribera & Lapedriza, 2019; Samek & Muller, 2019).

In particular, rather than AI researchers and system developers, more users from the non-expert public should be studied. The audience for a XAI system can be system developers (who are primarily interested in performance), clients (primarily interested in effectiveness or efficacy), professionals (primarily interested in work related outcomes), regulators (primarily interested in policy implications), and everyday users of the models (primarily interested in trust or accountability). Marginalized communities are most often ignored (Ehsan & Riedl, 2020). The nature and presentation of explanations can be significantly different depending on the audience, yet "few papers address the different AI literacy levels users may have and even fewer address the diversity of stakeholders and their different needs for XAI." (Gerlings et al., 2020, p. 5).

The emphasis on researchers and developers resulted in an emphasis on the technical strategies and techniques of XAI. As a result, too often ignored, minimized, or simply misunderstood are the issues and concerns of the everyday person as they use machine learning systems and encountered XAI systems. This study addresses this gap putting the focus on the non-expert, lay population. For the lay person in particular, explanations are a "learning process" requiring XAI to have an instructional and evaluation model as part of its design (Clancey & Hoffman, 2021). However, in most XAI research and implementations "teaching explanations tend to be forgotten" (Sheh & Monteath, 2018, p. 264). HCXAI as a pedagogical method also means "making explanations actionable" (R. Singh et al., 2021, p. 14), ensuring that users know how an explanation "is intended to be used" (Davis et al., 2020).

However, given the complexity of machine learning and XAI, researchers have cautioned that "we may be fine-tuning AI methods with elaborate bells and whistles that no human-ear can hear" (Keane et al., 2021). As a result, central to HCXAI are "efficiency" (i.e., the time it takes for the user to understand the explanation) (Rüping, 2006) and the "mental fit" with a user (Bibal & Frénay, 2016).

2.4.1 Principles of HCXAI

The principles of human-centered XAI, as presented in Mueller et al. (2021), "start with human-focused principles for the design, testing, and implementation of XAI systems, and implement algorithms to serve that purpose" (Mueller et al., 2021). Among these principles are the importance of context (regarding user objectives, decision consequences, timing, modality, and intended audience), the value of using hybrid explanation methods that complement and extend each other, and the power of contrastive examples and approaches. Developers are urged to "build explanatory systems, not explanations" recognizing the dynamic nature of intelligent systems and that XAI cannot simply be "one-off." Explanations are about "knowledge transformation and sense-making." The principles of HCXAI are presented in Table 1 along with brief descriptions extracted from the text. To highlight themes inherent in the principles, they have been categorized as key objectives (i.e., overarching principles; marked in red) and principles that have specific relevance to users, context, techniques, and evaluation. A full description of the principles is available in Appendix 1.

Table 1: Principles of HCXAI (Mueller et al. 2021)

Principle	Brief Description	Category
The property of being an explanation is not a property of statements, visualizations, or examples.	Explaining is a process [and] must be understood by a user to be effective.	Key Objective Users Evaluation
Work matters.	It is impractical to develop a useful and usable explanation system outside of a work context.	Context
The importance of active self-explanation.	Focus explanatory systems on information that empowers users to self-explain, rather than simply delivering an output of an algorithm.	Users Techniques
Build explanatory systems, not explanations.	Explanation[s] must be accompanied by other things (instructions, tutorial activities, comparisons, exploratory interfaces, user models, etc.).	Key Objective
Combined methods are necessary.	Multiple kinds of information can complement one another … both global and local explanations may be justifiable and reinforce one another.	Techniques
An explanation can have many different consequences.	Different explanations can have very different effects. The explanations should be tuned to the goal.	Users Context Techniques Evaluation
Measurement matters.	Designers should identify what conse-quences the explanation should have	Evaluation

	to develop an appropriate measurement and assessment approach.	
Knowledge and understanding are central.	The focus of explanation is on developing a better understanding of the system.	Key Objective
Context matters: Users, timing, goals	The best explanation depends on context: who the user is, what their goal is, when they need an explanation, and how its effectiveness is measured.	Users Context Evaluation
The power of differences and contrast.	First develop learning objectives for an explanatory system and identify the contrasts necessary to support those objectives.	Key Objective Techniques
Explanation is not just about transparency.	Contrast, global explanation and local justification, examples, explorable interfaces that permit hypotheticals, etc. will be necessary.	Techniques
The need for explanation is “triggered.”	Explanations are not always necessary, [they are] triggered by states such as surprise and violations of expectation.	Users Context
Explanation is knowledge transformation and sensemaking.	[It] is not just the learning or incorporation of information; it involves changing previous beliefs and preconceptions.	Key Objective Users
Explanation is never a “one-off.”	Users often need repeated explanations and re-explanations. XAI systems might benefit from considering the long-term interaction with users.	Key Objective Users

2.4.1.1 Key Objectives

The key objectives in the HCXAI principles highlight the foundational concepts that inform HCXAI as a whole and guide the other principles. Central to HCXAI is the call for “explanatory systems, not explanations.” Explanatory systems include more than expressions of veracity, confidence, or justification, they include “instructions, tutorial activities, comparisons, exploratory interfaces, [and] user models” that provide tools and a broader context for user understanding. While a key objective of HCXAI is “understanding and sensemaking,” the system approach is evident in the emphasis on explanation as a process, the need for long-term interactions with users not “one-off” events, and the principle that while information is important, “changing beliefs and preconceptions” is an expected outcome.

2.4.1.2 Users

The user focus in the HCXAI principles includes an expected emphasis on user objectives, the use of appropriate and contextual explanatory techniques, and the identification of “triggers” that prompt the need for explanations. However, the principle of “active self-explanation” offers a new strategy. Instead of “simply delivering an output of an algorithm,” the explanatory system provides information to allow the user to self-explain. This principle underscores the importance of user empowerment, information provision beyond an explanation, and a pedagogical strategy for HCXAI.

A key strength of HCXAI is to challenge dominant theories or narratives and reveal insights from marginalized communities (Ehsan & Riedl, 2020). For example, a study of XAI and accessibility looked at the needs of two largely unexplored groups with respect to XAI: aging users and users with mental health issues (Wolf & Ringland, 2019).

In the case of aging users, increasing attention was required to changes over time in the user's comprehension (technical and situational) and their levels of attention and concentration. XAI techniques responded to both cognitive decline and a user's desire to remain independent. In this way HCXAI disrupts some of the narrow thinking that has guided machine learning and XAI, and opens the field to explore issues from other disciplinary perspectives.

2.4.1.3 Context

Context is relevant to explanations regarding user objectives, timing, consequences of explanations, and "triggering" events. The principle that "work matters" indicates that "it is impractical to develop a useful and usable explanation system outside of a work context." Work in this context is a task and not a hypothetical engagement without definable goals and objectives.

2.4.1.4 Techniques

Contrasting and counterfactual explanations are prominent in the explanatory techniques mentioned in the principles and are among the most widely discussed in the XAI literature (DeVito, 2021; Miller, 2019; Mueller et al., 2021; Páez, 2019). This recognizes that users prefer "explanations for the *road not taken* — namely, why the model chose one result and not a well-defined, seemly similar legitimate counterpart" (Shen & Huang, 2021).

However, the HCXAI principles extend beyond the techniques conventionally associated with XAI by recommending "tutorial activities, comparisons, exploratory interfaces, user models." These emphasize explanation as a "learning process", support the principle of “self-explanation” and highlight the teaching role of HCXAI (Clancey & Hoffman, 2021; Sheh & Monteath, 2018).

2.4.1.5 Evaluation

Evaluation in the HCXAI principles is user focused with understanding and sensemaking as the key benchmarks. Complementing these are evaluation criteria that consider user objectives and the explanatory context. While the need for explanations may be “triggered” by specific events, the principles underscore the importance of a longer term (i.e., not a “one-off”) relationship with the user. In highlighting that different explanations have different “consequences” for the user, the principles acknowledge that explanations should be actionable. While explanations should increase understanding, they should also provide the basis for users to act on those explanations by making adjustments or changes to their use of the system.

2.5 Recommender Systems

Recommender systems “are software tools and techniques providing suggestions for items to be of use to a user ... [and] are valuable means for online users to cope with information overload and help them making better choices.” (Ricci et al., 2015, p. vii). Explanations are central to recommender systems to ensure users understand why they are receiving the recommendations these systems provide. As a result, the application of XAI techniques and strategies to recommender systems is widely discussed.

Recommender systems, as distinct from search systems, emerged in the early 1990s in order to filter newsgroups and email for the most relevant items based on user profiles (Goldberg et al., 1992; Resnick et al., 1994). Recommender systems are widely used in commercial systems (Aldrich, 2011) for such topics as travel (Bernardi et al., 2019; Bobadilla et al., 2014), books (Alharthi et al., 2018; Linden et al., 2003), movies (A. Liu et al., 2009; Steck et al., 2021; Tuzhilin & Koren, 2008; YuMin et al., 2018), research literature (Beel et al., 2013; Kudlow et al., 2019; Naak et al., 2008; C. Wang & Blei, 2011) and music (Kaminski & Ricci, 2012; McFee et al., 2012). Recommender systems are embedded in such commonly used and recognized services as Amazon, Spotify, Netflix, Facebook, Google, TikTok, and Booking.com.

Jannach et al. broaden the context of the "recommendation problem" and redefine it as the process to "find a sequence of conversational actions and item recommendations for each particular user that optimizes the overall goal over the specified timeframe" (Jannach et al., 2016, p. 102). The ubiquity of algorithmic recommendations has made recommender systems the "public face" of AI which "has settled deep into the infrastructure of online cultural life, where it has become practically unavoidable" (Seaver, 2019, p. 431).

Computational techniques used in recommender systems have evolved from basic data analytics (e.g., co-rankings) to current machine learning models (e.g., deep learning) and reinforcement learning (Bernardi et al., 2019; Jannach et al., 2011; Khanal et al., 2020; Portugal et al., 2018; Ricci et al., 2015; Singhal et al., 2017; Stål, 2021). The main approaches are content-based, collaborative, knowledge-based, or hybrid (Jannach et al., 2011; Ricoi et al., 2015), although variations and context specific approaches now include ontology-based, demographic-based, utility-based, context aware-based, trust aware-based, social-network systems, fuzzy-based systems, and group systems (Tarus et al., 2018). As with ensemble approaches in AI, recommender systems are increasingly hybrid. Innovations in the design and use of recommender systems continue because of increasing commercial adoption, larger and more diverse item databases, more explicit group and individual preferences, longitudinal models recognizing changes over time, and the application of emerging machine learning algorithms initially developed in other domains.

As the technical foundations of recommender systems innovate and evolve, a challenge to the understanding of their nature and processes is the opacity of contemporary AI. The “black box” problem arises from two general conditions: the inherent complexity of machine learning techniques applied to recommender systems (e.g., neural networks with billions of parameters) and the intellectual property (IP) protections that shield the underlying algorithms from public scrutiny (Biran & Cotton, 2017; Marcus & Davis, 2019; O’Neil, 2016; Pasquale, 2015).

Despite extensive research into recommender systems, few functional or conceptual models have been created (Jannach et al., 2016). A notable exception is a 2007 review of the recommender systems literature used to create a conceptual model of recommender systems (Xiao & Benbasat, 2007). Using the lens of decision support systems (DSS) in a consumer behaviour context, this research examines and tests 28 propositions about recommender systems as part of the model. Key dimensions of the model are: system characteristics, user characteristics, user-system interactions, product type, provider credibility, user outcomes, and user evaluation (Xiao & Benbasat, 2007, p. 140). This model has three significant limitations in the context of the present study. The DSS lens is more focused, and hence more limited, than the wider context of recommender systems. The 2007 publication date excludes the substantial increase in recommender systems research and application arising from advanced AI and broader commercial application. Finally, the authors specifically exclude consideration of algorithms as an aspect of their model. This last limitation is especially important given the centrality of algorithms to the behaviour of any recommender systems.

The present study uses the Spotify music streaming service as a representative example of recommendation systems. Spotify is discussed in more detail in Chapter 3.

2.6 Folk Theories, XAI, Explanations, and Machine Learning

Folk theories have been applied to algorithms in social media, news feeds, and online search (Bucher, 2017; Büchi et al., 2021; DeVito, 2021; DeVito et al., 2017, 2018; Eslami et al., 2016; French & Hancock, 2017; Karizat et al., 2021; Rader & Gray, 2015; Thomas et al., 2019; Toff & Nielsen, 2018; Ytre-Arne & Moe, 2021). Four studies have explored folk theories with respect to recommender systems (Colbjørnsen, 2018; Kulesza et al., 2012; Siles et al., 2020; Yeomans et al., 2019). Two of these articles, Colbjørnsen and Siles et al., focus specifically on Spotify.

Accurate folk theories have been shown to improve user performance with machine learning systems (Bos et al., 2019; Kühl et al., 2020; Kulesza et al., 2012; Schrills & Franke, 2020; Tullio et al., 2007; Villareale & Zhu, 2021). XAI has been shown to improve folk theories (i.e., make them more accurate) (Alipour et al., 2021; Cramer, 2010; Eslami et al., 2016; Graichen et al., 2022; Gunning & Aha, 2019; Kulesza et al., 2012; Ray et al., 2021). Research has demonstrated that folk theories can suggest the types and kinds of explanation most amenable and helpful to users but without specifically addressing XAI systems (Bos et al., 2019; Kühl et al., 2020; Schrills & Franke, 2020; Sonboli et al., 2021; Tullio et al., 2007). A small number of studies have suggested that folk theories can be used to improve XAI systems (Gentile et al., 2021; Ngo & Krämer, 2021; Villareale & Zhu, 2021; D. Wang et al., 2019).

The relationships among folk theories, XAI, explanations, and machine learning systems that informs most of the above research is depicted in the XAI framework (Figure 1) developed by the influential DARPA XAI research project (2017-2021) (Gunning et al., 2021).

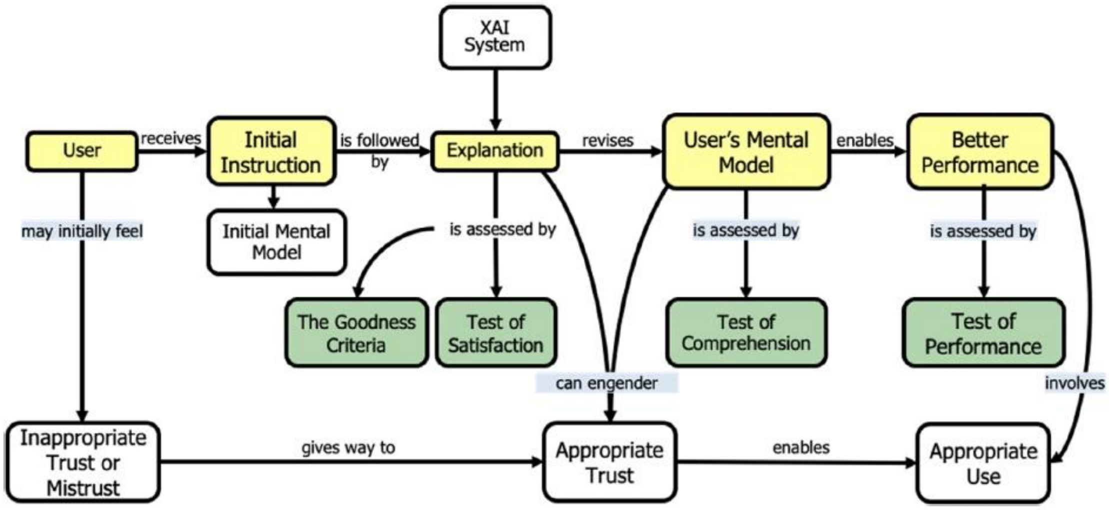


Figure 1: XAI Framework (from Gunning et al. 2021)

The DARPA funded projects found that "XAI is useful for measuring and aligning mental models for users and XAI systems" (Gunning et al., 2021, p. 8). Consistent with the framework, these projects identified how an explanation provided by an XAI system can improve a user's folk theory resulting in "better performance" through "appropriate trust" in the system and "appropriate use."

While valuable and important, what is missing from the framework, and the resulting research, is how the user's "initial mental model" can be used to influence the XAI system and the nature of the explanations provided (Koch & Fortes Rey, 2022). The link between folk theories and the XAI system, not present in the DARPA framework, allows XAI to "meet the user where they are" (DeVito, 2021, p. 339:4). The overall conclusion from a study of the folk theories of an experimental music recommender system was that "telling an end user more about how it [the recommender system] *does* work may help him or her tell the agent more about how it *should* work" (Kulesza et al., 2012, p. 10). Hoffman et al. suggest that "the evaluation of XAI systems can benefit from basically asking users to identify the triggers that motivated them to ask for an explanation" (Hoffman et al., 2019, p. 18).

2.6.1 Folk Theories of Algorithmic Systems

Three studies of folk theories of algorithmic systems are presented to illustrate differences in outcomes and themes: a study of multiple algorithmic systems (Facebook and Twitter) (French & Hancock, 2017), a study specific to Spotify (Siles et al., 2020), and a study examining general views on algorithms with no specific system referenced (Ytre-Arne & Moe, 2021). Other studies that elicited folk theories of algorithmic systems were not highlighted because they focused narrowly on Facebook (Bucher, 2017; Eslami et al., 2016), TikTok (Karizat et al., 2021), online dating (Huang et al., 2022), chatbots (Stoeckli et al., 2020), or management information systems (Tullio et al., 2007) resulting in specific contextual dependencies.

2.6.1.1 French & Hancock (2017)

A study of users of the algorithmic feeds of Twitter and Facebook identified four primary folk theories held across the platforms: the Rational Assistant, the Unwanted Observer, the Transparent Platform, and the Corporate Black Box (French & Hancock, 2017).

The Rational Assistant reflects positive beliefs that the feed “understands and prioritizes their interests.” Those who hold the Transparent Platform theory believe the feed is “unfiltered” and that curation of the feeds is largely in the control of the user. The Unwanted Observer reflects negative beliefs that the feed is “overreaching in their use of personal data to serve the company’s interests.” Those who hold the Corporate Black Box theory share the beliefs of the Unwanted Observer but add that the feed is “opaque” and “difficult to control.”

Two main themes emerge from these folk theories: agency and surveillance. Control (agency) is assigned to either the user (Rational Assistant and Transparent Platform) or the feed (Unwanted Observer and Corporate Black Box) as a surrogate for the providing company. The Unwanted Observer and Corporate Black Box believe they are being surveilled and exploited for corporate interests while the Rational Assistant and Transparent Platform believe any data gathered by the feed is used for their own interests.

2.6.1.2 Siles et al. (2020)

A study of Spotify users in Costa Rica explored folk theories from the perspective of “the platform’s place in their daily lives and social relations” (Siles et al., 2020, p. 3). Drawing from the notion of “data assemblages” which views algorithms in a sociotechnical context (Kitchin, 2017), this broader canvas for espoused folk theories resulted in more metaphorical descriptions of how “algorithmic systems” operate and affect the lives of participants.

Participants in the Siles et al. study were experienced users of Spotify who were comfortable with the operational aspects of the system and, presumably, had already formed theories of some of the structural and conceptual issues. However, including only experienced users potentially precluded novel theories from new or casual users. Many of the participants explained Spotify by making analogies to other systems (Collins & Gentner, 1987). In some cases, participants did this by declaring it the same as other systems (e.g., Netflix, Facebook, and YouTube) and in other cases by describing it as completely different (e.g., Apple Music). Spotify, and recommender algorithms more generally, were understood as part of a larger digital ecosystem.

Two main folk theories emerged. One theory personified Spotify, viewing the system as a “buddy,” a “hacker,” or even a “ghost” who surveilled users but for the beneficial purpose of music recommendation. Consistent with this view, participants attributed other characteristics to the system such as “intense,” “annoying,” “greedy,” and “intimate” (“It always knows what I want”). Through personification “users envision Spotify as a privileged social intermediary: its algorithmic recommendations are not only a way to form or strengthen a tie but also an intrinsic part of that relationship” (Siles et al., 2020, p. 6). These users explain that Spotify makes its recommendations through frequency of use, user moods during use, and the specifics of the music (i.e., characteristics of their favourite music). As a result, Spotify works by “the construction of patterns based on similarity with users who share sociodemographic characteristics with them” (Siles et al., 2020, p. 7). The authors suggest that personifying Spotify overcomes the reluctance to accept algorithmic recommendations noted in other research (Yeomans et al., 2019).

The other main folk theory views Spotify as a system of valuable resources (i.e., music) that must be trained to be personally useful. Participants use computational terms such as “control system,” “database,” and “code” to describe Spotify as a “feedback control system.” However, participants highlight that the system must be “taught” requiring “feedback” and attention to obtain the best results. This view sees Spotify as a partnership of user and system (human and technology) linked together, evolving over time, and focused on personalization.

These two folk theories view agency differently. Those who personify Spotify view its power as difficult to avoid and exerting social control over them, with users moving between “submission and resistance” (Siles et al., 2020, p. 11). Recommendations are “explicit, arrive constantly, and have neither context nor explanation” (Siles et al., 2020, p. 12). This latter observation holds even though Spotify does provide contextual cues and simple explanations (McInerney et al., 2018). Those who view Spotify computationally see a balance of power and agency between the user and the system. However, users must continually interact with the system in order to “hold up their end of the bargain” making regular use a requirement (Siles et al., 2020).

Siles et al. conclude that both folk theories center on issues of identity such that “folk theories become an expression of who the person is and wants to be seen” (Siles et al., 2020, p. 10). Engaging with Spotify’s recommendations is performative. Users explore, define, and reflect a “performed self” (Kant, 2020, p. 130) through their use of the system. Given the holistic perspective of this study, the authors stress that “people do not simply have folk theories; they have vivid stories about how they received recommendations that shaped their lives and selves" (Siles et al., 2020, p. 12). The folk theories people hold are a form of "storytelling" (Norman, 2013, p. 57). If users conceptualize folk theories as narratives, an underexplored option is to construct narrative explanations with stories as a form of abstraction (Diederich, 2018). This may be part of the strategy in using "effective anthropomorphism" as a means of helping users understand machines by endowing them with human traits (Bos et al., 2019). It also reinforces the use of play (Villareale & Zhu, 2021), a form of narrative exploration, as an explanatory strategy.

An analysis of tweets that referenced attitudes towards Spotify revealed folk theories that reinforce the findings of Siles et al. (Colbjørnsen, 2018). In these tweets Spotify users described their use of the system in personified, collaborative ways. Spotify was identified as "my algorithm" and referenced as a "personally assigned wizard" that some users were "in love with". This "pseudo-intimacy" came with a loss of agency as the tweets frequently alluded to "capitulating to the algorithm, for better or worse" (Colbjørnsen, 2018, p. 176). While not a user study in the traditional sense, the analysis of tweets does reveal that "people do communicate with algorithms, that they identify with them, argue and negotiate with them and speculate about their behaviour and characteristics" (Colbjørnsen, 2018)

2.6.1.3 Ytre-Arne & Moe (2021)

Tensions regarding agency are also evident in the folk theories about algorithms derived from a single question in a larger survey of media literacy among Norwegians (Ytre-Arne & Moe, 2021). The analysis resulted in five folk theories: "(1) Algorithms are confining, (2) Algorithms are practical, (3) Algorithms are reductive, (4) Algorithms are intangible, and (5) Algorithms are exploitative." The derived folk theories are based on an analysis of the open-ended responses to the survey question: "Do you see any negative or positive consequences of media companies' use of algorithms?"

There are important limitations to the Ytre-Arne & Moe study. The survey question is very general, it encourages polarized responses (negative/positive), and it does not draw from user experiences with a specific application (e.g., Spotify or another public facing algorithmic system). However, the resulting folk theories are consistent with those found in other studies.

Algorithms believed to be confining "narrow your world view" while algorithms described as practical are "helpful in sorting through cascades of information." Algorithms are reductive because they draw from "stereotypes and simplified notions" and "give crude and limited portrayals of human experience and identity", and they are intangible because they are "difficult to grasp, their power is opaque, and lack of precise insight causes negative feelings of suspicion and foul play." Algorithms are believed to be exploitative because they are "harvesting *your* information to use for *their* purposes."

Agency is a common theme in the folk theories. The user and the system (the algorithm) are viewed variously as places of cooperation, resistance, and acquiescence (e.g., "practical" and "transparent" but also "exploitive" and "unwanted") (Karizat et al., 2021). The locale and influence of that agency can shift. In some cases the user is in full control of the recommendation process, sometimes with disruptive intent ("my goal in life is to confuse the hell out of the Spotify algorithm" (Colbjørnsen, 2018, p. 176)), while at other times the user must fully "capitulate" to the algorithm. As Bucher notes, "While algorithms certainly do things to people, people also do things to algorithms" (Bucher, 2017, p. 42). A study of the folk theories of Facebook's News Feed found two surprising results regarding agency (Eslami et al., 2016). First, 62% of the participants were unaware at the beginning of the study that any algorithm at all was involved in News Feed. The user was in full control. Second, even following interventions that highlighted aspects of the algorithm, 12% of the participants believed the News Feed was completely random. There was no control at all.

2.6.2 Folk Theories and HCXAI

Four papers identify the folk theories of algorithmic systems and make specific connections to HCXAI systems: a study of players of AI games (Villareale & Zhu, 2021), a study of algorithmic news curation (Ngo & Krämer, 2021), a comparative study of proxy tasks and folk theories as alternatives to evaluate user understanding (Gentile et al., 2021), and a study that uses a cognitive framework to determine HCXAI design (D. Wang et al., 2019). While the folk theories in these papers are less fully developed than those reviewed above, they move beyond how folk theories relate to specific explanations and explore how folk theories can influence explanatory systems. A system approach to explanations is a key objective of HCXAI.

2.6.2.1 Villareale & Zhu (2021)

In a preliminary study of how players of AI-based games form folk theories, the researchers conclude that these folk theories can be applied to HCXAI strategies and techniques (Villareale & Zhu, 2021).

In the highly interactive setting of gaming, the researcher identified an "in the moment" folk theory about what the players believe will work in a specific situation. As a result, the development of player folk theories "iteratively cycles through an exploration and elaboration phase" suggesting that understanding this process "can help address the open problem of how users want to interact with XAI and when users may need an explanation" (Villareale & Zhu, 2021).

However, when these "in the moment" theories failed (i.e., did not produce the expected result), players differed in their assignment of the source of the misalignment. Some players believed they were responsible because they held an inaccurate or deficient understanding of how the system works (i.e., their folk theories), a finding consistent with other research into user self-blame for algorithmic inadequacies (Büchi et al., 2021), while other players believed the system operated incorrectly (i.e., their folk theory was sound and there was a problem with the system). Researchers found that users who identified their own responsibility in failure developed a more complex understanding of the AI game while those who focused on the AI as the source of failure had difficulty in formulating more detailed and complex folk theories to overcome the failure.

As a result, developers can "design failure to help users develop more accurate mental models of AI." Researchers recommended that HCXAI "could address this by supporting the redirection of attributions from the AI (i.e., entity) to the player (i.e., person) to effectively strategize with the AI's capabilities. Then, scaffolding explanations to gradually expose the users to other capabilities over time" (Villareale & Zhu, 2021). Similarly, failure in recommender systems, such as inadequate recommendations, present an opportunity for HCXAI highlight the inadequate folk theories that guide user behaviour.

HCXAI should consider "play" as an interactive method to facilitate the cycles of "exploration and elaboration" the researchers identified. This recommendation is consistent with the use of play as a pedagogical strategy in technology environments to assist in problem solving and ideation (Bogost, 2016; McGonigal, 2011). It is also consistent with play as storytelling and the use of narrative exploration in explanatory systems (Norman, 1983b; Siles et al., 2020).

2.6.2.2 Ngo and Krämer (2021)

This study of the folk theories of the Google News algorithmic news curation system identified eight "assumptions" about how the "inner mechanisms" work (Ngo & Krämer, 2021). The focus on mechanisms limited the elicitation of folk theories to operational characteristics of "white-box explanations" and excluded the more holistic algorithmic experience of the participants represented by "black-box explanations:"

> White-box explanations provide information about the input and output of a system and how it determines a certain outcome … Black-box explanations, in turn, provide information on the motivation of a system and influence user satisfaction and comfort of the system (Ngo & Krämer, 2021, p. 3).

Despite these limitations, the findings suggest how folk theories can inform HCXAI approaches. In particular, and uniquely, Ngo & Krämer focus on the implications of the user interface (UI) for HCXAI. To overcome information overload resulting from the complexity of the algorithmic systems, a UI that employs "hidden design" methods is recommended. These methods use unobtrusive visual cues and dropdown menus to signal the availability of explanations but not make them dominant in the UI space.

The researchers also focus on correcting the misconceptions and deficiencies in folk theories by empowering the user, not by trying to turn them into experts. By focusing on a limited number of specific, core misunderstandings (the folk theories identified two: inferences from user data and where user data is stored), explanations can be selective. Like DeVito (DeVito, 2021), Ngo & Krämer believe making users aware of a small number of algorithmic mechanisms and relationships will result a satisfactorily informed user.

2.6.2.3 Gentile et al. (2021)

This study concluded that folk theories rather than proxy tasks (i.e., human simulations of AI tasks) are the most effective way to evaluate the impact of explanations on the non-expert understanding of AI systems (Gentile et al., 2021). Folk theories are the result of real-world decision-making situations and provide an "assessment of human understanding and accuracy in a decision-making task." Folk theories, as they are modified by the explanations, provide a measure of the success of the XAI system.

However, the researchers also argue that users “with different levels of prior knowledge [of AI systems] are likely to have different explanation needs.” As a result, a “quantitative assessment of a users’ prior knowledge of AI systems” needs to be incorporated into an XAI system. While no such assessment tool is presented, this observation is consistent with the HCXAI principles that encourage a pedagogical approach to explanation, particularly the need for “hypotheses about the user’s knowledge and cognitive activity (student model)” (Clancey & Hoffman, 2021, p. 2).

Prior experience with other recommender systems or machine learning systems is a factor in the building and holding of folk theories about subsequent systems (DeVito, 2021; Grimes et al., 2021; Siles et al., 2020; Snead et al., 2015). Pre-existing and persistent folk theories about AI may complicate effective XAI by erecting conceptual barriers or alternatively facilitate XAI by providing useful analogies. In two studies, the accuracy of folk theories did not increase despite specific interventions to improve user understanding of the system (Kulesza et al., 2012; Tullio et al., 2007). The lack of improvement reinforces the difficulty in changing or adjusting folk theories arising from “persistent cognitive biases that influence what information we take in and consider” (Gelman & Legare, 2011, p. 391). To overcome this inertia, users “need additional, high-level feedback in order to adopt more correct structures” (Tullio et al., 2007, p. 32).

2.6.2.4 Wang et al. (2019)

This study describes a framework for “human-centered, decision-theory-driven XAI” that takes into consideration “how people reason, make decisions and seek explanations, and cognitive factors that bias or compromise decision-making” (D. Wang et al., 2019, p. 1). Drawing from the literature of psychology, philosophy, and decision-making theory, the researchers align human reasoning and reasoning heuristics with specific XAI strategies and techniques. While criticizing “unvalidated guidelines” such as those of Mueller et al.

regarding HCXAI (Mueller et al., 2021), Wang et al. acknowledge that these guidelines provide the “high-level objectives” for their “lower-level building blocks.”

Findings from this study support key objectives from the Mueller et al. HCXAI framework. Wang et al. recommend that XAI “support hypothesis generation” where explanations “allow users to generate and test hypotheses” thereby facilitating the HCXAI objective to allow users to “self-explain.” The researchers advise that not just “raw data” (e.g., factors or hyperparameters) but also “supplementary data” or “situational data” be available to users to provide additional context to the explanation and to support self-explanations. The emphasis on hypothesis generation, supplementary information, and “integrating multiple explanations into single explanations” reinforce the HCXAI objective of explanatory systems not solely explanations. The focus in Wang et al. on cognitive science as the basis for XAI is explored and extended in Taylor & Taylor (Taylor & Taylor, 2021). These researchers propose “a cognition-inspired approach to XAI” (Taylor & Taylor, 2021, p. 472). This approach seeks to discover new or enhanced approaches to XAI using the experimental methods of cognitive psychology rather than inferring from existing research, as does Wang et al.

While neither Wang et al. nor Taylor & Taylor discuss folk theories specifically, the elements of decision-making theory and cognitive psychology they explore with respect to XAI are the building blocks of folk theories (Gelman & Legare, 2011). The principles of HCXAI rest upon these elements.

2.7 Summary: Folk Theories, Recommender Systems, and Explainable AI

The intersection of folk theories and XAI can be seen in their respective definitions. Folk theories are the “intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems, which guide reactions to and behavior towards said systems” (DeVito et al., 2017, p. 3165). The focus is on explanation, guidance, and use. With respect to users, the DARPA definition of XAI is for users “to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners” (Turek, 2016). The focus is on understanding, trust, and use. Applied to recommender systems, folk theories provide insights and validation for XAI.

Drawing from the DARPA definition, HCXAI is not narrowly focused on system performance, veracity, or validity (although those are important, indeed essential). It takes a broader view of XAI which encompasses the larger sociotechnical context which includes user expectations and behaviours. The principles of HCXAI describe the requirements of user-centered explanatory systems and act as a benchmark against which folk theories can be assessed.

Researchers have argued that an explainee’s folk theory must be accurate and warn that “the consequences of engaging in the process of explanation and interpretation based on false premises can be catastrophic” (Palacio et al., 2021). While this may be important in highly consequential situations, this assumption challenges what is known about folk theories regarding consumer technologies held by the general population. Folk theories in these settings are often less than fully accurate, can include contradictory information, and yet remain functional (Norman, 1983b). While the alignment of system behaviour with user folk theories is desirable (Norman, 1983b), it has not always resulted in a positive impact on user performance (Schmettow & Sommer, 2016; Xie et al., 2017).

Folk theories can be seen as a bridge that “let us meet the user where they are in terms of understanding and literacy, regardless of how contradictory, sparse, or fragmented these understandings may be” (DeVito, 2021, p. 339:4). Understanding how a system works is a good and perhaps even a necessary thing. However, if users hold relatively intractable ideas about how a system works (i.e., their folk theories), then one way to achieve explainability is to make sure that XAI recognizes those folk theories (Riveiro & Thill, 2021).

3 The Spotify Music Streaming and Recommender System

3.1 Overview of Spotify

In order to capture the nuances and specific experiences of users in their formulation of folk theories, it is recommended that researchers focus on specific algorithmic systems and not generic systems or surveys across multiple systems (Castelo et al., 2019; Hamilton et al., 2014; Lomborg & Kapsch, 2019). As a result, this study selected Spotify as a specific recommender system.

Spotify is a music streaming service that began in 2006 as a P2P file sharing system. In 2014, with the acquisition of the music analytics company The Echo Nest (the.echonest.com), the service moved from human-curated to primarily algorithmically-created recommendations (Carlsson & Leijonhufvud, 2021; Eriksson et al., 2019; Fleischer & Snickars, 2017; Söderström, 2021a; Sun, 2018). Spotify Home (effectively the homepage of the website and app) and features such as Discover Weekly, Radar Release, Made for You, Song Radio, and autoplay all utilize algorithmic techniques to present users with both new and favourite music (Ciocca, 2017; Jebara, 2020; Lalmas, 2019; Popper, 2015). User satisfaction with these recommendations is central to the business model of the company. As noted in the Spotify Initial Public Offering (IPO) Prospectus,

> our ability to predict and select music content that our Users enjoy is critical to the perceived value of our Service among Users and failure to make accurate predictions could materially adversely affect our ability to adequately attract and retain Users, increase Content Hours, and sell advertising to meet investor

> expectations for growth or to operate the business profitably. (Spotify, 2018, p. 31)

Spotify is an ideal service within which to explore folk theories of machine learning because of its size, reach, experience, and relative transparency. The service, available in 184 countries, has ~400M monthly users offering over 82M songs and ~4B playlists (Spotify, 2021a). Approximately 60,000 new songs are uploaded to Spotify each day (Spotify, 2021d). It has been providing algorithmic recommendations for a substantial period of time and the recommender techniques used are generally known (Fleischer & Snickars, 2017). Unlike many companies utilizing recommender systems, Spotify engages publicly through blogs and academic conferences about the nature and challenges of its recommendation processes and infrastructure. Despite this, many of the specific algorithms and their parameterization remain trade secrets. While it is not possible to understand precisely how a recommendation is determined, many of the key concepts, strategies, and processes are known. The recommender techniques are varied, ranging from simple heuristics, to commonly used matrix factorization and collaborative filtering, and lastly to state-of-the-art deep learning neural networks and reinforcement learning incorporating extensive data elements (Chodos, 2019; Eriksson et al., 2019; Stål, 2021; Whitman, 2012). Perhaps most importantly, machine learning is central to both the business processes and the user experience. Tony Jebara, Vice President of Engineering, describes machine learning as “the heart of everything we do at Spotify” (Jebara, 2020).

Additional reasons for focusing on Spotify are the nature of the medium and the experience using it. Listening to music can be a passive experience or a focused and active experience. Music can absorb our attention or be almost subconsciously experienced. Unlike book or other retail recommendations, music recommendations and subsequent listening happen in real time and can be experienced repeatedly (even continuously). Evaluations of music recommendations are immediate and can differ, even for the same song, when the listening context is different. Finally, listening to music is an emotional experience as well as an aesthetic one, it “both shapes and reflects moods” (M. Park et al., 2020, p. 230). As a result, in assessing the nature and source of recommendations, user satisfaction is based on a more holistic, visceral perspective.

Unlike other algorithmic decision-making systems, such as those involving finances, health or employment, Spotify’s recommendations are only moderately consequential in the lives of users. A poor or inadequate music recommendation may be annoying, but its material impact is limited. While this might cause users to take Spotify’s algorithmic processes less seriously, reducing the focus on negative consequences may allow users to be more speculative and less guarded in their observations about the algorithmic recommendations.

Spotify has a substantial impact on music discovery because of its size and reach. This has positioned the service as a cultural intermediary, shaping and influencing music taste not merely reflecting the musical tastes of its users (Karakayali et al., 2018; Morris, 2015; Pelly, 2017; Prey et al., 2020). As a result, users view their interactions with Spotify in terms of an intimate, social relationship with some describing Spotify as a “buddy” in development and expression of their music tastes (Siles et al., 2020). The processes of personalization and recommendation must be viewed from this sociotechnical perspective to contextualize the folk theories of users.

Spotify, like most recommender systems, is comprised of many different algorithms working together. While it is incorrect to refer to Spotify’s algorithm in the singular, for the purposes of convenience and clarity when the singular is used throughout this study it refers to a cluster of algorithms employed in the service. Indeed, when users talk about Spotify, they almost exclusively refer to the “algorithm” not the “algorithms.”

3.2 Spotify Personalization

The layout of Spotify Home differs among the various apps and platforms, but it is generally comprised of Shortcuts, Shelves, and Cards. Shortcuts are quick links to previously accessed features. Shelves are horizontal collections of Cards. A Card is a collection of songs (a playlist).

The most personalized and algorithmically curated cards are the Discover Weekly playlist (recommendations to encourage exploration and new discoveries), the Release Radar playlist (newly released songs based on prior user activity and recommendations for new discoveries), and the Made for You shelf (a set of cards/playlists of familiar songs and new discoveries). Each of these results from Spotify's recommendation infrastructure (Ciocca, 2017; Popper, 2015). The extent of personalization causes Spotify to say "there is no 'one' true Spotify. Essentially, there are 248 million versions of the product, one for every user!" (Jebara, 2020).

3.3 User Information and Preferences

A key issue and potential concern for users is the nature, type, and extent of personal information that a recommender system obtains to make its recommendations. However, most users recognize that useful recommendations require regular feedback to the system: the more relevant and timelier the information, the more effective the recommendations.

Users provide information and feedback to Spotify through a variety of means. During the initial registration, users provide baseline demographic information. As part of its privacy policy (Spotify, 2020b), Spotify identifies a broad set of data elements it collects as a function of system use. These include search queries, date/time of activities, streaming history, user clicks, "likes" and "unlikes", dwell time on songs or playlists, playlists created, a user's Spotify library, browsing history, cookie data, IP addresses, device, network/device performance, browser, operating system, "non-precise location" (e.g., phone GPS data), sensor data (e.g., accelerometer or gyroscope), and third party services with which Spotify and/or the user interacts with (e.g., Facebook and other social media) (McInerney et al., 2018). Spotify also ignores certain information and feedback in specific circumstances. As would be expected, Spotify does not collect data when "private listening" is temporarily engaged. It also does not collect user data regarding children's music (when listened to on an adult account rather than Spotify Kids account), ambient tracks (e.g., the sound of a fireplace), and Christmas music (Douglas, 2018).

While Spotify emphasizes algorithmic recommendations based on these data, the service uses a combination of human and algorithmic curation (Fleischer & Snickars, 2017; Goldschmitt & Seaver, 2019). In a process it calls "algotorial" (algorithmic + editorial), Spotify employs over 100 playlist editors to create "hypothesis" playlists, tested with users in A/B studies, to assist in the training of its machine learning systems (Spotify, 2021d; Stål, 2021). A new drop-down menu appeared on Spotify applications in early 2021 clarifying how recommendations are made:

> Our personalized recommendations are tailored to your unique taste, taking into account a variety of factors, such as what you're what you're listening to and when, the listening habits of people who have similar taste in music and podcasts, and the expertise of our music and podcast specialists. In some case, commercial considerations may influence our recommendations. (Spotify, 2021c)

Editors also curate special purpose playlists, such as the RADAR playlists, which highlight new or largely unknown performers. Artists are able to influence these editors by using a Spotify pitching tool to highlight tracks for inclusion in future playlists (Spotify, 2021d). Spotify also leverages user created playlists to provide additional human curated content as input to algorithmic recommendations (Pichl et al., 2017; Popper, 2015). As a result, recommendations are "composed out of human and algorithmic parts that are constantly reconfigured into arrangements that make it difficult to distinguish between the human and the algorithmic at any level" (Goldschmitt & Seaver, 2019, p. 72).

3.4 Taste Profile

For each user, Spotify creates and updates a Taste Profile (Eriksson et al., 2019; Jehan & DesRoches, 2014; Popper, 2015). These profiles include clusters of preferred artists and genres, activity (clicks, likes, dwell time, playlist construction), and affinity scores for popularity and diversity (e.g., tolerance for exploration). Also included is a "taste-freeze" metric, defined as "the average active year of your favourite artists" (Whitman, 2012).

Spotify uses a variety of high-level descriptions to classify listeners. One, derived from the 2006 Phoenix 2 Project, is core to the Taste Profile (Celma, 2010, p. 46):

> *Savants*: passionate about music with extensive knowledge; judged to be 7% of the population
>
> *Enthusiasts*: keen about music but balanced with other interests; 21%
>
> *Casuals*: music is important but other things are far more important; 32%
>
> *Indifferents*: engage with music but are generally apathetic; at 40% they represent the predominant type of listeners

Another classification characterizes users as "specialists" who listen to similar songs and genres or "generalists" who listen to a more diverse set of songs and genres (A. Anderson et al., 2020). The focus on listening diversity (or lack thereof) is a recurrent theme in critiques of Spotify and in Spotify's own research efforts.

Spotify's extensive and intensive focus on data collection is evidenced in the description by Brian Whitman (a co-founder of The Echo Nest and co-developer of the Taste Profile) of music recommendations in terms of "scale" and "care" (Whitman, 2012). Scale responds to the challenge of the large music dataset (e.g., 80M+ songs) by using algorithmic methods over human curation. Care tempers this computational approach by seeking results which are more sensitive to the individual user: “This is more than activity mining as collaborative filtering sees it: it’s understanding everything about the listeners we can, well beyond just making a prediction of taste based on purchase or streaming activity” (Whitman, 2012). The operationalization of “care” in Spotify’s recommendations involves a complex and data intensive recommendation infrastructure and process that is characterized by the Echo Nest developers as “music retrieval from everything” (Jehan et al., 2010).

3.5 Popularity Bias

The Spotify mission is to “unlock the potential of human creativity by giving a million creative artists the opportunity to live off their art and billions of fans the opportunity to enjoy and be inspired by these creators” (Spotify, 2020a). In the context of its business model, this mission positions Spotify as a multisided market where the service is a platform bringing together different constituents with different objectives: users/listeners, artists, and music companies such as recording labels and music aggregators (Abdollahpouri et al., 2017; Eriksson et al., 2019; Galuszka, 2015; Mehrotra et al., 2018; Mehrotra, Shah, et al., 2020). The nature and requirements of this market have specific implications for Spotify’s algorithmic recommendations.

For example, popularity bias is a key challenge for all recommender systems. The more popular an item is (e.g., a song), the more likely it is to be recommended, resulting in an increase in popularity (Aggarwal, 2016; Jannach et al., 2011; Ricci et al., 2015). Without interventions, recommendations would overwhelmingly favour the popular and the familiar. However, many users like diversity in their listening experience and want recommendations for music that is new to them (Porcaro & Gomez, 2019; Slaney & White, 2006; Villermet et al., 2021). Similarly, emerging artists or those with limited visibility want to reach a new audience that might enjoy their work. Finally, music companies, the largest of which (Universal and Sony) are shareholders in Spotify, want to maximize the value of their music catalogues (Spotify, 2021a). This is the classic algorithmic challenge to exploit or explore (Berger-Tal et al., 2014) couched in terms such as diversity (Lalmas, 2019), novelty (Celma, 2010) or fairness (Mehrotra et al., 2018). Novelty is “the difference between present and past experience, whereas diversity relates to the internal differences within parts of an experience” (Castells et al., 2015, p. 882). Novelty accentuates the “long tail” of the music catalogue exposing users to less popular items over time while diversity increases the daily or weekly inclusion of new and different items within the regularly updated Made for You playlists.

Fairness, which is enabled through diversity and novelty, is for Spotify a business perspective: “blindly optimizing for consumer relevance may have a detrimental impact on supplier fairness” (Mehrotra et al., 2018, p. 2243). Fairness in this context balances the interests of the listener with those of the music companies, who are licensees as well as shareholders, ensuring that a broad presentation of their catalogue is exposed to users (J. Smith et al., 2020). Spotify uses an affinity and sensitivity metric, measuring user tolerance for relevance (i.e., accuracy) against fairness (i.e., diversity or novelty). Relevance and accuracy, in Spotify’s parlance, are measures of known user satisfaction. Relevant and accurate recommendations of music or artists are those that are consistent with the prior listening and satisfaction of users. The user affinity and sensitivity metric is used to optimize user satisfaction at levels that also satisfy music suppliers (e.g., artists and music companies) (Mehrotra et al., 2018; Mehrotra, Shah, et al., 2020). However, “not any random diversity is good diversity, and users who prefer diversity prefer personalized diversity” (Mehrotra, Shah, et al., 2020, p. 695). This receptivity to diversity within a “window” of what a user finds interesting or tolerable is algorithmically predictable (Abdollahpouri et al., 2021; Mehrotra, Shah, et al., 2020). While “personalized diversity” can appear to be an oxymoron, it captures the challenge of multi-stakeholder systems. Findings indicate that creating adaptive policies based on individual’s affinity to fair content results in a minor impact on relevance while positively impacting fairness and satisfaction. Of course, this notion of fairness may not extend equally to independent artists and those represented by smaller music providers (Eriksson et al., 2019; Pelly, 2017).

Fairness in the context of artists rather than music companies is operationalized through the creation of groups:

> We divide the artists into different groups based on their position in the popularity spectrum. Specifically, we consider the popularity distribution of all artists, and bin the artists into ten bins of equal size. In light of group fairness, a set of tracks is fair if it contains tracks from artists that belong to different bins. (Mehrotra et al., 2018, p. 2245)

The fairness of Spotify has been critiqued suggesting that the typical user experience, in part as a result of the influence of music industry ownership, is homogeneous, focused predominately on popular items, and intentionally absent of the long tail of music diversity (Celma, 2010; Chodos, 2019; Eriksson et al., 2019; Fleischer & Snickars, 2017; Snickars, 2017). However, different users (e.g., "specialists" or "generalists") expect recommendations matching their preferences whether diverse or not (A. Anderson et al., 2020).

3.6 Algorithmic Methods

Spotify uses collaborative filtering, content-based, and context-based recommendation methods (Ciocca, 2017; Jebara, 2020; Lalmas, 2019). Collaborative filtering tracks the co-clustering of users and songs by activity measures (e.g., likes, listening, playlisting). While content agnostic, collaborative filtering requires a certain amount of user activity to generate useable metrics. New users or new songs (i.e., those with limited or no activity metrics) constitute the "cold start" problem requiring the use of other recommendation methods.

Collaborative filtering methods typically use either nearest-neighbour, k-means algorithms or latent factor models. Nearest-neighbours and k-means algorithms identify relationships or similarities between items or users to make recommendations (D. Kim et al., 2007; Mittal et al., 2010). Latent factor models create feature vectors for both users and items. Matrix factorization is a widely used implementation of latent factor models (Koren et al., 2009); Spotify is known to have used a particular co-clustering algorithm (Dhillon et al., 2003; Lalmas, 2019). Matrix factorization has higher accuracy than nearest-neighbour, is a more efficient learning algorithm, and can easily integrate additional data elements as features such as audio signals for songs and explicit feedback from users (Koren et al., 2009).

Content-based approaches focus on metadata (e.g., descriptions, tags, web content, social media) and audio features (e.g., tone, timbre). Content-based methods interpret data through regression and categorization. Natural language processing (NLP) techniques, such as summarization algorithms, are applied to textual sources to extract relevant information for subsequent use and to create "cultural vectors" for artists and songs (Whitman, 2012). A critique of the Spotify process for extracting and analyzing data from websites uncovered evidence of how "trivial, random, and unintentional data enters into the data streams" (Eriksson, 2016). Audio features or audio signals are content representing the audio characteristics of the music such as tone, timbre, pitch, and tempo (e.g., low-level indicators). These features are used to compare and contrast songs as well as to infer semantic categories such as genre, danceability, energy, "speechiness", and mood (e.g., high-level indicators) (Dieleman, 2014; van den Oord et al., 2013). Spotify has identified and uses over 5,500 different genres in its music classification (G. McDonald, 2021).

Context-based approaches consider user context and environmental context. Weather, location, specific device used, and social setting are examples of environmental data collected and utilized. User context data, derived from activity sensors and external data, includes intent, time of day, day of week, physical activity level (from the phone's accelerometer), and location (GPS or IP address). Intent is inferred from online activity. Spotify tracks seven general intent categories clustered as active listening (e.g., discover new music, match music to mood) or passive listening (e.g., find background music, access my playlists, or saved music). Spotify jointly uses interaction data (e.g., dwell time, songs played, clicks, even abandoned sessions) and inferred intents to predict user satisfaction with recommendations (Mehrotra et al., 2019). Findings indicate that different interactions vary in importance across intents and suggest grouping user sessions by intent.

Machine learning including reinforcement learning, has become the dominant method for Spotify recommendations. The details of Spotify's current methods were outlined by Oskar Stål, Head of Personalization at Spotify (Stål, 2021).

Spotify creates a massive embedding space from data collected from users, relevant content from a wide variety of sources, and audio characteristics. Shared machine learning models are used for specific use cases (e.g., user affinities, music and artist similarities, and clustering across a variety of subjects). Models created for specific use cases (e.g., Discover Weekly, the Home page playlists) will use data from the shared models and specific input data (e.g., intent, time of day, taste preferences, many others). These models generate recommendations, and user responses and actions are tracked. The algorithms are retrained daily considering user behaviour.

Recently Spotify began using reinforcement learning to understand and manage the longer-term interests of users (and Spotify) (Hansen et al., 2021). Consistent with reinforcement learning terminology, Spotify describes users as having a "state" which reflects their current relationship to music and previous recommendations. Changing user states is the objective. Generally, this means moving users to states of increased satisfaction. In terms of training algorithms, Spotify identifies "ideal" users who appear to be most successful in using the system and maintaining high satisfaction. Data about these exemplar users are used by various models.

Rewards in reinforcement learning shape optimal strategies. The rewards in Spotify are user likes, music saved, playlists created, and other user actions. The objective is to "maximize the future accumulative rewards" hence "optimize for a long-term, fulfilling content diet, rather than a click or stream" (Stål, 2021). By modeling state changes in users Spotify has created models that create the probabilities of users moving from one state to another.

A final tool are the simulators that Spotify runs offline with access to substantially more data than the production models. Simulators predict how users will react to specific recommendations. Algorithms are trained against different simulations and then A/B tested with users in production settings. Algorithms ranked better by users, through the rewards accumulated, stay in production.

3.6.1 Audio Features and Semantic Analysis

Inferring genre or other high-level semantic descriptions from raw audio files (e.g., audio features or signals) is central to music recommendation, especially where user data is sparse. The Echo Nest makes extensive use of audio feature analysis as part of the Taste Profile. As a result, it is a core aspect of the Spotify recommendation process. For example, a “seed” song from a user can initiate a comparison of its audio features with those of other songs resulting in a recommended playlist with songs containing similar features. While the specifics of audio feature analysis at Spotify are not known, the general processes and techniques are widely understood.

The process begins by transforming audio from playback formats (e.g., .au, .wav .mp3) into a useable data format. FFT (fast Fourier transform) and MFCC (mel-frequency cepstrum) are typical algorithms for this with MFCC viewed as a more effective signal representation (Kumar et al., 2016). A study of different classifiers against both FFT and MFCC representations compared logistic regression, k nearest neighbour, support vector machines, decision trees, and LSTM RNN using the GTZAN labeled music dataset for training and testing (Tzanetakis & Cook, 2002). The LSTM RNN outperformed all the other algorithms.

Recently raw waveforms have been used successfully as input to “end to end” machine learning systems (Nam et al., 2019). Data representation and preprocessing algorithms such as MFCC for audio feature identification introduce elements of feature engineering or hand-designed representation requiring or utilizing domain knowledge.

Using raw waveforms in end-to-end neural networks eliminates human domain knowledge bias in preparing the formats for audio feature analysis. This research highlights the need for fully algorithmic and scalable solutions given the high dimensionality of the search space of music recommenders.

As with many aspects of recommendation processes, the extraction and use of audio features to infer semantic information raises issues of algorithmic bias. Chodos asks: “If recommendation systems are just ‘bad’ at recognizing salient features of non-Western music, does that count as a social justice issue?” (Chodos, 2019, p. 85).

3.6.2 Metadata

Metadata about songs, genres, and artists are used to train algorithms and to inform users or other Spotify stakeholders (i.e., artists, music publishers, music licensing organizations, owners). Despite its importance, the metadata practices in the music industry are “complex and broken” (Deahl, 2019) and these problems can materially impact the nature and performance of Spotify’s algorithmic processes. It is estimated that 25% of music publishing royalties are unpaid because of metadata problems (Molinder, 2018). Various individuals and organizations contribute metadata at different stages of the music creation and distribution process each with their own definitions, standards, and quality controls. Since different systems are often used at each stage, data is lost and/or misrepresented when metadata migrates from one system to another. However, repeated calls for metadata standards for music (Deahl, 2019; Molinder, 2018; Slattery, 2019) have ignored the extensive work done by Music Implementation Task Force for inclusion in the RDA Toolkit (www.rdatoolkit.org).

Metadata used by Spotify comes from a variety of sources in addition to those referenced above. Some of it is algorithmically created, such as the audio analysis and genre identification done by the Echo Nest toolset. Spotify also integrates metadata from its editors, directly from artists, and from music data companies. In the latter case, TiVo supplies a variety of metadata including credits for performers, creators, and producers. Somewhat surprising is the lack of folksonomies on Spotify, such as tagging, to incorporate user generated metadata. The naming of user created playlists is the only opportunity for users to apply a form of folksonomies (Besseny, 2020). The importance of metadata, as well as the general lack of quality control, is evident in the extent of metadata hacks perpetrated on Spotify. Metadata reflecting fake artist credits, fake collaborations, and playlist hacks are techniques used by bogus artists to deceive listeners and/or misdirect artist payments (Slattery, 2019).

While not metadata in the strictest sense, the “about” pages for artists are important for many music listeners since this information often serves the function that liner notes did previously with vinyl albums or CDs. These data are also used by the recommender algorithms to augment relationships among artists and to incorporate other contextual information. From the user’s perspective, the provenance of these pages can be problematic. Some pages are clearly authored by the artist themselves. Other pages resemble that for the jazz pianist Claude Bolling which has the bylines of “Rovi” and “arwulf arwulf,” neither of which is identified or is accompanied by a link for more information. Rovi is a music data company Spotify contracted to provide biographies of artist facts. It has since been acquired by TiVo which now provides similar information to Spotify although it is never credited with a byline. Arwulf Grenier is “arwulf arwulf,” a music critic, DJ, and staff member at the Technical Services dept at the University of Michigan. As with conventional metadata, these artist pages are part of the explanatory and descriptive ecosystem of Spotify.

3.6.3 “Bart”: A Key Spotify Algorithm

One of Spotify’s most important algorithms is a multi-armed bandit derivation known as *Bart* (BAndits for Recsplanations as Treatments) (McInerney et al., 2018). Multi-armed bandit algorithms are known for balancing exploitation and exploration (Vermorel &

Mohri, 2005) making them useful in Spotify's desire to balance relevance and accuracy with diversity, novelty and fairness. Exploitation recommends predictions with the highest user engagement (i.e., accuracy or popularity) while exploration recommends less certain but broader predictions (i.e., diversity or novelty). Both processes collect valuable data for future interpretation (e.g., likes or dwell time on those diverse and novel recommendations). As Jebara notes, "almost everything you see [on Spotify Home] is determined by the exploitation and exploration approach" (Jebara, 2020).

The neologism "recplanations" (a mashup of recommendations and explanations) highlights Spotify's analysis that user satisfaction increases with contextual explanations for specific recommendations (McInerney et al., 2018). *Bart* jointly optimizes for recommendations and explanations because "*why* is as important as *what* is being recommended" (Celma, 2010, p. 190). Explanations in Spotify are general and generic, including: "Because it's [a specific day of week]"; "Inspired by [a specific user]'s recent listening"; "Because it's a new release"; "Because [a specific user] likes [a specific genre]"; and "Because it's popular". Explanations can highlight specific moods or foci. These explanations are represented in the Spotify interface as "shelves" (i.e., collections of playlists).

Bart uses reinforcement learning (Sutton & Barto, 2018) with predicted user engagement (i.e., likes, dwell time) from the contextually explained recommendations as the reward model. *Bart* retrains the model at least daily in batch mode (i.e., offline) with performance improvements with the frequency of retraining. Higher user engagement was found for explanations regarding popularity, recent listening, and genre, and less for explanations related to mood and focus (McInerney et al., 2018, p. 36).

3.6.4 Machine Learning Infrastructure: Event Management

Spotify's machine learning infrastructure has evolved from custom coding with onsite servers and storage to approaches based on open source solutions (e.g., Apache Beam) coupled with extensive use of various cloud-based Google services and products such as TensorFlow and KubeFlow (Baer, 2019). A critical aspect of this infrastructure is the event delivery and management process (Maravić, 2016a, 2016b, 2016c).

The event management process takes data ("events") resulting from a wide variety of user actions (e.g., listening, searching, playlist creation, selecting favourites, location, time of day) and channels it into Spotify's data infrastructure for subsequent processing. The flexibility and scalability of the infrastructure was demonstrated with the rapid 2020 implementation of the "Shortcuts" feature (Spotify, 2020c). This project also highlighted Spotify's use of heuristics prior to (and some cases in preference to) developing machine learning models. For all the complexity in much of Spotify's recommendation process, sometimes it relies on simple, transparent heuristic methods. Models are "significantly harder to maintain, monitor, and debug in comparison to heuristics" (Spotify, 2020c). The heuristics for the Shortcuts project, drawn from an analysis of only two listening metrics, served as a performance benchmark for the subsequent machine learning model. The final model, determined after the use of different features, architectures, and hyperparameters, resulted in a 26.7% improvement over the heuristic model.

3.6.5 Assessing the Spotify Algorithmic Methods

Spotify's algorithmic tools and methods reflect a sophisticated infrastructure to gather, process, and utilize data from a wide variety of sources to make music recommendations. Consistent with the approach that machine learning is at the "the heart of everything," Spotify utilizes deep learning and reinforcement learning techniques that address the large scale and scope of its music database and user community.

Key algorithms place particular importance on balancing exploitation and exploration so that the recommendations reflect both known user preferences (exploitation) and diversity through previously unheard music (exploration). While the collection of a broad set of data signals from users and other sources is critical to this balance and to satisfactory recommendations, it is the event management system that manages these data in real time and processes them into data structures that make them algorithmically available. Data preparation involving collection, configuration, verification, and feature extraction accounts for approximately 80% of the effort to build and maintain machine learning models (Sculley et al., 2015).

The use of reinforcement learning techniques, with their focus on developing more comprehensive user models, illustrates Spotify's interest in building longer term relationships with its users. Reinforcement learning in recommender systems, unlike traditional collaborative filtering, creates predictive models that are more responsive to changes in user behaviour.

However, algorithmic methods are susceptible to bias and unfairness (Zehlike et al., 2022a, 2022b). Gender imbalance in Spotify recommendations, by promoting male artists disproportionately, has been attributed to its algorithmic methods (Eriksson et al., 2019; Werner, 2020). Analyzing and categorizing music as raw audio with end-to-end use of machine learning techniques responds to the challenge of managing a massive and growing music database. However, critics have noted that given the characteristics of the training data used in developing these models, there are concerns about the algorithmic analysis of non-Western music that misrepresents its features (Chodos, 2019).

Finally, while many of Spotify's algorithmic methods are known, much is hidden behind intellectual property protections that preserve competitive advantages and deflect potential consumer concerns. The latter is most obvious in speculation that Spotify's algorithmic methods capture and utilize signals about a user's emotional state. While denying its use (Elk, 2021), Spotify has patented algorithms that do just that (Huland, 2021).

3.7 User Experience: The Spotify Algorithm

While a number of studies have explored user satisfaction with Spotify recommendations (Garcia-Gathright et al., 2018; Hosey et al., 2019; Zhang et al., 2013), they have focused mostly on accuracy or alignment with user tastes. One user study revealed observations about the nature of the underlying algorithm (Kuoppa, 2018).

Although the sample was limited (8 participants), users in this study indicated that the Spotify algorithm was “biased to quantitative efficiency” favouring exploitation over exploration. In addition, while users felt that Spotify “provided them new ways to categorize and organize music it also seemed to make listening more incoherent or fragmented as well as decontextualized” (Kuoppa, 2018, p. 100). Listening to recommended playlists makes it easy to lose track of (or even not know) the names of the songs and the artists: “The music is then just a flow of unfamiliar pieces with no context that may have been provided by a friend or a music journalist” (Kuoppa, 2018, p. 100).

In this view the algorithms intervene between the listener and the database of songs to the extent that users increasingly rely on Spotify to “remember” what they liked and to shape their musical tastes. Listeners view Spotify’s algorithms and its recommendations as more than simply offering up suggestions. Despite being largely opaque “black boxes” that operate in ways mostly unknown to them, listeners have made them intimate partners in their lives. The sophistication of the Spotify recommendation engine and the effect it has on listeners makes this service an ideal example to pursue how users understand the processes of algorithmic recommendations.

3.8 Critiques of Spotify

While there are ongoing concerns about low levels of artist compensation (Dredge, 2020; Spotify, 2021b), other critiques focus on issues that have a material impact on the algorithms and the user experience. For example, the decision by Spotify to allow artists to influence their appearance in key playlists (e.g., Discover Weekly or in autoplay) in exchange for lower royalty payments has been critiqued as “streaming payola” (Rogers, 2020).

A study of the Spotify Radio feature (now called Song Radio) questioned Spotify’s recommendation effectiveness (Snickars, 2017). Spotify Radio algorithmically generates a playlist based on a “seed” song identified by the user. Despite interventions by users (e.g., likes, dislikes, skipping), the same songs and artists appeared in a loop. This “more of the same” caused the researcher to suggest “the recommendation ability of Spotify Radio is exaggerated. In fact, one might even argue that the claim of musical personalisation and the ability to be recommended an infinity of content to some extent is even untrue” (Snickars, 2017, p. 207).

A different concern was raised in a study of country music on Spotify. By continually refreshing a country music playlist (to obtain new recommendations), the result “privileges male artists and disadvantages everyone else” (J. E. Watson, 2019). The research noted that “before Spotify recommended one song by a female artist, the algorithm had recommended 121 songs by male artists” (J. E. Watson, 2019). The gendering of music genres (i.e., rock as male-focused, masculine, and white) is reinforced in Spotify recommender systems is “not coincidental, or innocent, but reinforcing patterns of power” (Werner, 2020, p. 88).

3.9 Conclusion

Spotify is described as “a new form of sociotechnical cultural intermediary” (Webster et al., 2016, p. 137). It is a leading music streaming provider with a diverse user community eager to talk about music, Spotify, and recommendations. Spotify is an ideal example of contemporary machine learning recommender systems. It has a large user base, maintains an extensive collection of music and artists, and makes use of a variety of machine learning techniques. Like other commercial recommender systems, Spotify is a “black box” since it is guarded about precisely how recommendations are made.

4 Eliciting the Folk Theories of Spotify Users

This qualitative study examines the folk theories of the users of recommender systems and how they can offer insights to enhance machine learning explainability. The Spotify music recommender system is used as a representative machine learning system. Recruited users of Spotify participated in a survey followed by individual interviews. Folk theories about Spotify are elicited and analyzed using statistical, factor, and thematic techniques. The derived folk theories are discussed in relation to the principles of human-centered explainable AI (HCXAI) to inform and enhance strategies and techniques for explainability in machine learning systems for the lay public. The following research questions inform this study:

> What are the folk theories of users that explain how a recommender system works?
>
> Is there a relationship between the folk theories of users and the principles of HCXAI that would facilitate the development of more transparent and explainable recommender systems?

4.1 Eliciting Folk Theories

Eliciting user folk theories can be difficult. Various verbal and written methods, such a "think-aloud" protocols or questionnaires, can provide incomplete information, contain erroneous information (people say one thing but do another), obscure belief structures, require reasons where users have none, and elicit user responses based on what they think the researchers want to hear (Norman, 1983b). However, self-reporting is still a common research method to explore user folk theories.

Doherty & Doherty categorize reflections in user self-reports as momentary (the "experiencing self"), retrospective (the "remembering self"), and prospective (the "future-oriented self") resulting in different perceptions of the system being examined (Doherty & Doherty, 2018). Questionnaires and interviews about user experiences and conceptualizations rely on retrospective reflection and highlight the concern that "it is not how well users remember their past experience which is of relevance to design but why certain details are reconstructed, and not others" (Doherty & Doherty, 2018, p. 68). The characteristics of retrospective reflection (e.g., past oriented, purposeful recollection, reconstruction rather than recall, lack of a temporal perspective, greater identification of social or emotional aspects, comparative thinking, and possible false causal attribution (Doherty & Doherty, 2018, p. 68)) need to be recognized, mitigated if possible, and acknowledged as limitations in the perspectives obtained.

Possible approaches to mitigate these concerns are the use of analogies and counterfactual questioning. Users often employ "analogies to map the set of transition rules from a known domain (the base) into the new domain (the target), thereby constructing folk theories that can generate inferences in the target domain" (Collins & Gentner, 1987, p. 247-248). In the case of understanding the information behaviour of a recommender system, the known domain of human information behaviour (i.e., a user's own information behaviour) becomes an analogy for making inferences. Counterfactual "what if" and "why not" questions have been used to probe a user's understanding by proposing alternative hypotheses and thereby uncovering beliefs, inferences, and causalities (Payne, 1991).

The folk theory research literature reveals a relatively narrow set of approaches, dominated by reflective interviews or questionnaires (Lin et al., 2012; O.-C. Park & Gittelman, 1995; Staggers & Norcio, 1993). Mixed method studies add direct observation and performance tests to assess the "experiencing self" as opposed to the "remembering self" (Makri et al., 2007; Michell & Dewdney, 1998; O.-C. Park & Gittelman, 1995; Sasse, 1997; Staggers & Norcio, 1993). Some of these studies used experimental or quasi-experimental methods (He et al., 2008; Sasse, 1991). In a review of methods for determining user folk theories, Sasse identified errors as "the most informative data about users' models" but noted that most studies merely "count errors rather than analysing them" (Sasse, 1991, p. 66). Overall the review found an "over-reliance on performance data and lack of ecological validity" (Sasse, 1991, p. 69).

The difficulty in determining users' folk theories is exemplified by two studies. A user study of how Twitter works was notable not for the accuracy or inaccuracy of the answers but the high instance of "unsure" responses to almost all the questions (Proferes, 2017). Most users didn't know how Twitter works and weren't able to or willing to describe a folk theory about it. Similarly, in a user study of the SPSS statistical software, few could articulate their own folk theories. The typical response was "I don't have one" (Staggers & Norcio, 1993, p. 597).

The difficulties of eliciting folk theories have not diminished their value although researchers are cautioned "to move away from the perception that truth is being sought" and focus rather on the utility of the research to further understand behaviour (Rouse & Morris, 1986, p. 360).

4.2 Mitigating the Limitations of Folk Theory Elicitation

With respect to eliciting folk theories, Hoffman et al. note:

> People may not be able to tell you "everything" about their understanding, and they may not be able to tell it well. But with adequate scaffolding by some method of guided task reflection, people can tell you how they understand an event or system, they can describe their knowledge of it, and the concepts and principles that are involved" (Hoffman et al., 2019, p. 9).

While this study uses the common elicitation practice of self-reporting, it attempted to mitigate some of the concerns of this method. Spotify users completed a survey followed later by an interview. Using two different elicitation methods, separated by several days or weeks, enabled participants to express their views differently and perhaps more clearly. It also allowed participants to reflect on the answers from the survey prior to and during the interview. This encouraged a greater degree of self-reflection. The separation of survey and the interview allowed the researcher to consider the survey responses in guiding the direction of the interview based on the known views of the user. The semi-structured interviews with users are designed using a form of laddering (Price, 2002; Reynolds & Gutman, 1988) that draws the participant into more detailed descriptions of their views. Factual questions were coupled with counterfactual questions, and these were grouped thematically to ensure key aspects of Spotify recommendations were considered.

The trustworthiness of research findings is always a concern. Member checks are "the most crucial technique for establishing credibility" (Lincoln & Guba, 1985, p. 314). Interview participants were asked to review and validate their interviews and to address any preliminary interpretations arising from them. Allowing participants to contest the emerging findings in a variety of fora (interview transcripts, post-interview clarifications) provided for "negotiated outcomes" where participants were able to "negotiate meanings and interpretations." Member checks are important because a "working hypotheses that might apply in a given context are best verified and confirmed by the people who inhabit that context" (Lincoln & Guba, 1985, p. 41).

5 Methodology

This study used a multistage process. Users of Spotify were surveyed and subsequently individually interviewed. The elicited folk theories from the survey and interviews were analyzed against the principles of HCXAI.

The intent of the user survey and interviews was to obtain the everyday experiences of a convenience sample of Spotify users. The everyday experiences of users reflects the "messiness of real life" (Braun & Clarke, 2006) complete with responses that are "complex, nuanced, playful, glib" (Kant, 2020). The survey and interviews were open to all Spotify users (both paid and free, ad-supported). While the paid version of the service offers more complex algorithmic processing for its recommendations, focusing solely on this group of users could limit participation on economic grounds and impact diversity.

In total 19 Spotify users were surveyed and then individually interviewed to allow for a more in-depth and interactive exploration of the Spotify recommendation algorithm. Participants were recruited through Twitter and received a $25 e-gift card. The survey, conducted using the Qualtrics survey tool, collected some baseline information about the usage of Spotify as well as initial responses to three key questions that were explored in more detail during the interviews. The individual interviews were conducted and recorded using Zoom. Interviews lasted approximately 60 minutes. The survey was analyzed using SPSS and the interviews were analyzed using NVivo.

5.1 Spotify User Survey

Following the guidance that a survey be "appropriately brief and simple to complete" (Hank et al., 2009, p. 257), the user survey consists of 7 closed, contextual questions, two open-ended questions, and a final section of 22 scalar statements. One open-ended question asks participants to describe how they think Spotify uses information to make personalized recommendations. The other open-ended question asks participants about what strategies they might use to influence (i.e., change) those recommendations. The scalar statements ask participants about the influence 22 data elements have on their Spotify recommendations. The order of the statements was randomized for each participant. Responses to the survey were analyzed using SPSS prior to the individual interviews and informed the direction and focus of the user interviews. The survey was pre-tested with three Spotify users. Some questions were revised for clarity as a result. Further information about the Spotify User Survey is detailed in Chapter 6.

5.2 Spotify User Interviews

As “purposeful conversations,” semi-structured interviews are recommended when researchers “are aware that individuals understand the world in varying ways. They want to elicit information on their research topics from each subject’s perspective” (Luo & Wildemuth, 2009, p. 233). This process allows the interview to follow the specific context and experience of the participant.

Research is divided on whether an interviewer needs to be an expert in the field involved in order to successfully elicit the ideas and concepts of those being interviewed (in this case machine learning systems) (Hove & Anda, 2005; Kvale, 1996). However, the practices of ethnographic researchers suggest methods for informed but non-domain experts to acknowledge their limitations and still conduct in-depth and probing interviews (Forsythe, 2001; Spradley, 1979). Combining “friendly conversations” with more specific, probing questioning allows for a more open rapport and an encouragement for more diffusive responses (Spradley, 1979). While it is important to seek breadth and detail in these interviews, “grounded theorists attend more to whether their participants’ accounts are *theoretically plausible* than whether they have constructed them with unassailable accuracy” (Charmaz & Belgrave, 2012, p. 352).

Content analysis is widely used in LIS (Armann-Keown & Patterson, 2020). Transcripts from the user interviews were analyzed using the NVivo qualitative data analysis software (version R1). Further information about the Spotify User Interviews is detailed in Chapter 7.

5.3 Recruitment

The recruitment criteria required that participants be Spotify users, 18 years or older, and a resident of Canada or the United States. This information was confirmed during the initial stages of recruitment. Professional colleagues from previous workplaces were excluded as were friends and relations.

A recruitment tweet was posted to the researcher's 2,958 followers on February 16, 2021. Tweepsmaps, a Twitter analysis service, reported that 68% of the followers were from Canada and 22% from the United States. Twitter reported that the tweet had received 3,639 impressions and 119 engagements. Other Twitter users had "liked" (5) and/or retweeted (10) the original tweet. The original tweet was retweeted on February 19, 2021 indicating that participants were still being recruited. Twitter reported that the retweet had received 928 impressions and 23 engagements. Other Twitter users had "liked" (2) and/or retweeted (2) the retweet.

From this activity, 11 Twitter Direct Messages (DMs) were received from potential participants. Interested participants were contacted (DM or email) with a link to the Letter of Information (LOI). At this stage, 10 were confirmed into the study. One DM was from a country outside the research protocol. Surveys and subsequent interviews were begun with this confirmed group.

Four participants who had completed their participation (i.e., both the survey and interview) agreed to inform friends or colleagues about the project. They tweeted, posted on their Facebook accounts, or emailed prospective participants. Having reviewed the Letter of Information, an additional 10 participants were confirmed.

By Feb 27, 2021, 20 participants were confirmed and the recruitment for Spotify users was terminated. After repeated efforts to engage User 1 in completing the survey, this user was removed from participation resulting in a final participant total of 19. See Appendix 2 for the recruitment tweets and Appendix 3 for the Letter of Information and Consent.

5.4 Limitations

Given the use of Twitter as the primary recruitment tool, this is a self-selected sample and therefore not necessarily representative of the population of Spotify users (Connaway & Powell, 2010). While the sample sizes is small (19 Spotify users), the purpose of qualitative research is “not to generalize from a sample but to develop an in-depth understanding of a few people” (Creswell & Plano Clark, 2011, p. 174). The survey and interviews revealed low variability in views and beliefs precluding the need for additional participants (Connaway & Powell, 2010). However, it is arguable that a larger sample would have uncovered additional perspectives.

The user interviews were conducted under COVID-19 pandemic restrictions. Instead of in-person interviews, Zoom was used. Remote interviews can obscure the nuances of in-person engagements and may have negatively affected the elicitation of folk theories and the exploration of XAI strategies. However, Zoom made it easier to conduct interviews over geographic distances.

This study uses Spotify as a representative recommender system and an example of machine learning. While different recommender systems and machine learning systems share some common features, the user experiences with Spotify, and the resultant folk theories, cannot be generalized to other recommender or machine learning environments.

Spotify, as with most recommender and machine learning systems, is rapidly evolving. Following the user survey and interviews in early 2021, Spotify has added new features and algorithmic processes. The observations of the users may no longer reflect the specific conditions of the current Spotify system.

5.5 Ethical Considerations

This study posed a minimal risk to participants where "the probability and magnitude of possible harms implied by participation in the research are no greater than those encountered by participants in those aspects of their everyday life that relate to the research" (Canadian Institutes of Health Research et al., 2018, p. 22). The names of users surveyed and interviewed were replaced with study IDs to ensure confidentiality. Users were given the option of reviewing the transcript of the interviews to ensure no sensitive information was revealed.

The survey and interviews with Spotify users received approval from the Western University Research Ethics Board on February 5, 2021 (see Appendix 4).

6 Spotify User Survey

6.1 Introduction

Following the guidance that a survey be "appropriately brief and simple to complete" (Hank et al., 2009, p. 257), the Qualtrics user survey consists of 8 closed, contextual questions, 2 open-ended questions, and a final section of 22 scalar statements. Two of the contextual questions ask participants about the nature of their music interests and whether they listen to a wide variety of music or only specific artists and genres (i.e., Generalists vs. Specialists). Five questions inquire about their use of Spotify: whether they use the paid or free version, how long they have been using Spotify, how often they listen to it, what kind of device they use to access the service, and if they are satisfied with the recommendations. The final contextual question asks about the nature of the recommendation system: "How do you think Spotify's personalized music recommendations are made?"

- Solely by algorithms
- Primarily by algorithms and partly by humans
- Primarily by humans and partly by algorithms
- Solely by humans
- Don't know

The two open-ended questions ask participants to describe how they think Spotify works and how they could shape the recommendations:

> "How does Spotify use information to determine the personalized music recommendations for you?"

> "What could you do to shape the personalized music recommendations you receive from Spotify?"

The scalar statements ask participants about the influence 22 data elements have on their Spotify recommendations ("To what extent do you think the following influence Spotify's music recommendations for you?"). Respondents ranked these data elements as "very important," "important," "somewhat important," or "not important." The order of the statements was randomized for each participant. Responses to the survey were analyzed prior to the individual interviews and informed the direction and focus of the interviews. See Appendix 5 for the Spotify User Survey.

6.2 Methods

Spotify users completed the survey using the Qualtrics system. The survey data was downloaded to an Excel file and extraneous variables were removed (e.g., personally identifying information, geolocation, email, etc.). The edited Excel file was uploaded to SPSS (version 27.010) for analysis. Descriptive statistics were obtained and analyzed. Cross tabulations were run on a series of variables.

The 22 statements rate data signals according to how much users believe they influenced Spotify recommendations were tabulated according to the scale provided ("very important", "important", "somewhat important", and "not important"). A frequency table was created, and the results analyzed.[1]

---

[1] User 2 noted in the interview that they had misunderstood a survey question (about factors that might influence a recommendation); they resubmitted their responses and the Qualtrics entry was edited to reflect the new answers. They had understood the question to mean what factors where important to her, not what is important to Spotify. Another user (User 9), during the interview, asked to change a survey response; this was done based on their instructions. The factor question regarding "Playlists I've created" was altered from "Very Important" to "Somewhat Important." The analysis of the survey responses was re-run to include these adjustments.

The open-ended questions “How does Spotify use information to determine the personalized music recommendations for you” and “What could you do to shape the personalized music recommendations you receive from Spotify?” were analyzed for key concepts. The responses to both questions were also uploaded to NVivo to be further analyzed thematically in conjunction with the user interviews.

Finally, a factor analysis was conducted on the 22 statements ranking the influence of these data signals using Q methodology. This analysis is described in more detail below.

6.3 Results

6.3.1 Participants

The Spotify User Survey has 19 participants. All are 18 years or older, live in Canada or the United States and are users of Spotify. The participants are experienced with Spotify with 95% having used the service for at least one year and 90% listening to it most days or every day. Most pay for the service: Paid (79%); Free (ad supported) (21%). See Tables 2 and 3 for details.

Table 2: Experience with Spotify

	Frequency	Percent
Less than 1 year	1	5.3
1 to 5 years	10	52.6
More than 5 years	8	42.1
Total	19	100

Table 3: Frequency of Spotify Use		
	Frequency	Percent
Less often than weekly	1	5.3
At least weekly	1	5.3
Most days	8	42.1
Every day	9	47.4
Total	19	100

The participants differ on how they access Spotify. None of the participants selected the option “On a smart assistant (e.g., Alexa, Google Home).” See Table 4 for details.

Table 4: Device Used to Access Spotify		
	Frequency	Percent
Computer	9	47.4
Smartphone	9	47.4
Other	1	5.3
Total	19	100

The musical interests and objectives of the participants differ: Generalists (“listens to a wide variety of artists and genres”, 58%) outnumber Specialists (“listens to mostly the same artists and genres”, 42%). Most participants describe their interest in music as either Keen or Passionate (90%). See Table 5 for details.

Table 5: Interest in Music		
	Frequency	Percent
Engage	1	5.3
Important	1	5.3
Passionate	8	42.1
Keen	9	47.4
Total	19	100

While equal numbers of respondents accessed Spotify on a computer (desktop/laptop) or a smartphone (47% in each case), Generalists preferred a computer (8 of 11), and Specialists preferred a smartphone (7 of 8). See Table 6 for details.

Table 6: Category * Device						
			Device			Total
			Computer	Smartphone	Other	
Category	Generalist	Count	8	2	1	11
		Percentage	42.1%	10.5%	5.3%	57.9%
	Specialist	Count	1	7	0	8
		Percentage	5.3%	36.8%	0.0%	42.1%
Total		Count	9	9	1	19
		Percentage	47.4%	47.4%	5.3%	100%

6.3.2 Satisfaction

Despite the differences described above, 79% of the participants indicated satisfaction with the recommendations that Spotify provides. However, differences in satisfaction were evident between Generalists and Specialists.

6.3.2.1 Generalists vs. Specialists

Prior research has demonstrated differences between Generalists and Specialists regarding their acceptance of diverse recommendations, and their behaviour and expectations with respect to the nature and performance of a recommender system (A. Anderson et al., 2020). Generalists, with diverse music interests, "who become more generalist over time tend to do so by drifting away from algorithmically-driven listening and gravitating towards user-driven listening (i.e., search). These results strongly suggest that algorithmic recommendations are associated with *reduced* consumption diversity" (A. Anderson et al., 2020, p. 2156). As a result, with their more narrow and specific musical interests, "classical recommendation models perform much better for specialists than for generalists" (A. Anderson et al., 2020, p. 2164).

While not statistically significant and generalizable, data from the survey indicates differences specific to those described as Generalists or Specialists. These categories of listeners have different musical interests and objectives in using services like Spotify. Generalists listen to a wide variety of artists and genres while Specialists tend to listen to the same artists and genres. As a result, these groups may have different expectations of Spotify's recommendations and have different perceptions of how the recommendation system works.

Only Generalists indicated that they were not satisfied with Spotify's recommendation (4 of the 11 Generalists: 21% of the participants). See Table 7 for details.

Table 7: Category * Satisfied Crosstabulation

Fisher's exact test (p = .105). Not statistically significant.

			Satisfied		Total
			No	Yes	
Category	Generalist	Count	4	7	11
		Percentage	21.1%	36.8%	57.9%
	Specialist	Count	0	8	8
		Percentage	0.0%	42.1%	42.1%
Total		Count	4	15	19
		Percentage	21.1%	78.9%	100%

6.3.3 How Recommendations are Made

With respect to how recommendations are made, none of the participants selected the options "Primarily by humans and partly by algorithms" or "Solely by humans." The principal role of algorithms over humans in making recommendations is indicated by 95% of respondents. See Table 8 for details.

Table 8: How Spotify Recommendations are Made		
	Frequency	Percent
Don't Know	1	5.3
Primarily by algorithms and partly by humans	7	36.8
Solely by algorithms	11	57.9
Total	19	100

While most respondents believe recommendations are made "solely by algorithms" (58%), there is no appreciable difference between the beliefs of Generalists and Specialists. The "Don't Know" response was removed. See Table 9 for details.

Table 9: Category * How Made Crosstabulation

Fisher's exact test (p = .786). Not statistically significant.

			How Made		Total
			Primarily by algorithms and partly by humans	Solely by algorithms	
Category	Generalist	Count	4	7	11
		% within Category	36.4%	63.6%	100%
	Specialist	Count	3	4	7
		% within Category	32.9%	57.1.%	100%
Total		Count	7	11	18
		% of Total	36.8%	57.9%	100%

Of the dissatisfied Generalists, 75% (3 of 4) believe that recommendations are made "solely by algorithms." No Specialists were dissatisfied with Spotify's recommendations. See Table 10 for details.

Table 10: Category * How Made & Satisfaction

Fisher's exact test (p = .453). Not statistically significant.

			Primarily by Algorithms & Not Satisfied	Solely by Algorithms & Not Satisfied	Total
Category	Generalist	Count	1	3	4
		% within Category	25%	75%	100%
	Specialists	Count	0	0	0
		% within Category	0%	0%	100%

6.3.3.1 User Beliefs about Spotify's Recommendations

Three sections of the survey inquired about participant beliefs regarding how Spotify's recommendations worked:

1. Participants were asked to rate each of 22 statements about data signals according to how much they thought each contributed to Spotify recommendations (see Figure 2 for a ranked list). All the data signals described in the statements are known to be, or thought to be, collected, and used by Spotify (Eriksson et al., 2019; Jehan & DesRoches, 2014; McInerney et al., 2018; Spotify, 2020b).

2. The open-ended question "How does Spotify use information to determine the personalized music recommendations for you?" probed both how

recommendations are made and what information is used in making those recommendations.

3. The open-ended question “What could you do to shape the personalized music recommendations you receive from Spotify?” probed how participants could take some control over the recommendations by intentionally shaping them.

Participants often prefaced their responses to the open-ended questions with phrases that signaled their lack of knowledge or uncertainty about recommender systems: “not sure exactly” (User 2), “I assume” (User 5), “best guess” (User 11) and “this is total guess work as its completely opaque to me how it all happens” (User 3). User 4 was concise and definitive: “I actually don’t know.”

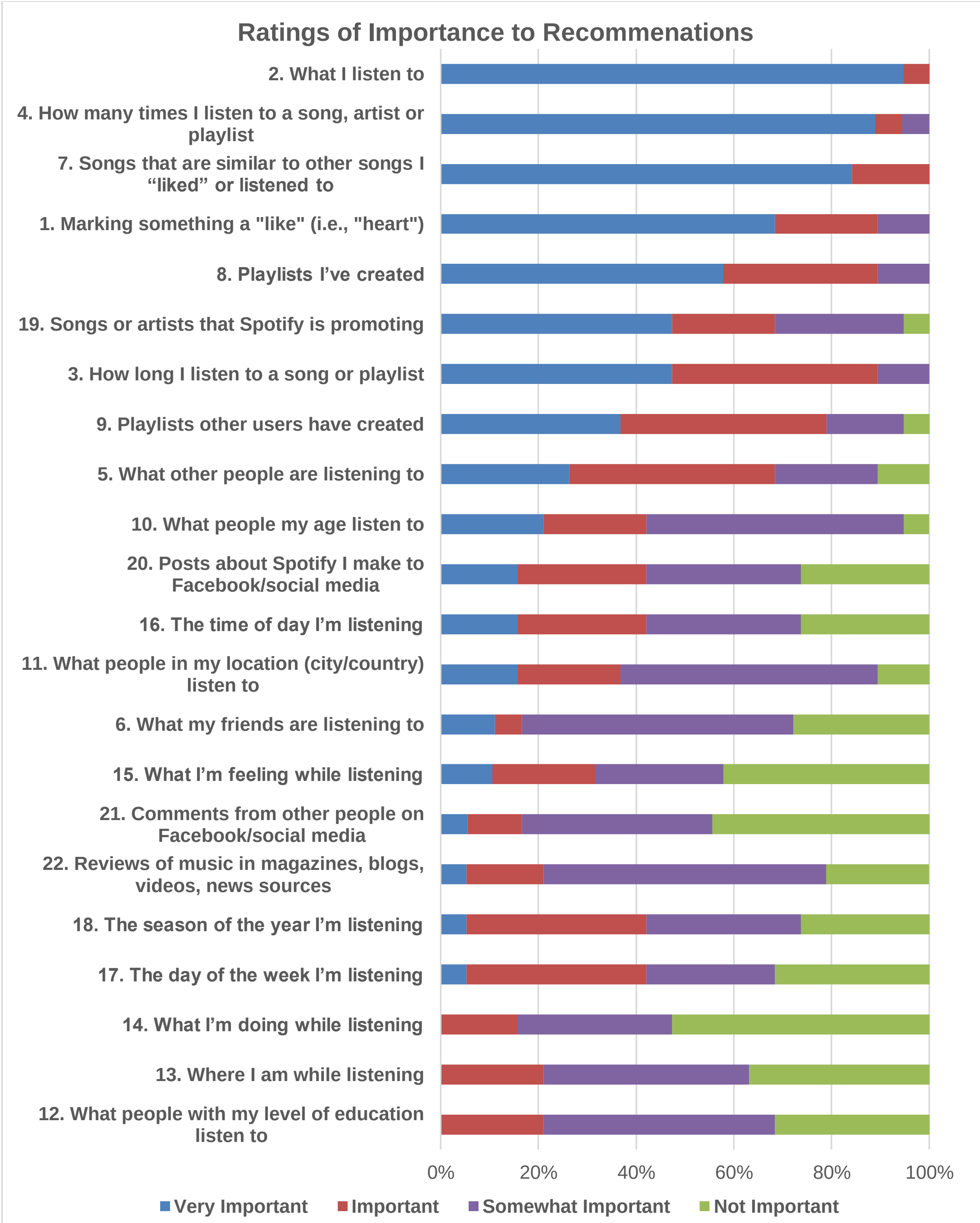


Figure 2: Ratings of Importance to Recommendations

6.3.3.2 Common Responses

The most common responses to how Spotify uses information identified a similar cluster of actions: what users were listening to (songs, artists, and genres), frequency of listening, skipping songs, “liking” (i.e., “hearting”) songs or playlists, creating playlists, and adding songs to their library. These were among the highest-rated statements: For example, the following were all rated “very important”: “What I listen to” (95%), “How many times I listen” (89%), and “Marking something a ‘like’ (i.e., ‘heart’) 68%. One user acknowledged the “cold start” problem (i.e., insufficient data on new users or newly added songs): “once there’s enough data about the type of genres or tags that you listen to, they’re able to suggest personalize music recommendations” (User 13).

While most users were uncertain about the recommendation process, User 18 identified a causal relationship between their actions and the results: **“**I can usually determine which parts of my listening history have influenced a recommendation.” User 19 noted the rationale for the information obtained and used by Spotify: “I’m providing Spotify with feedback as to my likes and preferences.”

What other people do on Spotify also provides important information used in personalizing a user’s recommendations. “What other people are listening to” and “Playlists other users have created” were both ranked “important” by 42% of the participants. However, “What my friends are listening to” was ranked only “somewhat important by a majority of participants (53%). Spotify “find[s] people that have similar habits to you, and then look[s] for recommendations based on what that group listens to” (User 11) and “fills in the gaps” (User 6). The playlists others create were mentioned by Users 5 and 16. User 11 believes that “other people’s listening habits are really what drives it.” Another example of the importance of “other people” is the high rating (47% “very important”) of “Songs or artists that Spotify is promoting.” Spotify promotions are made by music editors (i.e., “others”).

6.3.3.3 Shaping Recommendations

In response to how users might shape recommendations, users simply repeated the actions listed above: listen more, even “repeatedly” (User 13), to the same artist or genre, “like” more songs or artists, and create more playlists. New strategies included removal and avoidance: “remove the outliers” (User 12), “remove a song from a playlist” (User 15), and “skip a song within a certain amount of time” (User 15). In a unique suggestion, User 14 wants to signal their emotion state: “I think it could be cool to give feeling inputs to the algo and have it deliver a mix based on those things.”

User 20 believes different data signals result in different results: “‘Liking’ songs and bands sometimes will influence recommendations, but it seems like listening to bands sends a stronger message to Spotify.” User 6 concurs: “the more you listen to the same type of music, the more specialized the music becomes.” Using the “don’t play this” option (e.g., disliking something with the “Don’t like this song” or “Don’t like this artist” indicator) was noted by User 3 as a strategy to shape or alter recommendations, although User 20 cautioned that it “doesn’t seem to influence algorithms too much.” Some users expressed frustration in attempting to shape Spotify’s recommendations. Instead, they wanted to “game their algorithm” (User 10) or take actions aimed at “confounding any attempt to box my taste in” (User 4). Another user would open a private session to hide their normal preferences (User 15).

6.3.3.4 Explicit and Inferred Data Signals

With respect to important data signals, only one user mentioned the age and location (i.e., city/country) of the listener (User 11) and another referenced “demographics” generally (User 5). The statements regarding age (“What people my age listen to “) and location

(“What people in my location (city/country) listen to”) were both ranked only “somewhat important” by 53% of the participants. Another demographically related data element, “What people with my level of education listen to” also received low ratings: “somewhat important” (47%) and “not important” (32%).

Statements about data signals inferred about a participant received low ratings: “What I’m feeling while listening” (26% “somewhat important” and 42% “not important”), “Where I am while listening” (42% “somewhat important” and 37% “not important”), and “What I’m doing while listening” (53% “not important”). The last two statements were the lowest rated among all the statements. Similarly, statements about temporal data signals were lowly rated: “The time of day I’m listening” (31% “somewhat important” and 26% “not important”), “The day of the week I’m listening (26% “somewhat important” and 32% “not important”), and “The season of the year I’m listening (31% “somewhat important” and 26% “not important”).

6.3.3.5 Classification and Similarity

Participants believe Spotify categorizes and classifies music and artists in some manner although the various methods used were described in general terms. Spotify uses “tags ... or some adjectives” (User 13), it can “associate tags (some kind of metadata?)” (User 9), and it applies “key descriptors or tags ... and then recommendations are made that match those descriptors” (User 18). Part of this descriptive method includes relationships among and between songs and artists. This was described as “adjacency” (User 10), “associations” (User 7), and “related” (User 8). The “sound of a song” (User 14) was the sole mention of the characteristics of music (i.e., audio ID). However, a related statement, “Songs that are similar to other songs ‘liked’ or listened to”, is one of the highest rated statements with 84% of participants identifying it as “very important”.

Only two participants identified the weighting of information as influential. User 14 noted the “emphasis on recent listening” in recommendations. User 9 identified weighting influenced by popularity bias or used for financial advantage (“weighting towards recordings deemed to be popular and/or profitable to Spotify”).

The word “similar” appears frequently in participant responses. Similarity is central to many of beliefs about how information is used in recommendations: “similar habits” (User 11), “similar artists and genres” (User 5), “similar genres and eras” (User 19), “similar tastes” (User 19), “artists that are similar” (User 6), “similar tracks” (User 8), and songs or artists that “most closely resemble” each other (User 12). “Songs that are similar to other songs ‘liked’ or listened to”, is one of the highest rated statements with 84% of participants identifying it as “very important”. User 18 identified a result where notions of similarity were violated: “Recently Spotify ‘botched’ a recommendation by picking up on the fact that I had listened to Bach cello suites and Chopin piano pieces and gave me a strange cello arrangement of a Chopin piano piece. Those things are not equal.”

6.3.3.6 External Data

Most comments referenced data that was generated through the Spotify application. Only two users believe that Spotify uses information from external sources. User 10 mentioned external media sources (“market data” and “music sales”) and User 5 mentioned social media: “I’ve connected my Spotify account to Facebook, so there’s a load of demographic and interest-related information that could be used in a personalized recommender system.” Reflecting a similar lack of importance, “Reviews of music in magazines, blogs, videos, news sources” was rated “somewhat important’ (58%) and “not important” (21%) while “Comments from other people about music on Facebook or other social media” was rated “not important” (48%).

6.3.3.7 Data Signal Importance

The ratings of the importance of the 22 statements of data signals were compared for the following groupings using t-tests:

1. Music interests and objectives in using Spotify (Generalists vs. Specialists),

2. Device used to access Spotify (computer vs. smartphone),

3. How recommendations are made (solely by algorithms vs. primarily by algorithms and partly by humans), and

4. Whether respondents were satisfied with Spotify recommendations.

These characteristics are thought to result in differences regarding how respondents believe certain data signals influence recommendations. However, no statistically significant results were obtained[2].

6.3.3.8 Algorithms and Humans

In responding to how Spotify uses information, many participants discussed what, or whom, is using that information: algorithms, humans, or both. User 14 imagines a "vast recommendation algorithm" at the core of Spotify while User 16 sees a balance of "both algorithms and human curation." The human influence is important "especially in naming the [genre] categories and validating algorithmically suggested relationships between

---

[2] Multiple testing such as that conducted on the statements is known to introduce Type 1 errors. Before the Shapiro-Wilk tests were run, the Bonferroni correction was applied (Armstrong, 2014) resulting in a new alpha for significance of $p<0.0023$ (determined by dividing 0.05 by the 4 sets of 22 tests).

artists/genres" (User 3). User 5 makes distinctions about the roles of algorithms and humans:

> I answered, "solely by algorithms," because I don't think that there are people personally picking out songs for just me, but I do think there are humans involved in the overall process. Maybe there are people employed by Spotify in some capacity to curate genre playlists that the algorithms draw from to create more personalized recommendations, for example. Or maybe Spotify algorithms take user-created playlist data from their platform (and probably others) for machine learning.

Most participants believe Spotify's recommendations are made "solely by algorithms" (58%) while 37% believe they are made "primarily by algorithms and partly by humans."

User 18 summarized the general understanding of most participants about how Spotify uses information for recommendations and how those recommendations can be shaped: change, adjust or remove "anything that tells the algorithm to privilege one part of my data more than another." However, the observation of User 3 echoes the frustrations of many participants: "training the algorithm is a lot of effort."

6.3.3.9 Factor Analysis

Deriving folk theories requires an understanding of the diverse and often conflicting conceptualizations of recommender systems expressed by users. The need is not to identify clusters of like-minded users but rather to surface clusters of similar concepts or beliefs. These clusters or factors identify different subjective perspectives which serve as building blocks for folk theories.

Q methodology, developed by William Stephenson (Stephenson, 1954), is a factor analysis method that "inverts the R methodological tradition by employing persons as its variables and tests, traits or other items as its sample or population (of cases)" (Watts &

Stenner, 2012, p. 22). Q methodology "reveals the key viewpoints extant among a group of participants and allows those viewpoints to be understood holistically and to a high level of qualitative detail" (Watts & Stenner, 2012, p. 4). With respect to how users believe the Spotify recommender system works, this method will "identify the different subjective perspectives expressed across the group of participants" (Bailey et al., 2019, p. 7). This methodology has been used with a variety of LIS research topics (VanScoy, 2021).

While most often Q methodology requires participants to engage in a forced selection process from a series of statements, a non-forced process has been defended and subsequently widely used (Bolland, 1985; Brown, 1971). This study uses a non-forced approach.

The specific Q sorts used in this study are statements with participant ratings responding to the following question from the Spotify user survey: "To what extent do you think the following influence Spotify's music recommendations for you?". The following rating scale is used: Very Important (4), Important (3), Somewhat Important (2), and Not Important (1)

The Q sorts were analyzed with the KADE software package (Banasick, 2019). The default Q methodology removes cases with missing variables from analysis. The Q sorts for Users 4, 7 and 15 each have one missing variable. To include these in the analysis, the missing variables were imputed from the average rating of the other participants[3].

---

[3] User 4: "How many times I listen to a song, artist or playlist" (participant average = 3.8) was imputed to the rank 4 ("very important").

User 7: "What my friends are listening to" (participant average = 2.1) was imputed to the rank 2 ("somewhat important").

User 15: "Comments from other people about music on Facebook or other social media" (participant average = 1.7) was imputed to rank 2 ("somewhat important").

The Brown centroid factor extraction method was used with the number of factors to extract set at seven (Brown, 1980). The varimax factor rotation method was used (Akhtar-Danesh, 2017) and four factors (Factors 1, 2, 4, and 6) were identified. Eigenvalues measure statistical strength and variance. Watts & Stenner recommend that factors with eigenvalues bordering on the cut-off of 1.00 be retained (Watts & Stenner, 2012, p. 105-106). As a result, Factor 4 (0.9) and Factor 6 (0.9), are included in this study. These four selected factors together explain 71% of the study variance. Users 10 and 18 were correlated to Factor 7 with a low eigenvalue of 0.48 and were removed from further analysis. The eigenvalues of Factor 3 (0.14) and Factor 5 (0.14) are considered too low for the inclusion of these factors. See Appendix 6 for factor loadings and user clusters.

Relative ratings of statements identify for each factor both “consensus” statements and “distinguishing” statements. Consensus statements are similarly ranked between factors (i.e., across participants), while distinguishing statements for a specific factor are those statements whose rank is significantly different for that factor as compared to the other extracted factors (Rahma et al., 2020). Distinguishing statements are identified at the $p<0.01$ and $p<0.05$ level of significance. Watts & Stenner caution against focusing solely on the distinguishing statements arguing for the holistic perspective insisted on by Stephenson that viewed all statements in the context of the specific factor (Stephenson, 1954). As a result, “it is the interrelationship of the many items *within* the Factor 1 array [the factor example provided by the authors] that should ultimately drive our interpretation” (Watts & Stenner, 2012, p. 149). Watts & Stenner emphasize “the logic of abduction” encouraging researchers to “never pass blandly across an item [i.e., a statement and its rating] without considering its implications ... ‘What does it mean?’ ‘What is it trying to tell me?’” (Watts & Stenner, 2012, p. 155).

6.3.3.9.1 Consensus Statements

While the four factors are separated by their distinguishing statements, there are consensus statements that do not distinguish between *any* pair of factors (i.e., they are statements of agreement between two or more factors). The following nine consensus statements were identified: five of the nine reflect characteristics generally agreed to have a high degree of influence on recommendations (rated 3 or 4), while four receive lower ratings (two very low, ratings of 1-2), indicating general agreement that they are less important in Spotify recommendations. See Table 11 for details.

Table 11: Consensus Statements and Ratings

	Rating			
Statement	**Factor 1**	**Factor 2**	**Factor 4**	**Factor 6**
Marking something a "like" (i.e., "heart")	4	4	4	4
How long I listen to a song or playlist	4	4	3	3
How many times I listen to a song, artist or playlist	4	4	4	4
Songs that are similar to other songs I “liked” or listened to	4	4	4	4
Playlists I’ve created	4	3	4	4
Songs or artists that Spotify is promoting	3	3	3	4
The time of day I’m listening	3	2	1	1
Where I am while listening	1	1	2	2
Comments from other people about music on Facebook or other social media	1	1	1	2

Four of these consensus statements reflect the specific actions of users: marking something, listening to something (how long and how many times), and creating playlists. These statements were ranked either “very important” or “important”. The statement “Songs that are similar to other songs I ‘liked’ or listened to”, ranked “very important” in all the factors, identifies a matching process that determines similarity related to user actions (“liked” or listened). The positively rated statement “Songs or artists that Spotify is promoting” identifies the human influence in recommendations related specifically to a company strategy or campaign.

Two other consensus statements reflect temporal (“The time of day I’m listening”) or spatial (“Where I am while listening”) characteristics of the user. However, in these cases the consensus is generally negative (“not important” or “somewhat important”) indicating a belief about the limited influence of these data elements in recommendations. Similarly, the final consensus statement, “Comments from other people about music on Facebook or other social media”, ranked mostly “not important”, minimizes the influence of social media posts in making recommendations.

6.3.3.9.2 Factors

The following section provides an analysis of the four extracted factors (Factors 1, 2, 4 and 6). Each factor is given a descriptive title reflecting its key concepts. The p values are provided for distinguishing statements. Each statement is accompanied with its rating in parentheses. In the tables, distinguishing statements are marked “D” and consensus statements marked “C”.

6.3.3.9.3 Factor 1: About me and what I’m feeling

Factor 1 has an eigenvalue of 10.4 and explains 55% of the study variance. Seven participants are associated with this factor (Users 3, 6, 7, 11, 13, 14, and 16). As indicated in the survey results, these users are all satisfied with Spotify’s recommendations, but they are divided by their musical interests (4 Specialists; 3 Generalists) and how they believe recommendations are made (4 Primarily by algorithms and partly by humans; 3 Solely by algorithms). They mostly listen to Spotify on a smartphone (5 Smartphone; 2 Computer).

The single distinguishing statement for Factor 1 is “What I’m feeling while listening” (3; $p<0.01$). This statement is rated “important” while in all the other factors this statement is rated “not important”. The significance of emotional states with respect to Spotify recommendations is contentious. Users associated with this factor emphasize the influence of their specific actions on the recommendations they receive from Spotify. See Table 12 for details.

Table 12: Relative Rating of Statements in Factor 1

* indicates p<0.01; otherwise p<0.05

	Statement	Factor 1	Distinguishing (D) or Consensus (C)	Factor 2	Factor 4	Factor 6
8	Playlists I've created	4	C*	3	4	4
1	Marking something a "like" (i.e., "heart")	4	C*	4	4	4
3	How long I listen to a song or playlist	4	C*	4	3	3
4	How many times I listen to a song, artist or playlist	4	C	4	4	4
7	Songs that are similar to other songs I "liked" or listened to	4	C	4	4	4
9	Playlists other users have created	4		3	3	3
2	What I listen to	4		4	4	4

19	Songs or artists that Spotify is promoting	3	C	3	3	4
16	The time of day I'm listening	3	C	2	1	1
10	What people my age listen to	3		3	2	3
15	What I'm feeling while listening	3	D*	1	1	1
5	What other people are listening to	3		4	1	4
20	Posts about Spotify I make to Facebook or other social media	2		2	4	2
11	What people in my location (city/country) listen to	2		3	2	3
14	What I'm doing while listening	2		1	2	1
18	The season of the year I'm listening	2		2	4	1
6	What my friends are listening to	2		4	1	2
13	Where I am while listening	1	C*	1	2	2
12	What people with my level of education listen to	1		2	3	2
21	Comments from other people about music on Facebook or other social media	1	C	1	1	2
17	The day of the week I'm listening	1		2	3	1

22	Reviews of music in magazines, blogs, videos, news sources	1		1	2	3

The statement, “Songs that are similar to other songs I ‘liked’ or listened to” (4), is predicated on user actions (“liked” or listened to) but it also indicates the importance of a matching process performed by the system. Users in this factor are divided about whether that matching process is solely algorithmic or primarily algorithmic and partly human.

A secondary group of influences on recommendations is centered on the actions of other people: “Playlists other users have created” (4), “What people my age listen to” (3), and “What other people are listening to” (3). “Songs or artists Spotify is promoting” (3) also indicates the importance of others, in this case the music editors at Spotify. While other people are clearly influential in the recommendations for these users that does not extend to their friends. “What my friends are listening to” (2) is rated only “somewhat important”.

While “Comments from other people about music on Facebook or other social media” (1) is rated as “not important”, consistent with the focus on “me”, the statement “Posts about Spotify I make to Facebook or other social media” (2) rated higher. However, in conjunction with the low rating for “Reviews of music in magazines, blogs, videos, news sources” (1), users associated with this factor find little influence from external media in the recommendations provided by Spotify.

Two demographic measures, “What people in my location (city/country) listen to” (2) and “What people with my level of education listen to” (1), might be expected to align with the focus on “me” (i.e., people like me). However, they are rated with little or no importance suggesting that such data elements carry limited informational value for recommendations.

The rating of “The time of day I’m listening” (3) suggests that temporal and spatial characteristics are important to users associated with this factor. For example, time of day often correlates with where someone is (e.g., work vs. home) and what they are doing (e.g., working vs leisure or domestic tasks). However, “What I’m doing while listening” (2), “Where I am while listening” (1), “The day of the week I’m listening” (1), and “The season of the year I’m listening” (2) are all rated either “not important” or only “somewhat important.” This suggests that temporal influences are limited to differences throughout the day and are unrelated to location or activities.

The users associated with this factor believe the recommendations in Spotify are “about me and what I’m feeling.” Their active and explicit signals to Spotify through what they do on the system are the highest-rating influences on the recommendations they receive. While other factors share this perspective, what distinguishes Factor 1 is the importance of what the listener is feeling. Emphasizing the influence of emotions represents a deeper level of understanding “me”, the listener, in providing relevant recommendations. This raises many questions about whether, and if so how, Spotify collects and utilizes such data.

Spotify provides many playlists associated with moods and feelings (e.g., “Alone Again,” “Sad Hour,” “IDK,” “Feelin’ Good,” “Life Sucks,” and “Happy Beats”). Does Spotify infer emotions from users listening to these to recommend other artists or songs? Are other signals, such as audio characteristics (e.g., tempo, tone, key signatures) or lyrical content, collected to understand emotions? If so, is this accomplished by algorithms, human editors, or both? Users associated with this factor are all satisfied with Spotify’s recommendations, but they are divided on how they are made.

While other people are important in influencing recommendations, it seems counterintuitive that friends are not. This might suggest that users don’t believe Spotify knows who their friends are. While certainly plausible, the close links possible between Spotify and various social media platforms (e.g., Facebook, Instagram) makes the identification of friends possible. In addition, Spotify provides many opportunities in the application for users to identify friends. Alternatively, it is possible that users simply believe they don’t share musical interests with their friends and as a result their friends would have a marginal, if any, influence on the user’s recommendations.

Given the extensive location tracking, inherent and optional, in all the digital devices used by participants, the belief in the lack of importance of location seems naive. Perhaps the belief is that Spotify can't infer from these data any relevant information. It is possible that some statements about data signals were rated "not important" or only "somewhat important" because users believe their information value is insufficient to influence recommendations. Unlike the explicit actions of users, data such as day of the week, where the listener is, or what the listener is doing cannot be directly linked to user preferences. Inferred insights are believed less valuable than directly measured events.

Similarity, in the context of the statement "Songs that are similar to other songs I 'liked' or listened to", implies a process to seek and match songs which will be relevant or of interest to the user. In the case of Factor 1, it must also incorporate the implications of emotional states. Similarity is experienced viscerally by users, but Spotify treats it as primarily computational. Users associated with Factor 1 are all satisfied with Spotify's recommendations but divided about how those recommendations are made suggesting differing views about the respective roles of algorithms and humans in creating recommendations.

6.3.3.9.4 Factor 2: About me and what my social group is listening to

Factor 2 has an eigenvalue of 1.1 and explains 6% of the study variance. Two participants are associated with this factor (Users 12 and 15). As indicated in the survey results, these users are divided regarding their musical interests (1 Specialist; 1 Generalist), are satisfied with their recommendations, use a computer to access Spotify, and believe the recommendations are made solely by algorithms.

The three distinguishing statements for Factor 2 are "What my friends are listening to" (4; $p<0.01$), "What people my age listen to" (3; $p<0.01$) and "What people in my location (city/country) listen to" (3; $p<0.05$). In all the other factors these statements are rated lower. Factor 2 is distinguished by an emphasis on the importance of data signals from others in influencing the recommendations users receive. Highly rated statements in this factor emphasize the importance of the explicit actions of individual users. See Table 13 for details.

Table 13: Relative Rating of Statements in Factor 2

* indicates p<0.01; otherwise p<0.05

	Statement	Factor 2	Distinguishing (D) or Consensus (C)	Factor 1	Factor 4	Factor 6
1	Marking something a "like" (i.e., "heart")	4	C*	4	4	4
3	How long I listen to a song or playlist	4	C*	4	3	3
4	How many times I listen to a song, artist or playlist	4	C	4	4	4
7	Songs that are similar to other songs I "liked" or listened to	4	C	4	4	4
2	What I listen to	4		4	4	4
5	What other people are listening to	4		3	1	4
6	What my friends are listening to	4	D*	2	1	2
8	Playlists I've created	3	C	4	4	4
9	Playlists other users have created	3		4	3	3
19	Songs or artists that Spotify is promoting	3	C	3	3	4

10	What people my age listen to	3	D*	3	2	3
11	What people in my location (city/country) listen to	3	D	2	2	3
16	The time of day I'm listening	2	C	3	1	1
20	Posts about Spotify I make to Facebook or other social media	2		2	4	2
18	The season of the year I'm listening	2		2	4	1
12	What people with my level of education listen to	2		1	3	2
17	The day of the week I'm listening	2		1	3	1
15	What I'm feeling while listening	1		3	1	1
14	What I'm doing while listening	1		2	2	1
13	Where I am while listening	1	C*	1	2	2
21	Comments from other people about music on Facebook or other social media	1	C	1	1	2
22	Reviews of music in magazines, blogs, videos, news sources	1		1	2	3

The statement, “Songs that are similar to other songs I ‘liked’ or listened to” (4), is predicated on user actions (“liked” or listened to) but it also indicates the importance of a matching process performed by the system. Since all the users associated with this factor believe that recommendations are made solely by algorithms, this similarity matching process is entirely algorithmic.

Also highly rated are statements that align with the distinguishing statements for this factor: “What other people are listening to” (4), “Playlists other users have created” (3), and “Songs or artists that Spotify is promoting” (3). While what others do on the system is important, what others say in various media is not. For example, “Comments from other people about music on Facebook or other social media” (1) and “Reviews of music in magazines, blogs, videos, news sources” (1) receive a low rating. Even a statement that ostensibly reflects data signals from the user (i.e., is about “me”), “Posts about Spotify I make to Facebook or other social media” (2) received a low rating consistent with the negative view of the influence of media sources on recommendations.

Many of the statements which reflect passive or implicit data signals from or about the user have low ratings: “What I’m feeling while listening” (1), “What I’m doing while listening” (1), Where I am while listening (1), “The time of day I’m listening (2), “The day of the week I’m listening (2), and “The season of the year I’m listening (2).

The users associated with this factor believe the recommendations in Spotify are “about me and what my social group is listening to.” Their active and explicit signals to Spotify through what they do on the system are the highest-rating influences on the recommendations they received. While other factors share this perspective, what distinguishes Factor 2 is the importance of the social group to which these users belong. This focus on “others”, including “my friends”, is centered on what this social group is listening to. However, additional statements also reflect the influence of other activities. For example, also highly rated are the playlists others create and, in the case of Spotify (another instance of “others”), the promotions made by Spotify’s music editors.

It is curious that the perception of the influence of others in recommendations does not extend to what others do or say in various external media (“Comments from other people about music on Facebook or other social media" and "Reviews of music in magazines, blogs, videos, news sources"). Even what the listener posts about Spotify in social media (i.e., another example about "me") received a low rating. It appears that users associated with this factor either believe there is low information value in data signals from the media, it is not collected by Spotify, or they believe it is not possible for the algorithm to interpret or infer information from these sources to be used in the recommendation process.

Another group of statements reflect passive and implicit data signals about the user, such as temporal or spatial characteristics, emotional states, activities, and locations. All of these received a low rating. While users associated with this factor believe the recommendations are about "me", they do not include in that view information collected passively about themselves. Except for emotional states, these data are easily and widely collected by Spotify's algorithms. As with data signals from external media, it may be that users do not believe the algorithms, which they believe are solely responsible for the recommendations, are able to infer anything useful from these data.

Creating recommendations that reflect similarity with songs previously "liked" or listened to is important to the users associated with this factor. Since all these users believe the recommendations are created solely by algorithms, the matching process, and the definition of what is similar, is entirely algorithmic. From this perspective the algorithms are another example of an "other" that influences the recommendations. It is possible that Spotify editors are viewed by users as algorithmic not human, thereby maintaining a consistent belief about how recommendations are made.

Data signals prominent in Factor 2 are those viewed as easily identified and which lack anything contentious or difficult to interpret. Hence similarity could be solely algorithmic because this is the only type of data an algorithm could collect and manage. In this context it bears remembering that all users in this factor are satisfied with Spotify's recommendations.

6.3.3.9.5 Factor 4: About me, my expressed taste, and not that of others

Factor 4 has an eigenvalue of .9 and explains 5% of the study variance. Three participants are associated with this factor (User 4, 8, and 17). As indicated in the survey results, this group is divided regarding their musical interests (1 Specialist; 2 Generalists), their satisfaction with the recommendations (2 Satisfied; 1 Not Satisfied), and how they listen (1 Smartphone; 2 Computer). They agree that the recommendations are made solely by algorithms.

The four distinguishing statements for Factor 4 are "Season of the year I'm listening" (4; $p<0.01$), "Posts about Spotify I make to Facebook or other social media" (4; $p<0.05$), "What other people are listening to" (1; $p<0.01$) and "What my friends are listening to" (1; $p<0.05$). The latter two statements are rated "not important". In other words, a distinguishing characteristic of this factor is the rejection of the influence of others in making recommendations. This view is further emphasized by the low rating of "Comments from other people about music on Facebook or other social media" (1), "Reviews of music in magazines, blogs, videos, new sources" (2), and "What people in my location (city/country) are listening to" (2). However, in contrast to the minimal influence of others, "Playlists other users have created" (3), "What people my age are listening to" (3), and "Songs or artists Spotify is promoting" (3) are all rated "important."

The distinguishing statements in this factor are rated differently in the other factors. Statements regarding season of the year and posts about Spotify received lower ratings (either "not important" or "somewhat important"). The negatively rated distinguishing statements in this factor were rated higher in other factors with "What other people are listening to" receiving a "very important" rating in Factors 2 and 6. Highly rated statements in this factor emphasize the importance of the explicit actions of individual users aligning with the distinguishing statement regarding posts the user makes on social media. See Table 14 for details.

Table 14: Relative Ratings of Statements in Factor 4

* indicates p<0.01; otherwise p<0.05

	Statement	Factor 4	Distinguishing (D) or Consensus (C)	Factor 1	Factor 2	Factor 6
8	Playlists I've created	4	C*	4	3	4
1	Marking something a "like" (i.e., "heart")	4	C*	4	4	4
4	How many times I listen to a song, artist or playlist	4	C	4	4	4
7	Songs that are similar to other songs I "liked" or listened to	4	C	4	4	4
2	What I listen to	4		4	4	4
20	Posts about Spotify I make to Facebook or other social media	4	D	2	2	2
18	The season of the year I'm listening	4	D*	2	2	1
3	How long I listen to a song or playlist	3	C	4	4	3
9	Playlists other users have created	3		4	3	3
19	Songs or artists that Spotify is promoting	3	C	3	3	4
12	What people with my level of education listen to	3		1	2	2

17	The day of the week I'm listening	3		1	2	1
10	What people my age listen to	2		3	3	3
11	What people in my location (city/country) listen to	2		2	3	3
14	What I'm doing while listening	2		2	1	1
13	Where I am while listening	2	C*	1	1	2
22	Reviews of music in magazines, blogs, videos, news sources	2		1	1	3
16	The time of day I'm listening	1	C	3	2	1
15	What I'm feeling while listening	1		3	1	1
5	What other people are listening to	1	D*	3	4	4
6	What my friends are listening to	1	D	2	4	2
21	Comments from other people about music on Facebook or other social media	1	C	1	1	2

The statement, "Songs that are similar to other songs I 'liked' or listened to" (4), is predicated on user actions ("liked" or listened to) but it also indicates the importance of a matching process performed by the system. Since all the users associated with this factor believe that recommendations are made solely by algorithms, this similarity matching process is entirely algorithmic.

While “The day of week I’m listening” (3) is rated “important” other temporal/spatial statements are viewed as “not important” or only “somewhat important”: “The time of day I’m listening” (1), and “Where I am while listening” (2). However, “The season of the year I’m listening” (4) is highly rated and is a distinguishing statement of Factor 4.

Two of the statements which reflect passive or implicit data signals from or about the user have low ratings: “What I’m doing while listening” (2) and “What I’m feeling while listening” (1). These inferred data signals are not regarded as having a significant influence.

The users associated with this factor believe the recommendations in Spotify are “about me, my expressed taste, and not that of others.” Their active and explicit signals to Spotify through what they do on the system are the highest-rating influences on the recommendations they received. While other factors share this perspective, what distinguishes Factor 4 is the publicly expressed taste of the user on social media and a rejection of the influence of others, including friends. A curiosity is the distinguishing statement “The season of the year I’m listening” (4).

Except for “The day of the week I’m listening” (3) and the season of the year, temporal and spatial statements receive a low rating. Similarly, statements reflecting passive or implicit data signals were rated either “not important” or only “somewhat important.” It appears that users associated with this factor either believe there is low information value in data signals from these sources, it is not collected by Spotify, or they believe it is not possible for the algorithm to interpret or infer information from these sources to be used in the recommendation process.

Creating recommendations that reflect similarity with songs previously “liked” or listened to is important to the users associated with this factor. Since all these users believe the recommendations are created solely by algorithms, the matching process and the nature of what is determined to be similar is entirely algorithmic.

What to make of the distinguishing statement regarding season of the year? It is possible that the importance of both “day of the week” and “season of the year” are related through a connection to religious observances and holidays. This would make certain days of the week more significant (e.g., Saturday for Jews, Sunday for Christians) and certain seasons, especially December (the “holiday season”), more important and noteworthy in terms of recommendations from Spotify.

6.3.3.9.6 Factor 6: About me and the opinions of others

Factor 6 has an eigenvalue of .9 and explains 5% of the study variance. The five participants associated with this factor (Users 2, 5, 9, 19, and 20) are a diverse group. As indicated in the survey results, these users are divided in all the key characteristics: musical interests (2 Specialists; 3 Generalists), how they access Spotify (3 Smartphone; 2 Computer), whether they are satisfied with the recommendations (4 Satisfied; 1 Not Satisfied), and how those recommendations are made (2 Solely by algorithms; 2 Primarily by algorithms and partly by humans; 1 Don’t Know).

The distinguishing statement of Factor 6 is “Reviews of music in magazines, blogs, videos, news sources” (3). This statement is rated either “not important” or “somewhat important” in all the other factors. The emphasis on the opinions of others in Factor 6 is augmented by the following statements: “Comments from other people about music on Facebook or other social media” (2) (a low rating but the highest in any of the factors) and “Songs or artists that Spotify is promoting” (4). Highly rated statements in this factor emphasize the importance of the explicit actions of individual users. See Table 15 for details.

Table 15: Relative Rating of Statements in Factor 6

* indicates p<0.01; otherwise p<0.05

	Statement	Factor 6	Distinguishing (D) or Consensus (C)	Factor 1	Factor 2	Factor 4
8	Playlists I've created	4	C*	4	3	4
1	Marking something a "like" (i.e., "heart")	4	C*	4	4	4
4	How many times I listen to a song, artist or playlist	4	C	4	4	4
7	Songs that are similar to other songs I "liked" or listened to	4	C	4	4	4
2	What I listen to	4		4	4	4
19	Songs or artists that Spotify is promoting	4	C	3	3	3
5	What other people are listening to	4		3	4	1
3	How long I listen to a song or playlist	3	C	4	4	3
9	Playlists other users have created	3		4	3	3

10	What people my age listen to	3		3	3	2
11	What people in my location (city/country) listen to	3		2	3	2
22	Reviews of music in magazines, blogs, videos, news sources	3	D	1	1	2
20	Posts about Spotify I make to Facebook or other social media	2		2	2	4
6	What my friends are listening to	2		2	4	1
13	Where I am while listening	2	C*	1	1	2
12	What people with my level of education listen to	2		1	2	3
21	Comments from other people about music on Facebook or other social media	2	C	1	1	1
16	The time of day I'm listening	1	C	3	2	1

15	What I'm feeling while listening	1		3	1	1
14	What I'm doing while listening	1		2	1	2
18	The season of the year I'm listening	1		2	2	4
17	The day of the week I'm listening	1		1	2	3

The statement, “Songs that are similar to other songs I ‘liked’ or listened to” (4), is predicated on user actions (“liked” or listened to) but it also indicates the importance of a matching process performed by the system. Users in this factor group are divided about whether that matching process is solely algorithmic or primarily algorithmic and partly human.

While not specifically opinions or commentary, the positively rated statements “What other people are listening to” (4), “Playlists other users have created” (3), and “What people my age listen to” (3) suggest that what others do act somewhat like opinions. Users appear to view these data signals as opinions that are then used by Spotify to inform the recommendations the users receive. Users associated with this factor differentiate between in the influence of “others” and “my friends”. “What my friends are listening to” (2) rates substantially lower than “What other people are listening to” (4).

Given the focus of the users associated with Factor 6 about the importance of externally published opinions or commentaries, it might be expected that the statement “Posts about Spotify I make to Facebook or other social media” (2) would rate higher than “somewhat important.” The views of others carry more influence than the views that users themselves post.

Three of the statements which reflect passive or implicit data signals from or about the user have low ratings: “What I’m doing while listening” (2), “Where I am while listening” (2), and “What I’m feeling while listening” (1). These inferred data signals are not regarded as having a significant influence.

All the temporally related statements were rated “not important”: “The time of day I’m listening” (1), “The day of the week I’m listening” (1), and “The season of the year I’m listening” (1). Users associated with Factor 6 appear to discount temporal data signals as valuable in creating recommendations. While Spotify actively collects and uses this information, users in this factor do not see a direct connection between these data and effects on recommendations.

The users associated with this factor believe the recommendations in Spotify are “about me and the opinions of others.” The active and explicit signals to Spotify through what users do on the system are the highest-rating influences on the recommendations they receive. While other factors share this perspective, what distinguishes Factor 6 are the opinions of others. The distinguishing statement, “Reviews of music in magazines, blogs, videos, news sources” (3), suggests the importance of external data sources and expert commentary to recommendations from Spotify.

This factor places an emphasis on human recommendations through publications (reviews, posts, other examples of music journalism or commentary) as a primary influence on Spotify’s recommendations. This view holds despite the belief by users associated with this factor that Spotify’s recommendations are made either solely or primarily by algorithms. Users associated with Factor 6 believe there to be a significant human presence in a process dominated by algorithms. However, it is important to remember that these human influences are collected and mediated through natural language processing (NLP) algorithms that directly inform recommendations.

Statements reflecting passive or implicit data signals were rated either “not important” or only “somewhat important.” All temporal statements were rated “not important.” It appears that users associated with this factor believe there is low information value in data signals from these sources, it is not collected by Spotify, or they believe it is not possible for the algorithm to interpret or infer information from these sources to be used in the recommendation process.

Creating recommendations that reflect similarity with songs previously “liked” or listened to is important to the users associated with this factor. Users in Factor 6 are divided about how recommendations are made indicating differing views about the respective roles of algorithms and humans in determining similarity. This is especially significant given the prominence Factor 6 gives to human influences in providing data signals for recommendations through opinions and commentary.

6.3.3.9.7 Factor Analysis Summary

The factor analysis of the 22 statements regarding the influence of specific data signals on Spotify’s recommendations resulted in four factors exemplified by the distinguishing statements for each group. These factors are the subjective perspectives of these groups of users regarding their beliefs about how the recommendation process works:

Factor 1: “About me and what I’m feeling”

Factor 2: “About me and what my social group is listening to”

Factor 4: “About me, my expressed taste, and not that of others”

Factor 6: “About me and the opinions of others”

While the emphasis on “about me” (i.e., my actions and the resulting influences on recommendations) is common across the factors, the distinguishing statements reflect differences about Spotify’s recommendations. The users associated with Factor 2 emphasize the importance of the actions of other listeners like them (e.g., friends, those of a similar age or in their location). In contrast, users associated with Factor 4 reject the influence of others both in their actions on the system and through external media.

Instead, these users place an emphasis on their own posts to social media. Factor 6 foregrounds the opinions of others (e.g., reviews, social media posts, Spotify promotions) as a distinguishing feature. This factor includes the importance of what others do on the system as a form of opinion sharing (e.g., creating playlists). Users associated with Factor 1 are distinguished by the view that what the user is feeling influences recommendations.

6.4 Discussion

6.4.1 How Recommendations are Made

Participants disagree about how recommendations are made. While 95% emphasize the prominence of algorithms over humans (recommendations are made “solely by algorithms” or “primarily by algorithms and partly by humans”), the majority believe recommendations are made “solely” by algorithms (58%). Since it is known that Spotify’s recommendations are made “primarily by algorithms and partly by humans” (Fleischer & Snickars, 2017; Goldschmitt & Seaver, 2019; Pichl et al., 2017; Popper, 2015; Söderström, 2021a, 2021b), most of the respondents hold an inaccurate view. None of the participants selected the options “Primarily by humans and partly by algorithms” or “Solely by humans.” The respondents in this study do not believe that humans play the main role in how Spotify’s recommendation system works.

Since Spotify’s recommendations are made primarily by algorithms and partly by humans, the inaccurate view denies the role that humans play in the recommendation process. Human influences in Spotify function in direct and indirect ways. What people do on the system impacts data used by the algorithms. The playlists of other users are indirect human influences (Pichl et al., 2017; Popper, 2015). Similarly, data signals derived from external media insert human perspectives and opinions into the recommendation process (Jehan et al., 2010).

In more direct ways, Spotify's music editors hand craft playlists. Editors, working with system developers, create playlists to serve as training data for machine learning systems (Dai et al., 2020; Mehrotra, Shah, et al., 2020; Spotify, 2021d). Developers of the algorithms make decisions regarding their evaluation and optimization that directly impact the resulting recommendations.

Users who believe both algorithms and humans are involved in making recommendations ascribe specific and distinct roles to each. For example, humans craft genre training data, name the genre categories, and validate algorithmic decisions while algorithms generate the actual recommendations. When User 5 described these different roles, they noted that humans "curate" while algorithms "create". Curation suggests a creative, professional activity while creation, in this context, is merely mechanistic. Humans are conceptually engaged in the recommendation process while algorithms, in a more operational manner, construct the outcomes.

In a 2021 addition to a drop-down menu item on the Spotify application, the recommendation process was outlined. This is one of the few, and itself relatively obscure, places where Spotify informs users:

> Our personalized recommendations are tailored to your unique taste, taking into account a variety of factors, such as what you're listening to and when, the listening habits of people who have similar taste in music and podcasts, and the expertise of our music and podcast specialists. In some cases, commercial considerations may influence our recommendations. (Spotify, 2021c)

However, from a user perspective, this "algotorial" process (i.e., algorithms + human editors) (Spotify, 2021d; Stål, 2021) results in a recommendation process "composed out of human and algorithmic parts that are constantly reconfigured into arrangements that make it difficult to distinguish between the human and the algorithmic at any level" (Goldschmitt & Seaver, 2019, p. 72).

In focusing on the role of algorithmic decision-making in the recommendation process, users have contrasting views about the capabilities of the algorithms. This contrast is evident in the rating of “What I’m feeling when listening” where 42% view it as “not important” while 32% rated it as either “very important” or “important.” Whether Spotify can collect data regarding feelings or emotional states and further, whether it is able to use this algorithmically, is controversial and contentious (Crawford, 2021; Elk, 2021; Stark & Hoey, 2021). However, nearly a third of participants attributed these capabilities to the Spotify recommendation algorithms, whether solely or primarily.

6.4.2 Active and Explicit vs. Passive and Implicit Data Signals

Respondents believe their active and explicit actions on the Spotify system have the most influence on the recommendations they receive. For example, “What I listen to” is rated “very important” by 95% of the participants and “How many times I listen to a song, artist or playlist” is rated “very important” by 89% of the participants. The most common responses to the question about how Spotify uses information to create recommendations identified the active and explicit actions of the user (e.g., listening time, “liking” or “hearting” a song or artist, skipping a song, creating playlists).

In contrast, passive and implicit data signals (i.e., those data collected in the background and do not require any user actions) are among those with the lowest ratings, reflecting the least influence on recommendations. As examples, “What I’m doing while listening” is rated “not important” by 53% of the respondents and “Where I am while listening” is rated “not important” by 37% of the respondents. In response to the survey question regarding how Spotify uses information, there are very few mentions of passive and implicit information, such as location, age, and demographics, and none related to activities, time, season, intention, or emotional states. This information is believed to be unimportant, not collected, or not able to be translated into recommendations.

The consensus statements from the factor analysis reinforce this perspective. The highest rated consensus statements indicated active and explicit listener actions (e.g., “liking” something, listening length, listening time, and playlists created). The lowest rated consensus statements reflect passively collected data signals (e.g., location, time of day).

6.4.3 Inferred or Interpreted Data Signals

In emphasizing direct user actions as the principal influences on recommendations, participants underestimate the importance of inferred or interpreted data signals that are widely collected and analyzed by Spotify (Jehan & DesRoches, 2014; Mehrotra et al., 2019; Spotify, 2020b). These are the passive and implicit data signals derived from various sources. For example, Spotify uses machine learning techniques to infer user intentions: “When users interact with the recommendations served, they leave behind fine-grained traces of interaction patterns, which could be leveraged for predicting how satisfied was their experience, and for developing metrics of user satisfaction” (Mehrotra et al., 2019, p. 1256).

An example is the location of a listener as determined by several techniques (e.g., cell towers, GPS, IP address). Since where you are (workplace, home, frequent locations) can correlate to what you are doing (work, domestic activities or leisure and other pursuits), Spotify makes inferences regarding recommendations a listener might like for different situations based on historical data held about the user (Eriksson et al., 2019).

Given the low ratings of inferred or interpreted data signals and their general absence from responses to how Spotify uses information or how to shape recommendations, participants appear to believe that they are either not collected, not important or difficult to interpret as signals for recommendations. The latter suggests that users place certain limitations on the capabilities of algorithmic decision-making. In this view, there are data that cannot be collected or data that the algorithms cannot successfully interpret.

Spotify’s “music retrieval from everything” strategy draws data from a vast array of signals and sources. All the data signals described in the statements participants rated are known to be, or thought to be, collected, and used by Spotify in its recommendation processes (Eriksson et al., 2019; Jehan & DesRoches, 2014; McInerney et al., 2018; Spotify, 2020b). Users appear to be unaware of, or to underestimate the design and ability of, Spotify’s recommendation process to find, extract, and utilize information beyond that directly available through the Spotify application.

6.4.4 Agency

The consensus arising from the survey data and the factor analysis is that recommendations are “about me” and that the active and explicit actions of the listener have the most influence on recommendations. While this emphasizes the control and direction of the user, participants also believe there are ways their agency is limited.

A widely held view is the importance to recommendations of songs or artists that Spotify is promoting. This consensus statement from the factor analysis, is highly rated (“important” or “very important”) by most respondents (68%) and referenced by User 9 in their response to the question about how Spotify uses information in recommendations. Spotify adjusts its recommendations by directly inserting specific artists or songs into recommendations or by altering the weighting of artists or songs to increase their rating and appearance in playlists (Rogers, 2020; Spotify, 2021c, 2021d). It may be possible to reconcile these beliefs if the rating reflects the importance of promotions by Spotify in making recommendations but not the desirability of these interventions.

The limitations of user agency are illustrated by responses to the question regarding how users could shape Spotify’s recommendations (see section 3.2.2). Users described few strategies to purposefully shape recommendations. They also expressed uncertainty about the effectiveness of those techniques (e.g., skipping or disliking songs). In apparent frustration, some users proposed confusing the system (“game” or “confound”) to shift its perspective. With respect to purposefully shaping recommendations, Spotify puts users in a passive not an active role. It collects a vast amount of data to create recommendations but provides few techniques or actions for the user to explicitly alter or shape them. Spotify wants to control the training of the algorithm, rather than allowing the user to participate more explicitly in that process. As User 3 observes, “training the algorithm is a lot of effort.”

6.4.5 Similarity

The statement “Songs that are similar to other songs I ‘liked’ or listened to” is one of the highest rated and is a consensus statement in the factor analysis. Similarity was also referenced with respect to songs, artists, genres, users, taste, habits, and eras in response to the question about how Spotify uses information in recommendations. Users want things that are similar and believe that Spotify can determine that things are similar.

From one perspective the importance of the similarity of songs or artists reinforces the folk theory that the system responds primarily to the explicit actions of users (e.g., “likes” and listening frequency). In this view, similarity equals what users have indicated they enjoyed previously matched against what others indicated they also enjoyed. Finding similar songs or artists is a matching process that reflects an understanding of co-occurrence as the core of collaboration-based recommender systems (e.g., “I like this song you like. I might like this other song you like”).

From another perspective, the importance of similarity foregrounds the complexity of the term “similar” and how it is enacted by the algorithms. Users believe Spotify understands and acts upon the idea of “similarity” with respect to both music and user preferences. As such, Spotify understands differences and similarities among types of music (genres, categories, etc.) and artists. It knows, or infers, a user’s musical taste or objectives in listening (e.g., Specialists and Generalists).

The opposite of similarity, difference (not similar, not equal), is also important as indicated by the reaction of User 18 to the recommended cello arrangement of a Chopin piano piece: “those things are not equal.” What is different, and how that is understood and determined, is as important as what is similar.

Similarity is experienced viscerally by users, but Spotify treats it as primarily computational. This raises questions such as what is defined as similar and by what or whom (algorithms, humans, or both), why does this denote similarity, and how is this similarity detected? Users differ on whether this process is solely algorithm or primarily algorithmic and partly human. The emphasis on similarity as a matching process obscures the more complex musical and listener analysis undertaken with the machine learning processes of Spotify where similarity is operationalized as the cosine similarity of vectorized music and user data.

6.4.6 Collaboration-Based and Content-Based Recommendations

In responding to the questions about how Spotify uses information in recommendations and how users could shape those recommendations, participants described or implied characteristics regarding the type of recommender system used by Spotify. While Spotify employs a model-based, deep learning recommender system design, participants mostly describe a collaboration-based and content-based system.

Participants believe Spotify uses “tags” or “descriptors” to categorize music and users. Recommendations “fill in the gaps” resulting from “matches” and “comparisons” and from items recognized “adjacent” or “related”. The process resembles a contingency table or a Boolean search: “If other users tend to like Bands X, Y, and Z, and I'm known to like X and Z, Spotify will recommend music by Band Y” (User 19). User 13 described the “cold start” problem which is a central challenge of collaborative-based recommender systems (Rana et al., 2020; Ricci et al., 2015). While User 5 does mention machine learning and User 3 refers to “training the algorithm”, the dominate representations of how recommendations are made are co-occurrence, search and retrieval, and a database model. These views reflect the characteristics of collaboration-based and content-based recommender systems (Jannach et al., 2011; Ricci et al., 2015).

In fact, Spotify uses model-based, machine learning techniques, specifically neural networks and reinforcement learning, which attempt to create models of users and music that learn a holistic and long-term view of a user's preferences and music characteristics (A. Anderson et al., 2020; I. Anderson et al., 2021; Dai et al., 2020; Mehrotra, Bhattacharya, et al., 2020; Mehrotra, Shah, et al., 2020). Model-based recommender systems, developed over the past decade, differ significantly from collaborative-based and content-based systems.

Users reflect an understanding of an older, and less powerful, recommendation method that is based largely on transparent notions of automated co-occurrence matching. The complexity and opacity of machine learning systems may be barriers to reconceptualizing how contemporary recommender systems work and the implications of that for the recommendations users receive.

6.4.7 Feelings and Emotional States

Some users believe what they are feeling influences the recommendations they receive from Spotify.The influence of feelings or emotions in the determination of recommendations is contentious with beliefs divided by level of importance, how recommendations are made, and satisfaction with those recommendations. While many participants (42%) are clear that emotions are "not important" to recommendations, a sizeable group (32%) believe "What I'm feeling while listening" is "very important" or "important." "What I'm feeling" is the distinguishing statement for Factor 1.

Does Spotify attempt to collect and understand emotional states? If so, how is this translated into recommendations? Can algorithms alone understand emotional states or is this only possible because humans are involved? Spotify can infer emotional states from the use of emotionally named and themed playlists (e.g., “Sad Songs”, “Happy Days”, and “Angry Rock”) and extract emotional clues from status updates on Facebook. Going further, Spotify has registered a patent to identify a user’s emotional state from audio input (e.g., from a personal assistant) (Huland, 2021) and researchers have developed a deep learning system to classify the emotional content of Spotify’s music database (de Quirós et al., 2019). The possible use of these technologies has resulted in widespread criticism and concern from musicians (Schwartz, 2021), human rights organizations (Oribhabor et al., 2021), and a broad coalition of creators (Stop Spotify Surveillance, 2021). Spotify has responded indicating it has no current plans to use the patent and underscored their commitment to privacy and socially responsible business practices (Elk, 2021).

It is not unexpected that feelings or emotions would be raised with respect to Spotify. Music is an emotional, even a visceral, experience. Listeners both seek and experience affect in their listening. User 14 thinks it would be “cool to give feeling inputs to the algo” to affect recommendations. However, if emotions have an influence in making recommendations, participants believe it is dependent on whether humans have played an explicit role in their creation. Satisfied users who believe feelings are important to recommendations are more likely to believe they are made with some human influence (i.e., “partly by humans”). Algorithms alone are generally viewed as unable to accomplish this.

6.4.8 Dissatisfied Generalists

Most participants are satisfied with Spotify’s recommendations (79%). However, research suggests that Generalists will be less satisfied with recommendations than are Specialists (A. Anderson et al., 2020). While the results are not statistically significant, in this study over a third (35%) of Generalists are dissatisfied compared to none of the

Specialists. Dissatisfied Generalists (Users 4, 9, 10 and 18) are more likely to believe recommendations are made “solely by algorithms”, deny the importance of feelings in making recommendations, and use a device which limits recommendation services to them and restricts Spotify’s access to certain data signals that influence recommendations.

Consistent with the prior research, this suggests that the interests and objectives of these users do not align with the way the recommendation system works. The “popularity bias” of recommender systems may be a determining factor by providing less diversity in recommendations than Generalists would prefer.

Given that all the Generalists are experienced Spotify users (91% have used Spotify for over a year), it may be that these listeners employ techniques or “hacks” to attempt to manipulate the algorithms in service to their interests. Dissatisfied users may be expected to ask for more, and more specific, explanations of how the system performs.

6.5 Conclusion

Respondents in this study are “music lovers:” 90% describe themselves as passionate or keen about music. According to the Phoenix 2 Project, people who are passionate (“savants”) and keen (“enthusiasts”) about music constitute only 28% of the population (Jennings, 2007, p. 46). Respondents are experienced users who have been listening to Spotify for at least a year (95%) with 42% listening for over 5 years. They make frequent use of Spotify with 90% using it every day or most days. While Spotify has a free, ad-supported version, 79% of respondents subscribe to the paid service. Most participants (79%) are satisfied with the personalized recommendations that Spotify provides.

While users disagree about how recommendations are made, all privilege the role of algorithms, either as “solely” or “primarily” responsible. Most users believe that recommendations are made “solely by algorithms,” although this is inaccurate since it is known that Spotify uses both algorithms and human influences and interventions to make recommendations.

Users believe their active and explicit actions are primarily responsible for the recommendations they receive. Passive and implicit data signals, especially those that require inferences and interpretation, are viewed as having lesser or no influence on recommendations.

Users hold beliefs about agency, the nature of similarity, and whether feelings are an important data signal. Agency is largely viewed as contested where both user and system vie for influence. Users believe Spotify can recognize and act upon notions of similarity between and among music, artists, and other users. Whether Spotify uses feelings and emotional states as a data signal is contentious. Users hold different beliefs about this, including its desirability.

The Spotify user survey identified user beliefs (i.e., their folk theories) about how the recommendation system works. These folk theories guide their use of the system and are explanations for how the system behaves. The Spotify user interviews provided opportunities to probe these folk theories in more detail and to draw out how understanding folk theories could inform XAI strategies.

7 Spotify User Interviews

7.1 Introduction

Advocates of a user-centered approach to XAI (Abdul et al., 2018; Doshi-Velez & Kim, 2017; Miller, 2019; Mueller et al., 2021) underscore the importance of the questions that users might, or should, ask in assessing trust and accountability. What are the key user questions that an XAI explanation should address? How should these questions align with a user's folk theories, reasoning processes or assessment methods? By examining and leveraging these questions, an interview strategy was devised for eliciting the folk theories of Spotify users.

Lim et al. identified five "intelligibility questions":

What: What did the system do?

Why: Why did the system do W?

Why Not: Why did the system not do X?

What If: What would the system do if Y happens? and

How To: How can I get the system to do Z?

They concluded that "explaining *why* a system behaved in a certain way, and explaining why a system did *not* behave in a different way provided most benefit in terms of objective understanding and trust" (Lim et al., 2009).

Wang et al. incorporated these questions into a more comprehensive framework linking specific reasoning processes to appropriate XAI strategies or techniques (D. Wang et al., 2019). The framework includes consideration of cognitive biases that result in reasoning errors. While purported to be a “user-centric” view of XAI, the framework is grounded in a decision support context that would more appropriately guide explanations for users who are experts in the domain of the AI. A lay population is not the focus of this user-centric view.

To address “real-world user needs” for explainability, Liao et al. created an “XAI question bank” collecting questions a user might ask about an AI (Liao et al., 2020, 2021). Specific questions were grouped into six general areas of focus: **Data** (i.e., training data), **Output**, **Performance**, **How** (i.e., how a global model works as opposed to a specific instance), **Why/Why not**, and **What if/How to be that/How to still be this** (i.e., getting different or the same results given changes in input). On the premise that “the suitability of explanations is question dependent,” the question bank focuses on the differing needs of users in specific contexts with specific concerns. The questions in each area of focus are structured to move from the general to the specific, enabling a probing appropriate to the explanatory need.

Importantly, these questions were determined by system developers not users. They reflect what developers, not users, believe are important questions. Despite this, the questions echo the results of earlier research that engaged specific user populations.

The question bank was used as part of the user interview to nudge the participants from general observations about the Spotify recommender process to more detailed information about specific aspects. This sequential refinement process allowed the interview to follow the lead of the participants while encouraging elaboration.

The interview with Spotify users contains two key questions. Question #1 addresses “what does the system do to create recommendations?” while question #2 addresses “what can I do to influence recommendations?” See Appendix 7 for the Interview Guide. In order to focus participant responses on elements of machine learning, follow-up questions directly or indirectly referenced the three key machine learning functions: representation, evaluation, and optimization (Domingos, 2012, 2015). Moving from the general to the specific, follow-up questions sought a deeper understanding of concepts raised by the participant. Counterfactual or contrastive questions broadened the conversation by probing areas unexplored by the participant.

Key question #1 is “How does Spotify use information to determine the personalized music recommendations for you?” This is a “How” question focused on the global recommendation model of Spotify but centered on the experience of the user. Follow-up “Why” and “Why Not” questions probed contrastive and counterfactual alternatives (Miller, 2019). Questions about the training data used (“Data” questions) explored system transparency and limitations.

These possible follow-up questions, used by some but not all interviews, are grouped by a focus on machine learning functionality:

a) Representation

1. What kind of data does Spotify learn from?
2. How does Spotify learn? For example, humans learn by a variety of methods: repetition, trial and error/reward, exploration (play, curiosity), from others, from self-direction
3. How much data is Spotify trained on?

b) Evaluation

1. What criteria does Spotify use to make recommendations?
2. What constitutes success in Spotify recommendations?
3. Why is a song a good match for you?

c) Optimization

1. How does Spotify determine the best recommendation for you? What information is being given priority?
2. Spotify has over 400 million users (some free, some paid), music by millions of artists, is owned by the 3 largest music publishing and distribution companies in the world, and, for the free version, is supported by advertisers. Given these different stakeholders, whose interests does it prioritize? How does it do that?

Key question #2 is “What could you do to shape the personalized music recommendations you receive from Spotify?” This question focuses on the relationship between input changes and output results. (i.e., “What if”). Follow-up questions probed the role of the user, the agency they possess, and how recommendations are a sociotechnical assemblage of user and system.

These possible follow-up questions, used in some but not all interviews, are grouped by a focus on machine learning functionality:

a) Representation

1. What data is Spotify *not* using?
2. What are the limitations of the data Spotify uses?
3. How do you retract or remove information from Spotify about you and your interests?

b) Evaluation

1. Spotify creates and maintains a Taste Profile about you. What do you think is in the Taste Profile? What does Spotify use it for? Do you think your Taste Profile is accurate? Why/Why not?
2. Your musical interests and tastes shift, over time, sometimes frequently. How does Spotify recognize this and respond to it?

c) Optimization

1. How often does the system make mistakes?
2. How accurate are Spotify’s recommendations?
3. Are the Spotify explanations of why you are being recommended playlists etc. useful? If so, why; if not, why not? What would a good explanation be like?

7.2 Methods

All user interviews were conducted and recorded via Zoom. Throughout the interviews, participants were reminded of their responses to survey questions (participants were encouraged to download their responses after they had completed the survey; a link was provided to a PDF version of their answers). In all cases users provided additional details and observations when prompted with their survey responses. In some cases, users reconsidered aspects of their response making different assessments of how Spotify used information or how they could shape recommendations.

The raw audio of the interviews was transcribed by Transcript Heroes, an external service (see Appendix 8 for the confidentiality agreement). Transcripts were checked against the raw audio of the interviews to correct errors or misinterpretations by the transcribers. Personal identifications in the transcripts were removed or concealed. As part of a "member check" validation, users were given the opportunity to review their interview transcript and make changes, adjustments or clarifications that would better represent their ideas and opinions. Most participants (13 of 19) acknowledged receipt of the transcript and requested no changes. User 11 clarified a reflection on Google's impact regarding surveillance. User 19 made extensive additions to sections of the interview regarding personal data collection by Spotify, algorithmic analysis of classical music, and efforts by Spotify's recommendations to expand music listening. Four users did not respond to the member check.

The transcriptions are verbatim with indications of pauses (e.g., "..."), reactions (e.g., [laughs]), and interjections by the interviewer or participant (e.g., "oh"). However, verbal tics such as "um," "uh" or "you know" were removed for clarity. The word "like" presented unique challenges. For many people, "like" is a verbal tic (e.g., "I was like listening to Spotify all the time"). These uses of the word occur frequently in the interviews and have been removed to avoid confusion with other uses. In the context of

Spotify, where users can “like” a song (i.e., mark it with a heart symbol to indicate enjoyment and creating a signal to Spotify’s algorithm), “like” and “liked” are actions in the system. For example, “I liked the new Taylor Swift song” or “Every week I like everything in the Discover Weekly playlist.” Where the words “like” and “liked” refer to actions on Spotify, they have been placed in quotation marks (i.e., “like” and “liked”). The other uses of the word are apparent from their context.

Following the interview and any follow-up, participants were sent a $25 e-gift card from Indigo Books and Music. One participant (User 15) asked that a donation be made to a specific charity instead and this was done. All user interviews were concluded by March 5, 2021.

7.2.1 NVivo Analysis

Using thematic analysis, data from the Spotify user interviews were interpreted within a constructivist framework. Widely used in LIS research, constructivism views individuals as “actively constructing an understanding of their worlds” (Bates, 2005, p. 11) and understands “knowledge production as the creation of mental models” (Talja et al., 2005, p. 83). More specifically, this analysis is informed by Personal Construct theory (Kelly, 1955) where constructs (i.e., mental models or folk theories) are “created by an individual, personally. Its reality exists, not in the things themselves, but in the interpretative act of the individual person” (Fransella, 2016, p. 1). In Personal Construct theory, located within the constructivist framework, people “experience their understandings as representations of a presumed external world” (Raskin, 2016, p. 34). From an LIS perspective, this approach “takes individual searchers [or Spotify users for example] and their interaction with information retrieval systems as its research object and takes the view that work tasks provide the primary context for information behaviour” (Talja et al., 2005, p. 92). These perspectives are consistent with the objective to elicit and describe the folk theories of Spotify users.

While constructivism views language narrowly as “a neutral vehicle for reporting observations and a (more or less clear) window to the speaker’s mind”, this study adopted a more constructionist approach to the understanding and interpretation of user interviews where language “is constitutive for the construction of selves and the information of meanings” (Talja et al., 2005, p. 93). As a result, this study used latent thematic analysis that attempts “to identify or examine the *underlying* ideas, assumptions, and conceptualizations—and ideologies—that are theorized as shaping or informing the semantic content of the data” (Braun & Clarke, 2006, p. 84).

The thematic analysis proceeded through five stages: data familiarization, coding, theme development, reviewing themes, and defining themes (Braun & Clarke, 2006; Terry et al., 2017). The process followed the “checklist of criteria for good thematic analysis” (Braun & Clarke, 2006, p. 96). The full corpus of interviews was reviewed multiple times before coding proceeded (i.e., checking transcripts against the raw audio, editing transcripts to remove personal identifications and normalize certain phrases, and a final close reading while taking general notes).

Coding is a process of both data reduction and synthesis (Terry et al., 2017). Using NVivo, an iterative process of coding and recoding was undertaken, focusing on key concepts, consolidation, patterns, and finally the identification of themes (Jackson, 2019; Saldaña, 2021). See Appendix 9 for the codebook used in this analysis.

However, “data are not coded in an epistemological vacuum” (Braun & Clarke, 2006, p. 84). As a result, an LIS colleague of the researcher, familiar with machine learning and XAI, recoded two of the most heavily coded interviews (Users 2 and 13). The objective was not intercoder reliability with an acceptable Cohen’s Kappa score but rather to “yield concepts and themes (recurrent topics or meanings that represent a phenomena)” that the researcher had missed or interpreted differently (N. McDonald et al., 2019, p. 72:13).The focus was on trustworthiness with the coding review as an assessment of dependability (Lincoln & Guba, 1985; O’Connor & Joffe, 2020). The coding review did not suggest augmenting or altering the codebook. However, it did highlight uncoded or undercoded sections relevant to key issues regarding user perceptions of their actions, subtle forms of resistance to the recommendation algorithms, and the impact of the survey on ideas raised by users in the interview.

The themes derived from the interviews are indicators of the folk theories held by the participants about how Spotify's music recommendations are made. Themes are "meaningful patterns in the data" (Morgan, 2018, p. 340) derived from "the conceptual linking of expressions [i.e., codes]" (Ryan & Bernard, 2003, p. 88). Rather than themes being "emergent" (Charmaz, 2014), they are the result of an "active process of pattern formation and identification" (Terry et al., 2017, p. 13) and partly "a question of prevalence" (Braun & Clarke, 2006, p. 82). The identification and review of themes resulted from an iterative analysis "to ensure that the themes work well in relation to the coded data, the dataset, and the research question" (Terry et al., 2017, p. 16).

7.3 Results and Discussion

The following sections report on the analysis of the interviews focusing on five areas:

1. the objectives and goals users had for using Spotify's personalized recommendations,
2. whether users were satisfied with Spotify's recommendations
3. how users believe recommendations are made,
4. user beliefs about specific factors, concepts, information, and techniques related to the recommendations, and
5. special groupings of users with specific characteristics.

7.3.1 User Objectives and Goals

Why do people use Spotify? What do they want from it? Confronted with questions about how the recommendations work, what information is used, and how they could shape those recommendations, many users indicated that they simply wanted Spotify to provide them with music to listen to:

> "I would rather just go and listen to my music" (User 12),
>
> "I'm there for the music and I just click continue on" (User 2), and
>
> "I just want the endless library" (User 18).

For many, Spotify is "a playlist machine" (User 8) and an "environment of constant music" (User 4) where users "put it on cruise control" and let the system "take the wheel" (User 5). They are untroubled by how the system works and the implications for their listening: "I often don't spend a lot of time actually thinking of what I'm listening – it's just there" (User 5).

Whatever the implications, users have struck a "bargain" with Spotify (User 19) and for most it's a "bargain that I've struck with them [which] I'm good with" (User 3). It is a bargain that sees issues such as privacy and algorithmic manipulation as inconsequential because it is "just music" (Users 7, 8, 12, and 16) and "just my music listening habits" (User 17). However, with respect to the mining of their social media, User 15 draws a line: "that's not worth the exchange, to make better recommendations by mining everything I've said."

Some users describe a changed attitude towards music listening because of personalized recommendations and a streaming service. There is little desire to search for more information or engage beyond the minimum required interaction: "you just kind of click a button and it's all there. I don't think there's as much motivation to have to go out and look up what they're all about" (User 17). User 16 described a link between using recommendations and a more passive approach to listening:

> I feel like lately my musical experience has been a little more passive and impersonal. In another phase of my life, I would have been more active and actually searched and looked through. But I think it’s less – I don’t know, it’s harder to do so when there’s so much. And I guess that’s the point of curation, and playlists, and recommending things.

While the discussion during the interviews raised curiosity in some to think more deeply about how Spotify works, it might have been short lived, as indicated by User 3 at the end of their interview:

> I'm going to spend the next little while just poking around at Spotify. See what else I can see that has been invisible to me. And then also probably put some of this aside and just go forth listening to my things that I listen to.

7.3.2 Satisfaction

Most participants are satisfied with Spotify’s recommendations (79%), although in some cases that satisfaction is qualified: “I’m reasonably satisfied with them” (User 11); “I’m just not unsatisfied” (User 5). One reason given for user satisfaction is alignment with the user’s interests: Spotify does “a good job of matching my music tastes” (User 12) and is “good at anticipating what kind of music I would be into” (User 14). Another reason is the importance of discovery: “I’ve stumbled upon things that I would have never found [by myself]” (User 14) and “I use it more and more because they keep recommending me more and more things that I find I really like” (User 19). User 8 is clear about their role regarding the algorithms: “I find that I just have trained those so well that on both those playlists [Discover Weekly and Release Radar] I like almost every track so much.”

For User 2, Spotify “gives me all the stuff I always liked, along with some newbies” but they are concerned that they don’t know what is “missing” from the recommendations. This is a “black box” problem. What are they not being recommended that they might like? Without naming them as such, User 2 wants some relevance and recall data. In a similar manner User 16 is concerned that Spotify “silos me into a particular style” and an “acoustic echo chamber.” It hasn’t facilitated “musical exploration.”

7.3.3 How Recommendations Are Made

User 3 noted that “an algorithm doesn’t enjoy listening to music ... an algorithm doesn’t understand the emotional or cultural role that music plays in our lives ... even if an algorithm could make the best recommendations possible, what does that mean?” Since all users believe recommendations are made either “solely” or “primarily” by algorithms, this question is at the center of how people perceive recommendations.

7.3.3.1 On Knowing and Not Knowing

When asked to describe how Spotify makes its recommendations or how they could be shaped, it was common for users to indicate their lack of knowledge. Some were clear about their lack of knowledge (“I don’t know”), some expressed uncertainty (“assume,” “guess”, “presume”, “imagine”) and others gave conditional responses (“it might be”).

The vocabulary of users acts as a proxy for their general knowledge of algorithms and artificial intelligence. While the terms “algorithm(s)” were used by all participants and “AI” or “machine learning” by some, few other general AI terms were used (e.g., neural networks, deep learning) and none of the specialized vocabulary of AI (e.g., vectors, matrix, reward, optimization). User 5 is unique among the participants in providing a broad but generally accurate description of machine learning:

> I'm only sort of superficially aware of how machine learning works but my understanding of that is essentially it's like a really complicated statistical model of you, feed it in a bunch of data, it looks for trends and patterns and tries to correlate them between say different people, different types of behaviour. So if we have all of this mass of data connected to a certain individual or profile, like me for example, they can take all the things I do or listen to or are interested in and so on and stuff and compare my data profile with the data profiles of other people.

Most participants in this study expressed little or no technical knowledge of algorithms or machine learning. When asked directly about their general lack of understanding of the technology most were unconcerned: "if I really wanted to know, I could find out. I think that that's how a lot of people feel." (User 12). The specifics and details are unnecessary because "knowing the basics I think is enough for this platform" (User 17).

User 13 feels differently: "I find it unsettling that I don't know specifically how it works ... you are being surveilled without fully understanding how that is happening, or what is being done to surveil you, I think is totalling concerning." Knowing the specifics is important because there's "an element of agency involved [so] I can make choices about what information they gather and what they don't."

While User 4 concurs, "it does concern me", they acknowledge "it doesn't concern me enough to do the work and figure out the whys and the wherefores of it". Spotify (and other machine learning systems) are complex. Trying to understand them requires a cognitive effort that many are unwilling or unable to do. However, User 4 assumes that they, rather than the system, must "do the work" to reveal the "whys and wherefores."

7.3.3.2 Solely by Algorithms

In the survey 11 participants (58%) indicated that Spotify's recommendations are made "solely by algorithms." During the interviews, 6 participants changed their view and now believe that recommendations are made "primarily by algorithms and partly by humans." Five participants maintained their original view (Users 4, 9, 11, 12, and 13). As a result, only 26% of participants held at the end of the interview the inaccurate view that recommendations are made "solely by algorithms." This group includes 2 dissatisfied Generalists (Users 4 and 9) and a Generalist (User 11) who is tentatively satisfied ("reasonably satisfied ... it's decent enough"). The remaining two are both satisfied Specialists (Users 12 and 13).

While maintaining a belief in the sole role of algorithms, User 11 acknowledges that humans have an indirect influence:

> I recognize that other people [i.e., other users] put playlists together on Spotify and share those. But even the visibility of those is algorithmically bound. So the way that I look at it is even if the data is being generated by humans it's fundamentally how you're seeing it and what's being chosen to be shown to you is all completely algorithm.

For many, the belief in the exclusivity of algorithms is based on the scale of Spotify. With over 80M songs and 400M users, providing human created or even human influenced recommendations is "just a bit too much for a team of people to be able to do"; it involves "literally no human" (User 4).

Despite relying on the algorithm for their recommendations, these users have an elementary view of how the algorithm works. Even though Users 12 and 13 mention the phrase "machine learning", these and other users in this group offer mostly non-technical, metaphorical, and general terms regarding how recommendations are made. The algorithms “scroll through” the data (User 4), finding “patterns” (User 11) using a “formula” (User 13), “categorize” genres (User 12) and utilize “some sort of value” to decide (User 11). Information is “filtered” (User 4) and “crunched together” (User 11). The algorithm “pushes or deters” things (User 12). The result is a “really weird Venn diagram” (User 13). User 11 describes this algorithmic process as “probably smart enough.”

Given this, it is unsurprising that these users acknowledge that they “don’t know specifically how it works” (User 13), “don’t know the ins and outs of the Spotify algorithm” (User 12) and find their lack of knowledge “unsettling” (User 13). While User 11 believes recommendations are made with “two information buckets ... critical reviews from reviewers and ... user response”, they “wouldn’t be sure how it would aggregate all of that, and then how it would determine how useful that would be on an individual basis.” Echoing the beliefs of many in this group, User 4 says “we’re not so much into getting under the hood, you know. We’re fine as long as it works. We’ve moved on from the guy in the garage. We don’t want to know how things are, we just want the car.”

While many participants in this study didn’t notice the Spotify explanations at all, users who believe recommendations are made solely by algorithms not only noticed the explanations, but also used them to validate the recommendations. User 12 wants “to see if they [the explanations] were right.” User 4 sees the “Made for You” explanation as “a bit of a challenge to see whether it actually is [accurately reflecting their interests].” For these participants, the explanations are less about “why”, “why not” or “how” than they are a competence or accuracy verification on the performance of the algorithm.

User 13 believes that an algorithm, rather than a human, was making recommendations because it felt less like being surveilled (“part of me is scared to potentially know that there’s a lot of humans in the background”). User 13 indicated that an algorithm has “certain limitations” about what it can know and do; these limitations offer “a comfort thing” in providing some degree of privacy. Ironically, it is not the power of the algorithms to create satisfactory recommendations but their limitations and their deficiencies that instill a sense of confidence or trust in the algorithms.

User 13 was asked if they could change data in their Taste Profile would they be concerned it would have a negative impact on recommendations (e.g., less relevant, too popular, less satisfactory):

> let's say I do edit this information and Spotify starts giving me recommendations of things that I just don't like, would I necessarily call that a negative thing, or associate a value judgment of it being bad? I would not say that it would be bad. I would not be worried that my experience with Spotify would be ruined. More than anything, I would just think like, 'OK. That's fine.' Like a small perk that I had before, isn't happening as frequently now.

User 13 acquiesces to the algorithm as if it isn't accountable to their preferences. User 11 suggests "it's less trust and it's more a willingness to lose." They delegate agency and power to the algorithm in part because "we're not exactly fully cooperating here because Spotify is still doing a lot that we don't necessarily know" (User 13). User 4 imagines how Spotify would describe itself: "this is who we are, take us, this is what we provide, this is what we do."

7.3.3.3 Solely by Algorithm Users Who Changed Their Beliefs

During the interviews, 6 participants changed their view and now believe that recommendations are made "primarily by algorithms and partly by humans" (Users 5, 6, 8, 15, 17, and 18). This group includes 3 satisfied Generalists (Users 5, 8, and 15), 2 satisfied Specialists (Users 6 and 17), and a dissatisfied Generalist (User 18). As a result of these users changing their beliefs, 68% of the study participants now believe the accurate view of how recommendations are made.

When questioned about their beliefs during the interview, users in this group acknowledged that humans were involved in the "overall process" (User 5), "in the back" (User 8) and as "product managers" (User 17). Humans are "changing the weightings"

(User 15) and are able to “tweak the algorithm” (User 6). Spotify music editors defined genres and created playlists as training data for the algorithms (Users 17 and 18). More directly, Users 6 and 18 recognize that developers (humans) created the algorithms and are responsible for what they can and cannot do.

The identified human roles and influences were primarily in oversight (validation and verification of the algorithms) and model training. The limited role for humans was largely because of the scale of Spotify: “I just couldn’t imagine humans are going in there and looking at people as individuals ... that’s why I thought it was like mainly algorithms, just because of the scale of it” (User 8).

It is significant that these users altered their beliefs from the survey. The interviews identified contradictions or deficiencies in their beliefs that they could no longer support. Folk theories can offer “resistance to counter-evidence” (Gelman & Legare, 2011, p. 391; Kulesza et al., 2012) allowing for “inconsistencies, gaps, and idiosyncratic quirks” to persist (Norman, 1983b, p. 8). In this case, however, these users choose to modify their beliefs. Seemingly a small change (from “solely” to “primarily”), this represents significantly different beliefs about recommendations.

7.3.3.4 Primarily by Algorithms and Partly by Humans

For those who consistently believed that the recommendations were made primarily by algorithms and party by humans, the key considerations are about the distinguishing roles and how they interact. As with those who retracted the exclusive role of algorithms, this group sees the human role in terms of oversight and training with the algorithm in a user-facing, decision-making role.

According to the participants, humans “create lists and the algorithms mine the lists” (User 10). Humans are responsible for “naming the categories and validating algorithmically suggested relationships” (User 3). While primarily algorithmic, recommendations are from “human-generated metadata” (User 7) and it is a human that makes the “initial judgement and trains the algorithm” (User 16). Algorithms can identify “a trend” but humans can detect “habits that people have” (User 14).

For some, the human role is limited. As User 6 noted, “there might be a human to tweak the algorithm. But it's much more of a maintenance thing rather than an active participatory type of thing.” Agreeing, “the more I think about it, the less human it has to be,” User 19 believes an algorithm creates better recommendations “because it's more resourceful in some ways”. In a nuance provided by this group, humans “decide what the algorithm will do” (User 3) and “determine what the algorithm will decide is what” (User 16). Deciding “what is what” and what an algorithm “does” suggests an ontological perspective.

User 14 quantifies those divisions of responsibility with respect to different recommendation products: recommendations are 75% algorithms and 25% humans, but the playlists are 75% humans and 25% algorithms. The algorithms choose the songs in the playlists but the aggregation of them into a playlist (i.e., genre, mood, category) is directed by humans. In the view of User 20, users are the humans involved: “I felt like I was the human in that situation trying to influence the algorithm”; the algorithms are “25% me and 75% Spotify.” User 16 sees a collaboration between the user and Spotify editors and developers in developing the algorithm: “I’m training the AI ... [and] the staff are further curating ... there’s a feedback loop.”

The nature, description, analysis, and recommendation of genres are key points of contention for many participants. Since it is viewed as “impossible to have a machine completely discern genre ... it's too cultural and too messy” (User 16), there is a “relatively important human component ... [in] naming the categories and validating algorithmically suggested relationships” (User 3). While algorithms can “track the genre, and they measure the beats, beat count, and they try to identify the mood of things” (User 16), “humans are better at making connections between genres” (User 3).

Despite the complementary roles of algorithms and humans, algorithms have specific weaknesses or limitations. Algorithms can “turn numbers and suggest relationships [but] I wouldn't trust it to absolutely know that someone who likes X will like Y" (User 3). Given the complexity of classical music, User 19 doesn't "trust the sophistication" of an algorithm to recommend that genre to them. As a result, they don't listen to classical music on Spotify.

The relationship between algorithms and humans is symbiotic and rests on their respective strengths. Algorithms "identify patterns that aren't visible to humans ... things that weren't anticipated ... [however] "if we totally relied on data and algorithms, we would lose things" (User 3).

7.3.3.5 Shaping Recommendations

In response to the survey question about how users might shape Spotify's recommendations, participants were unsure, offered few specific strategies, and often expressed frustration about the lack of options and the questionable effectiveness of what was provided. The interviews continued those themes. Users 13, 14 and 20 were unaware there was a way to dislike a song or artist. User 2 doesn't think removing a playlist as a strategy to shape recommendations "would really delete the information from Spotify."

The frustration at being unable to satisfactorily shape the recommendations revealed several adversarial strategies. User 10, who stated in the survey that they wanted to "game their algorithms", acknowledged in the interview that this was because "I don't like them." The adversarial strategy of User 2 was "putting your computer on mute and just letting the pop music roll for days." Other strategies do the opposite by using a "private session" where your activities are temporarily not tracked. With respect to their interest in classical music, User 19 has chosen to "hide it from Spotify" because the "complexity" of this music is not well understood by the system. User 13 admired the strategy of User 6, naming their playlists after food (e.g., "leek and chive soup"), because "what is the algorithm going to do with a title like onions and chives?".

Unlike many participants who are concerned about the amount of personal data Spotify collects, for User 10 the problem in shaping recommendations is insufficient personal data: “if Spotify sent me a questionnaire today that asked me more personal questions about my lifestyle, there's more data points to pull together, then that might help.” They wanted Spotify to recognize that “we’re not all the same individual units.”

7.3.4 Data Signals

During the interviews, participants discussed a variety of data signals and their beliefs about how these influenced the recommendations they received. These observations, for the most part, applied to beliefs about recommendations made solely by algorithms as well as those made primarily by algorithms.

7.3.4.1 Active and Explicit Actions

Consistent with the survey results, during the interviews users repeated and reinforced the importance and significance of active and explicit data signals in how Spotify makes recommendations (e.g., ‘liking,’ listening, frequency and dwell time, and creating playlists). User 3 provided a concise overview and underscores the comprehensiveness of the data capture:

> It records everything that any user does, in any interaction. Every half-completed search, every click, every “I spent 30 seconds on this page, I spent a minute on that page,” whatever it is. They will have recorded absolutely every interaction any user has ever had with their interface.

User 11 views these interactions as tracking user behaviours and responding with system behaviours: “what’s your behaviour immediately after being recommended a song. And I think it compiles all that to determine how successful a recommendation is. And then if it is successful, it repeats that behaviour."

While active signals are important, they are not, for some users, deliberate attempts to convey information to Spotify: "I don't really think that consciously about, OK by saying this I am telling the algorithm something. It is pretty incidental what they get. I just know that they get it" (User 3). Others seek very deliberate ways to "message" the algorithm: "Recommendations can be made solely by the title of the playlist that you're providing ... and so that's what I meant by messaging" (User 13). Naming playlists is one of the few ways Spotify users can employ techniques like folksonomy and tagging to express personal views or categorizations.

As with the generally low importance placed on passive and implicit data signals noted in surveys, users in the interviews continued to express doubt about these: "The system doesn't know what I'm doing. I know what I'm doing, and I choose based on knowing what I'm doing and then I tell it what I'm doing" (User 10), and "if you're not directly telling it then I don't know if it would fully know exactly what you're doing" (User 17).

7.3.4.2 Feelings and Emotional States

Participants are divided about the importance of how they are feeling to the recommendations they get from Spotify. In the survey 42% thought it "not important" but 32% deemed it "very important" or "important". The interviews confirmed this division and allowed participants to expand on the desirability of feelings as a data signal and how Spotify might capture and use that data.

Music choices are used for mood regulation (Lonsdale & North, 2011; Roe, 1985; Thoma et al., 2012) and to match emotional states (North & Hargreaves, 1996). Unhappy people are more likely to listen to sad songs (Saarikallio & Erkkilä, 2007). However, while music listeners are skeptical and cautious about inferences regarding emotional states, they are also interested in the possibilities (Cowen et al., 2020).

The use or increased use of mechanisms or inferences to determine a user’s emotional state has been widely criticized (Oribhabor et al., 2021; Schwartz, 2021; Stark & Hoey, 2021; Stop Spotify Surveillance, 2021) suggesting that these techniques are “extracting more about a person than they choose to reveal” (Crawford, 2021, p. 167).

One participant cautioned that assessing the moods of users “depends on how you’re conceptualizing mood” (User 15). The definitions and boundaries of moods or emotional states are fluid at best. In addition, inferring a user’s mood is a “false equivalency” (User 18) since “I listen to ludicrously depressing stuff when I’m very happy and the other way around” (User 11). However, “I think that they try very hard to know what you’re feeling and use that as something to serve up music” (User 3). In disagreeing with the notion that Spotify uses feelings as a data signal, some are not confident in their beliefs: “I don’t think that it really could get a beat on how I was feeling or anything like that. [But] I wouldn’t be surprised if I’m wrong” (User 11).

While User 15 says their emotional state is “none of their business”, Spotify operates differently, devoting considerable effort to creating a vast number of playlists grouped around mood or emotional states and recommending these playlists to users. Data on a user’s feelings would be “massively informative for what kinds of music that would be effective as a recommendation” (User 5). Some users are happy to comply: “I think it could be cool to give feeling inputs to the algo and have it deliver a mix based on those things” (User 14). Spotify currently provides many playlists focused on specific emotional states (there are dozens for “sad”, “angry” or “jealous” for example). While matching user mood with an appropriate musical mood is desirable for some (“I really do like it when it matches,” User 2), others are concerned about manipulation by recommendation: “perhaps if you start with a very depressing playlist maybe it does try to correct you in a more positive direction” (User 17). User 15 thinks Spotify’s motivation is less in “manipulating or understanding my emotional state [and more] as a way of provoking continuing engagement.”

In discussions about assessing a user's emotional state, User 3 observes that "the only cues that it's getting are the ones that I'm feeding it." This reliance on the active and explicit actions of the user contrasts with the more widely held belief that Spotify infers the feelings of users from the music they are listening to (Users 5, 13, and 17). However, this inference process is viewed as difficult and imperfect: "there's a lot of teasing out that would need to be done to fully articulate what Spotify does to assume mood in a person" (User 13) and "I think it thinks that it knows better than it does" (User 3).

7.3.4.3 Feedback and Learning

Most users are aware that their feedback is essential to creating effective and satisfying recommendations: "Spotify only works because they [listeners] are teaching it to work", (User 19). Users are teachers and Spotify is a learner. User 10 compares the cycle of feedback and recommendation to buying and wearing a new pair of pants: "I make them my own by using them." Others express this in more conventional terms: "I feel like I'm feeding it, it feeds me" (User 19); "I'm training the AI ... the staff are further curating, and the human playlist creators are also training the AI and there's a feedback loop going on" (User 16); "if I like a band I'll 'like' them or I'll 'like' the song. I'm in the habit of doing that because I know it'll change the algorithms" (User 20). Feedback and learning (e.g., "teaching" and "training") are often linked in the responses of users.

While users recognize that a recommender system is a "bargain" between user and system, participants differed on the extent and value of their feedback. Some were engaged and pleased: "I am a very active user of tailoring and training Spotify so that I can spend ultimately less time having to search for new music because it will give it to me already" (User 8). Others are less inclined and questioned the value: "Training the algorithm is a lot of effort" (User 3); "it takes a lot of time [to train the algorithm] and I want it to work faster and more intuitively" (User 10). User 15 doubts the widespread access to personal data is "worth the exchange, to make better recommendations."

The time and cognitive effort in training Spotify is also questioned: "how much work am I willing to put in? Well, some ... [but] I am not willing to spend time randomly ... that all takes a lot. It's a project" (User 10). User 3 acknowledges the cost/benefit when the recommendations are good enough: "spending the time to do all of those actions that I laid out to get the absolutely pristine version of me telling it what I like through all the ways that I think that I can tell the black box what I like. I don't need to do that because I'm getting out of it what I want."

Several users describe their activity on Spotify as unpaid labour: "my choices, my preferences, are being harvested for their algorithm ... [and this is] the product people are paying for" (User 15). User 19 saw this as "doing some of their work" for which they "don't get compensation." Unlike users who want options and means to contribute more information to Spotify (largely to improve their recommendations), User 3 noted that "I'm already doing labour in terms of data creation for them. I'm not going to volunteer to do more of it if there's no benefit back to me." Spotify is perceived as an exploiter.

Users were often unclear about how their feedback affects the system. User 15 noted that "I'm probably not training the machine learning algorithms that Spotify is using in an optimal way" because the various actions they take (e.g., liking a song, adding it a playlist, later deleting it, and then adding it to another playlist). Given all these data signals "what does that mean?" (User 15). User 9 was also skeptical: "whatever it is that they're getting from it isn't useful ... so that doesn't make me think that giving them more is going to improve things that much". Others think the deficiency is theirs: "It requires probably more about me" (User 10); "I don't know if I've intentionally tried to train the machine" (User 16). Some users are unclear if their feedback has been received: "Spotify is really good at adding them, but not so keen on removing songs" (User 20).

Users 9, 10, and 15 expressed frustrations at the limited ways they can interact with Spotify's algorithms and want acknowledgement of what actions had an effect, if any at all: "give me a bigger vocabulary and then make it meaningful. Then prove to me that you've heard me" (User 10). The desire for a "bigger vocabulary" recognizes the generally passive and largely imprecise data signals users can provide. It also suggests the need for different modes of interaction; modes that enable nuance as well as clarity (i.e., “meaningful”). The urging that Spotify “prove to me” suggests the need for explanations of cause and effect, and echoes the comments of others that Spotify is “this giant black box” (User 13) and “a complete black box” (User 3).

7.3.4.4 Monitoring: Location, Time, and Inference

Location can be correlated to what a user is doing and how understanding of that might influence the recommendations they receive. User 14 described this:

> I think it definitely knows when I’m at work and what I listen to. ‘Cause I listen to a lot of instrumental when I’m at work, or study playlists, lo-fi, hip-hop beats, or whatever. And I don’t see any of that recommended now. It’s Saturday.

Even though some “don’t think it has a location tracking for me,” (User 15), most users understand that “it probably does have ways to find that out” (User 17). However, many feel that location is not useful in determining their recommendations:

> I do understand that it would know where I am, because I mean all smart phones know where you are at all times. But I still don’t really know how that would affect my recommendations … I don’t really listen – I don’t change my listening habits based on where I am” (User 6).

Temporal data (e.g., time of day, season of the year) can also be correlated to what a user is doing or to other factors to influence recommendations. A common observation on the influence of temporal data was about Christmas: “you know in December you’re going to be getting a lot of Christmas records” (User 4). User 12 acknowledges that they listen to different music different times during the day (e.g., while at work during the day, at home in the evening). However, User 5 is skeptical about how this affects recommendations: “I couldn’t think of a good way for Spotify to both get that information and to make use of it in terms of the recommendation.”

The interview discussion of temporal influences caused User 13 to revise her beliefs about these data signals:

> It was like jarring to me to think about the fact, Oh my God. They're thinking about when I access the app and coming up with particular reasons why on those days, or during those times, if it's something to help me fall asleep, if it's something to help me wake up, if it's because I just needed a pick-me-up at this particular time of day, whatever it may be. It was just a little bit shocking to me.

As a result of this User 13 wanted to adjust their survey rating on the importance of temporal data signals to recommendations from "not important" to "quite important."

7.3.4.5 Data Collection Techniques

Users expressed beliefs about a variety of data collection techniques and how these data influenced recommendations. Prominent among these were audio ID, external media, the influence of other listeners, age, and Spotify's Taste Profile for each user.

The idea of an "acoustic ID" or "an audio fingerprint" (User 16) was identified by only a few users as a technique to identify "sonically similar" (User 14) recommendations. While User 17 believes Spotify has "some kind of formula or something that translates an intangible wavelength into an algorithm", User 15 was more descriptive and specific:

> Spotify bought one of those companies that does the music DNA mapping where they say we're going to run a sonogram analysis algorithm and say, oh, here are the properties or the characteristics of this audio wave and we will say that these are the qualities of that audio wave, and then we'll look at the audio wave of song B and say, oh, here are qualities that are shared by it.

The application for this "DNA mapping" was not just to determine similarity but also to preclude "jarring key changes" in the playlists "so that it's not super dissonant" (User 7). User 16 is confident that because of such mapping the algorithm can "measure much more scientifically" the genre of a song.

The survey responses illustrated the low rating of external media (e.g., social media, journalism) in influencing recommendations, although "Reviews of music in magazines, blogs, videos, news sources" is the distinguishing statement in Factor 6. This trend continued in the interviews. Most users don't believe external media are "an incredibly strongly influential connection" (User 5). Users 6, 8, and 13 acknowledge the use of external media for reviews and recommendations, but believe they don't "hold more weight" than what they are listening to or doing on the app. For User 6, "I always thought it [Spotify] was more self-contained."

However, critical reviews of music in magazines and other media have value. User 11 believes that "critical reviews" are one of the two important "buckets" of information responsible for recommendations ("listener response" is the other). User 2 thinks external media might provide important contextual information. They listened frequently to an artist who was subsequently charged with a series of sexual assaults. They stopped getting recommendations after this became public and wondered if the recommendation process is "actually that socially aware or if their algorithms take into account things like this." User 9 views the quality and quantity of critical reviews as a deficiency. They want more input from music critics: "a more catholic sourcing would be good, I think ... I think that they do use some kind of that, of music journalism source. And I would like to think that they would use more."

The importance of the actions of other listeners (e.g., what others listen to, playlists they create) in influencing recommendations is repeated and reinforced in the interviews. While Spotify does let users "follow" other users (and in turn receive recommendations based on the interests of these users), some users want a more personal relationship with other users for the purposes of getting better recommendations:

> I don't know where all those other people are. I don't know, there's no community. There are no clusters of communities where I can say, "Oh I know some of those people or I've heard some of those people and they're all talking to each other musically," that would be nice. (User 10).

The importance of "others" rarely includes friends. Some think the system is aware of who their friends are ("I don't know whether or not Spotify has access to my contacts [i.e., my friends]. Would I be surprised if they do? No," User 13) while others doubt it ("I don't think it knows who my friends are", User 19).

Users believe that certain data signals have a greater weight or influence on recommendations. Adding a song to a playlist "holds more weight" than 'liking' it (User 13) and skipping a song (i.e., a short dwell time) is less influential than clicking the "I don't like this song" button (User 17). External media (user posts and music reviews) are less important than what they, the user, listens to (Users 8 and 13).

Users believe weighting occurs within the algorithms: "there are product managers who are looking at the algorithm and know how it's done and are changing the weightings" (User 15). This weighting "amplifies" (User 19) certain songs or artists and "newer material is prioritized" (User 8). Many users identified a "weighting towards recency" (User 15). User 14 sees human hands in adjusting the weighting of data elements to shape recommendations in specific directions but do so in a way that is "algorithmically organic". This hand coding is viewed as opaque to the user even if it materially effects the recommendations. It preserves the perception of algorithmic responsibility and obscures the possible manipulation by Spotify for promotional reasons (e.g., financial agreements to highlight certain artists or songs).

User 15 notes that "your age is a good predictor of what people like" and User 20 concurs suggesting that "for Spotify not to consider age at all would be silly. Because certainly there are people that are nostalgic for their high school days and their middle school days." However, User 10 sees Spotify's use of age, which is collected when you register for Spotify, as a barrier to diversity: "There's an age thing. It's very ageist." Age is disproportionally determining recommendations, locking users into a specific era of music.

Spotify maintains a comprehensive Taste Profile on all its users (Whitman, 2012) with “a nuanced understanding of each portion of your taste” (Heath, 2015). This includes preferred genres, listening habits, demographic details, and an array of other data. However, only a limited portion of the profile is made available to users.

The high satisfaction rate suggests that Spotify’s Taste Profile is reasonably accurate, and users seem to agree: “How well does the Taste Profile know you? Decently, I would say ... I think it’s pretty good, at most people” (User 11); “I actually do think mine is pretty accurate. And I determine that accuracy by how much I like the recommendations it gives me in the weekly playlists” (User 8). For some, this was a bit of a surprise: “I did at one point feel like ‘Wow, Spotify knows me’” (User 16). Others acknowledge the profile but question its insights: “some of the stuff I like is a little bit arty I guess. So I know what I like about it, but I’m sure that Spotify doesn’t know what I like about it” (User 19).

However, many would like to “see behind the curtain and see what they have me pinned as” (User 11). One concern is simply transparency, “the right to know” (User 12) and another is reciprocity, “if someone’s going to have all this information, I don’t love it. But if they’re going to, I want to benefit and see it too. I want to know” (User 2). User 15 is concerned that he “cannot review and edit and challenge and delete” the profile so they can correct “a mistake” or tell Spotify to “ignore this signal.”

7.3.5 Concepts and Processes

Users held beliefs about various processes and concepts that influence the recommendations they received. Key among these were agency, privacy and surveillance, categorization and similarity, anthropomorphizing, and stakeholders.

7.3.5.1 Agency

Agency involves questions of power and control. Recommender system are a partnership of users and a system, and as a result, issues of agency are especially relevant. Many users are happy “letting Spotify take the wheel,” putting it on “cruise control” (User 5) and “let it guide me where it goes” (User 16). Despite this they still have “the degree of autonomy that I want in exercising my taste” (User 15). Others feel constrained and want to exercise greater control either adversarially by “gaming” the system (User 10) or by greater knowledge about how Spotify works. In the latter case, User 13 recognizes that the partnership is “not a completely symbiotic relationship and we’re not exactly fully cooperating here because Spotify is still doing a lot that we don’t necessarily know.” However, knowing more about how Spotify works is anticipated to “bring me some comfort in thinking that I have a level of agency involved.”

7.3.5.2 Privacy and Surveillance

Since recommender systems require disclosure of personal information, user tracking, and ongoing feedback, there are concerns about privacy and surveillance. However, “there’s a surrender of personal information that it needs in order to make recommendations that you want. I think that’s part of the deal. And that’s a world that I’ve accepted” (User 20). User 7 calls this “a balance” and “an exchange” of personal information and music access and User 19 concurs: “I had to make a judgement call on whether I wanted their service or whether I wanted my privacy and for the moment the service won over the privacy.”

For many users the decision to provide personal information and activity was not problematic. They viewed the data gathered and the implications of its use as innocuous: “the worst thing that they can do is look for patterns in my musical habits. And that’s data I’m willing to give up” (User 11). The privacy and surveillance concerns were minimal because “it’s just my music-listening habits” (User 17), “my music associations” (User 10), and “it’s music at the end of the day. It’s different. It’s not going to break the world” (User 14).

Other users have significant concerns about the “deal.” User 13 is concerned “that my music preferences say so much about me” and User 15, invoking consent, notes “it’s none of their business unless I’ve told them that I want them to monitor that stuff.” While User 2 is concerned about privacy they are also insistent that this extent of personal data directly and visibly benefits them:

> I don’t like that they know me, I don’t like that they’re collecting data, I don’t like that they also make assumptions about me that are incorrect. I don’t like that they know so much about me. I would either prefer them not know anything, but if they’re going to assume they know all this stuff, if they’re going to know me and provide catered services, at least let them be right ... if someone’s going to have all this information, I don’t love it. But if they’re going to – I want to benefit and see it too. I want to know.

Concerns about privacy caused users to explore options to moderate the impact. User 13 wants access to “one giant list and it simply just had a switch beside each one that just said, ‘Yes, you are allowed to access this information. No. you are not allowed to access this information.’” User 16 imagines a similar opt-in/opt-out process: “if I had the chance to opt out of certain forms of data collection, I’d definitely want to get to know what that is.” Increased knowledge about how recommendations work is linked to privacy protection and the mitigation of surveillance:

> I would definitely have more ease of mind in knowing how it worked and for multiple reasons. Because then, at least I have some level of understanding of, now I know that I’m definitely being surveilled. But now I also understand in what ways I’m being surveilled as well. And maybe there’s an element of agency involved in that where I can make choices about what information they gather and what they don’t” (User 13).

Many online platforms like Spotify have been identified with "surveillance capitalism" and their priority to capture user behaviour for profit (Zuboff, 2018). However, users in this study are clear, Spotify is "not Facebook" (User 14); "I'm a lot more concerned about Facebook than I am about Spotify" (User 20). In part this is because "it's music at the end of the day. It's different. It's not going to break the world" (User 14). Spotify "does feel different, it does feel like I have more control" (User 19). Where there was concern about Spotify, it was specifically about low artist compensation (Dredge, 2020; Ferraro et al., 2021; Jacob, 2021). In reference to this User 4 noted "an immediate backlash against it [Spotify]" and that the company "does has a sort of an odd corporate menace to it."

This mostly ambivalent attitude towards Spotify was echoed in user comments about trust. While the high satisfaction rate (79%) and listener longevity (42% had been listeners for over 5 years) imply trust in the system, few users mentioned trust in the interviews. When trust was raised it was qualified: "I wouldn't quite say trust. I think it's less trust and it's more ... a willingness to lose" (User 11). This acceptance of poor or unsatisfactory recommendations ("to lose") suggests tolerance if not trust. Given that users are defined by data, somewhat ironically, User 2 doesn't "trust them as a data company" but trusts them to "know what you like."

7.3.5.3 Categorization and Similarity

Most users believe Spotify categorizes music, and to a lesser extent users, using "a range of tools and categories and classifications" (User 15). Music is "categorized" and "catalogued" (User 12) in a "database" (User 5). Users mentioned a range of elements captured in this process: mood, tone, genre, era, male and female voices, length of song, and Acoustic ID. This "metadata" is described in "tags," "descriptors" and "labels." Spotify applies "values or properties" (User 15). Relationships between and among artists are captured. This includes biographical information such as an artist's "spin off projects"

(User 2), the “opening acts” when bands tour (User 6), and if various artists “created music together” (User 8). Categorization of listeners was less detailed. Users are “clumped” (User 6) and “bucketed” into types of music listeners (User 14). All the metadata “get jumbled up in some algorithm somewhere” (User 12).

Users described various sources for this metadata: Spotify, artists, producers, algorithms, music labels, music journalists, and social media. User 9 feels “there’s a bit of laziness about whether they get that information or not, on how accurate it might be.” With respect to song credits, “that information is so terrible so often. It’s so many songs just appear not to have been written by anybody.” User 7 is also concerned about metadata quality: “I’m hoping that they get it from reputable sources, like provided by the artists themselves ideally.” Because of these issues, User 6 wants to be able to “edit the metadata” to add relevant information and correct errors as you could on iTunes.

The quality and effectiveness of music categorization was raised particularly with respect to musical genres. Music is “messy ... it doesn't fit neatly into categories” (User 16). An algorithm “can’t work in these sort of boundary-less genres” (User 3) but can “when artists work in clearly defined genres and clearly defined sounds” (User 7). User 17 believes genre categorizations begin with “anchor points” determined by music editors:

> It seems like they have their grounding artists for the genre whatever the genre may be, so these are the people who have defined or who exemplified the genre. So they become the anchor points and then they fill it in with other creators out there.

The known problem with music industry metadata (Deahl, 2019) is echoed by User 19 who doesn’t “trust it” since “the metadata on classical music is a mess.”

The notion of similarity and similar things or people are described in terms of “likeness” (User 10), “adjacency” (User 13), and “proximity” (User 10). Music that has the same “lyrical quality” (User 2), “structure” (User 8) or is “sonically similar” (User

14) is identified by a “sound algorithm” (User 14) and rated by a “similarity index” (User 15).

Two users (Users 5 and 13) visualized similarity in terms of a Venn diagram. The algorithms create “this really weird Venn diagram where sometimes the circles align almost the entire diameter of the circle and other times it’s just a fraction of the diameter and they align” (User 13). However, as User 2 notes, “it’s hard to describe what I think is similar, because it could mean different things.” User 18, who described the recommendation of a Chopin cello transcription as “not the same thing” and “wrong”, notes that “where AI fails us, when it does fail us – for all its usefulness – it’s because it doesn’t know that it’s not the same thing.”

User 10 isn’t interested in what Spotify’s definition of similarity is or how it works: “I don’t know what it is particularly, and I don’t spend a lot of time thinking about it and I don’t really care.” For them what matters is whether they recognize something as similar in what Spotify is presenting as similar: “is the similarity a similarity that works for me?” (User 10).

7.3.5.4 Anthropomorphizing

Users endow Spotify with human characteristics. User 14 notes that the app says “good morning, good night, good afternoon, whatever, and then it pops up with six recommendations for you.” Spotify “knows my music taste better than most people I know” and their recommendations are “gifts” for which the user is “thankful” (User 2). Spotify “knows me” (User 16) and is “someone [that] made me a mix tape” (User 14). Spotify “seems to lack imagination” when a user receives unsatisfactory recommendations. (User 4). User 8 uses Spotify as part of her part-time job as an exercise instructor. In this instance the relationship is not as a friend but as a colleague: “I would never have thought of it as a business partner but in the context of how I use it, it very much is.”

User 14 believes Spotify actively encourages anthropomorphization, although some users resist this notion. User 3 notes “I’m not in a relationship with this algorithm.” Other users “don’t feel very warm and fuzzy about [Spotify]” (User 10) and are not part of a “symbiotic relationship” with the system (User 13).

7.3.5.5 Stakeholders

Spotify is a multisided or multistakeholder platform. In addition to users, it must respond to the interests of a variety of groups (e.g., artists, music producers, music labels, and, for the free version, advertisers). When asked how Spotify prioritizes or balances these groups, users had different perspectives. User 17 believes Spotify prioritizes “the user ... [then] the licensing and record label company ... artists are lower down.” Financial reasons were important for many users: “whatever makes them the most money” (User 2) and “based on how much money you’re giving Spotify” (User 19). In this respect User 3 is clear: “Owners, first and foremost and forever.” The influence of other stakeholders, particularly large music labels, caused User 20 to question what shaped their recommendations the most, their actions or stakeholder needs: “my listening preferences may not play as large a role as I might think.”

Spotify is clear that “in some cases, commercial considerations may influence our recommendations” (Spotify, 2021c). This occurs by deliberately inserting songs or artists into recommendations or adjusting parameter weights to preference promoted songs in playlist rankings. Most users are aware of this (Users 3, 11, and 14) and only one user pejoratively invoked “payola” (User 15). The effect is, as User 18 notes, “when in doubt [Spotify will] give me the thing they’re being paid to promote.” User 16 hoped “if there was any backend deals or stuff. It’d be nice if that was made more transparent and explicit.” In fact, in 2017 Spotify explicitly identified sponsored songs and provided a setting to disable it (default was to enable). Currently there is no such identification and setting (Deahl, 2017).

7.3.6 Generalists vs Specialists

Following the shift of six users during the interviews from “solely by algorithms” to “primarily by algorithms and partly by humans”, Generalists and Specialists now have similar views on how recommendations are made (see Table 16).

Table 16: How Made by Category

	Solely by algorithms	Primarily by algorithms and partly by humans
Generalists	3	8
Specialists	2	5

While their views on how recommendations are made are similar, the objectives and goals of these groups are not. Specialists have narrow expectations: “I’m not expecting Spotify to introduce me to my new favourite artist” (User 3); “I’ve never really had the want or the desire to broaden my horizons” (User 12). Generalists have diverse expectations: User 15 has “a pretty capacious comfort zone so I tend to be pretty catholic in my musical tastes” and “why do recommendations if all you wanted to listen to is the same” (User 18).

Creating diverse and novel recommendations are key objectives for Spotify since these qualities are strongly linked to user engagement (Mehrotra et al., 2018; Mehrotra, Shah, et al., 2020). However, Generalists and Specialists have different views on what constitutes diversity and novelty (A. Anderson et al., 2020; Waller & Anderson, 2019). For User 13, a Specialist, it means recommendations that are only “a little bit more outside your comfort zone” while User 5, a Generalist, is looking for “people I have never searched for or are often people or groups I've never even heard of." For two Generalists, the lack of diverse recommendations raises concerns about being "siloed" (User 16) and of Spotify "not paying attention" to their preferences (User 4).

While prior research has shown that Generalists prefer search over recommendation (A. Anderson et al., 2020), this difference was not evident among participants. User 4 preferences search, indicating that they "never found anything that way [using recommended playlists]" but User 5 is more typical of Generalists in this study indicating that searching for an artist or something specific was "fairly rare."

Generalists feel Spotify "silos me" (User 16) and the "hit rate" of relevant and interesting songs is not high (User 11). Specialists find "they tend to stay within the genre groupings [that the listener prefers]. I don't think that the recommendations I've ever had strayed too far from those" (User 17). Specialists prefer being "siloed" while Generalists would like Spotify to "stray" more.

When asked how they might shape recommendations, Specialists and Generalists had very different strategies. Specialists would "remove the outliers from your listening habits" (User 12), "listen to particular artists or songs repeatedly" (User 13), "be selective about what I use Spotify to listen to" (User 17) and listen to a playlist of "only one band" (User 20). As User 6 notes regarding the recommendations they received, "the more you listen to the same type of music, the more specialized the music becomes." Generalists would listen to "a new genre of music extensively" (User 11) and do "anything that tells the algorithm to privilege one part of my data more than another in profiling me" (User 18). Two Generalists, both dissatisfied with Spotify, took a more adversarial approach. User 4 would "leap from genre to genre, listening to wide a variety of sounds therefore confounding any attempt to box my taste in." Similarly, User 10 would attempt to "game their algorithms." Specialists attempt to reconfirm their interests with various techniques while Generalists attempt to disrupt the algorithm seeking a different result from the recommendations.

User 18 has different life roles with different objectives for Spotify: parent, student, teacher, and listener. They want to "control which persona and even which facet of persona” is represented in the recommendations. Like the desire of User 10, also a Generalist, for “a bigger vocabulary”, this illustrates that Spotify’s “music retrieval from everything” (Jehan et al., 2010) may have a detailed characterization of genres (over 5000 according to (G. McDonald, 2021)) but a limited view of listeners and their characteristics and objectives. Generalists and Specialists have different objectives and expectations for the recommendations Spotify provides. These differences suggest the need for explanatory systems to acknowledge and respond appropriately.

7.3.6.1 Dissatisfied Generalists

Only four users are dissatisfied with Spotify’s recommendations, and they are all Generalists. These dissatisfied Generalists are mostly experienced users: Users 9 and 18 have used Spotify for over 5 years and User 10 for over a year. Only User 4 is a new listener with less than one year’s experience.

Important to Generalists are diverse recommendations. They want new music, and not just from artists or genres they have listened to previously. Generalists are looking for diversity, surprises, and challenges to their normal listening experience. The want “unpredictable recommendations” (User 10). From a Generalist perspective, “why do recommendations if all you wanted to listen to is the same” (User 18). However, prior research has indicated that current recommender systems, particularly collaboration-based systems, favour the objectives of Specialists and penalize those of Generalists (A. Anderson et al., 2020).

Spotify’s recommendations are described as “fairly predictable” (User 10), “outside what I need” (User 18), and “disappointing” (User 9). Reasons for this include a lack of diversity and novelty (“it follows a path that I could follow myself,” User 10) and a poor understanding of user expectations (“they’re not really dialed in”, User 9). The expectation of these Generalists is for surprises and the unexpected but relevant: “I expect leaps. So why doesn’t it?” As a result, Spotify isn’t “the slightest bit reliable for coming up with the odd bit of interesting" (User 9). After using Spotify for less than a year, and being frustrated with the lack of diverse recommendations, User 4 declared "they might have given up on me."

Only two users, both dissatisfied Generalists, mentioned randomization as a recommendation technique. Randomization is an important algorithmic technique to encourage exploration over exploitation (e.g., in multiarmed bandits). User 10 believes they can direct Spotify to be random ("I have it on random play") although what this user is referring to is actually "shuffle" play within a specific playlist. However, for this user it "doesn't seem to work that well" because "I never hear [a particular] song." User 4 believes Spotify makes random recommendations but "at some point it shouldn't be random. At some point they should get to the point where they know you." For this user, randomness is a response to the cold start problem. With more data signals from the user, randomness should no longer be required. This user doesn't regard randomness as an exploration and discovery technique.

Dissatisfied Generalists are concerned that Spotify is not using, or using effectively, the information they have provided. Spotify doesn't seem to be "paying attention to the broadness of what it is that I'm liking" (User 4) and "whatever it is that they're getting from it isn't useful" (User 9). Despite the diversity of data that User 18 provides to Spotify "they're confused [and], when in doubt, give me the thing they're being paid to promote". The lack of confidence in Spotify's use of information "doesn't make me think that giving them more is going to improve things that much" (User 9).

User 9 thinks Spotify should use "a more catholic sourcing" of information and "be informed by people who listen to music, even critics." For some the answer is not simply more information or information from critical sources but information that is richer in detail and more nuanced. For User 10 this is a challenge to Spotify: "give me a bigger vocabulary and then make it meaningful. Then prove to me that you've heard me." Prove it because "I feel like I do invest time, so how much time? You can suck all my time."

7.4 Conclusion

During the interviews, Spotify users confirmed, amplified, and in some cases adjusted the views and beliefs they expressed in the survey. With a group of users altering their beliefs, most users (68%) now believe Spotify's recommendations are made "primarily by algorithms and partly humans" in accordance with actual Spotify practice. The survey allowed users to elaborate on the respective roles and responsibilities of algorithms and humans. Algorithms had the primary responsibility for recommendations while other users or Spotify editors had mostly oversight roles. The human role was minimized.

The importance of active and explicit data signals was reinforced. However, beliefs about how users can shape or alter recommendations revealed both a lack of techniques to produce changes and an uncertainty about the efficacy of those techniques. Passive and implicit data signals, such as external media, other users, location tracking, and activity tracking, were viewed as having lesser or even no importance. While objectives in using Spotify matter (e.g., generalists are looking for diverse music while specialists want familiar music), the broader differences seen in prior research were not as evident in these interviews.

Most users either didn't notice the explanations that Spotify provides about recommendations, or they felt they were unimportant. However, users had many questions or concerns about algorithmic practices, data privacy, surveillance, tracking or inference of emotional states, and Spotify policies for which they expected answers or explanations.

Users have little understanding of the machine learning technology underlying Spotify's recommendations and generally have few concerns about this lack of knowledge. Many acquiesce to the system, allowing it to "take the wheel" and make decisions irrespective of user actions or interests. Connected to this is the anthropomorphizing of Spotify. Spotify is "my algorithm." It "knows me", says "good morning", and provides recommendations as "gifts."

Users have a variety of beliefs about how the similarity of music, artists, and other users affects recommendations. Music is “sonically similar” because of algorithmic audio analysis. Artists and users are categorized and classified with metadata to indicate similarity. The co-occurrence of user behaviour (e.g., “liking” the same song) is taken to signify similarity of musical taste.

Users have different beliefs about agency. Some are concerned about limited agency, particularly with respect to the techniques available to manage recommendations, while others are happy to put Spotify “on cruise control” and let the system make all the decisions. A few users actively resist Spotify by adopting a variety of adversarial strategies. Most see the relationship as a “balance” or a “bargain” where agency is shared and contested.

Users also hold beliefs about Spotify as a company. Spotify was “not Facebook.” Most users attributed few malicious motives to the company. However, users held views about Spotify’s artist compensation levels, priorities for the company, commercial imperatives, and lack of user control and customization.

8 Folk Theories of the Spotify Music Recommender System

8.1 Introduction

User beliefs elicited from the survey and interviews can be synthesized into a set of folk theories about how Spotify makes personalized recommendations. The following folk theories are presented as verbs from the user perspective. Each verb is accompanied with a short description.

Expressing folk theories as verbs acknowledges Dervin's concern that "we focus primarily on entities, not process; on nouns, not verbs" (Dervin, 1993, p. 51). Moving "from nouns to verbs; and from nounings to verbings" (Dervin, 1994, p. 377) … "mandates a focus on the hows of human individual and collective sense-making and sense-unmaking, on the varieties of internal and external cognizings, emotings, feelings, and communicatings that make, reinforce, challenge, resist, alter, and reinvent human worlds" (Dervin, 1999, p. 731).

Seven folk theories were identified: complies, dialogues, decides, surveils, withholds and conceals, empathizes, and exploits. It is important to remember that individual users will hold some but not all these folk theories and some may hold contradictory beliefs depending on the context.

8.2 Folk Theories

8.2.1 Spotify Complies

Spotify complies with my directions to create recommendations.

Spotify responds to me and my preferences.

In the survey, the highest rated data signals that users believed influenced recommendations were those that represented active and explicit actions by users. These actions were viewed as directives to which Spotify complies. For example, rated as "very important" by users are "What I listen to" (95%), "How many times I listen to a song, artist or playlist" (89%), and "Marking something a 'like' (i.e., 'heart')" (68%). Four of the nine consensus statements in the factor analysis reflect the specific actions of users: marking something, listening to something (how long and how many times), and creating playlists. Each of the four factors extracted emphasized that the recommendations were "about me" and actions taken on the system. Additionally, the statement "Songs that are similar to other songs I 'liked' or listened to" is one of the highest rated data signals that users believe influences recommendations and is a consensus statement in the factor analysis.

Where the data signals were implicit, inferred, or obtained from external sources, users indicated lesser or no influence on recommendations. This includes data signals about education level, location, activity, temporal data, and external data (e.g., from magazines or social media). The only exception is time of day.

Users perceive themselves as in control: "The system doesn't know what I'm doing. I know what I'm doing, and I choose based on knowing what I'm doing and then I tell it what I'm doing" (User 10) and "if you're not directly telling it then I don't know if it would fully know exactly what you're doing" (User 17). User 3 observed that "the only cues that it's getting are the ones that I'm feeding it." User 14 wanted those cues to include their emotional state: "I think it could be cool to give feeling inputs to the algo and have it deliver a mix based on those things." Users believe they have agency over the type and extent of information gathered by the system to make recommendations. User 15 noted "it's none of their business unless I've told them that I want them to monitor that stuff."

Training the algorithm is not reciprocal. A user "tells" the algorithm (User 18) about "my 'likes' and preferences" (User 19). Users believe they can trace the compliance of the system: "I can usually determine which parts of my listening history have influenced a recommendation" (User 18).

8.2.2 Spotify Dialogues

Spotify dialogues with me about music.

Together we find recommendations of old favourites and new music.

Users believe that Spotify engages them in a dialogue. The user-Spotify dialogue is expressed in a variety of ways. It is a form of reciprocity: "I'm feeding it, it feeds me" (User 19). It is a teacher-learner relationship: "I've trained my Spotify really well" (User 8) and "Spotify only works because they [users] are teaching it to work" (User 19). It is a "feedback loop" (Users 16 & 19). User 10 invokes a process that underscores the importance of experience and comfort in the dialogue by comparing it to buying and wearing a new pair of pants: "I make them my own by using them."

There are other ways Spotify encourages and receives “messages:” “Recommendations can be made solely by the title of the playlist that you’re providing ... and so that’s what I meant by messaging” (User 13). Naming playlists is one of the few ways Spotify allows users to employ techniques like folksonomy and tagging to express personal views or categorizations. The messaging can be “anything that tells the algorithm to privilege one part of my data more than another” (User 18). User 20 takes specific actions (e.g., ‘like’ a song or artist) “because I know it’ll change the algorithms.”

When recommendations were not satisfactory, users often identified a breakdown in the dialogue. User 15 noted that “I’m probably not training the machine learning algorithms that Spotify is using in an optimal way” because of the various actions this user takes (e.g., “liking” a song, then adding it to a playlist, and later deleting it from that playlist and adding it to another playlist). User 10 believes Spotify “requires probably more about me” and User 16 questioned whether they are effectively communicating: “I don’t know if I’ve intentionally tried to train the machine.” User 10 wanted a different type of dialogue: “if Spotify sent me a questionnaire today that asked me more personal questions about my lifestyle, there’s more data points to pull together, then that might help.” They wanted Spotify to recognize that “we’re not all the same individual units.”

Agency is shared, and sometimes contested, but the roles are different. Humans (users and Spotify editors) “curate” while the algorithms “create” (User 5). Curation is a creative, intellectual activity while creation, in this context, is mechanistic. Users believe they “have more control” than they do on Facebook (User 19). They can quantify that control: the recommendations are “25% me and 75% Spotify” (User 20).

8.2.3 Spotify Decides

Spotify decides for me what music I am recommended using black box processes.

Most users are unaware of how Spotify works, are largely unconcerned about this, and are happy to let the system determine the recommendations. Whether users believe the recommendations were made “solely by algorithms” or “primarily by algorithms and partly by humans,” users believe the system operates independently from them according to its own objectives and intensions. User 13 summarizes the belief: “it’s all this giant black box, I don’t know anything and there’s nothing I can do about it either.” As a result, Spotify “pushes or deters” recommendations (User 12) according to its own objectives not those of the user.

Users were open about their lack of knowledge of recommender systems and algorithmic decision-making. They were clear (“I don’t know”), uncertain (“assume,” “guess, “presume,” “imagine”) and conditional (“it might be,” “best guess”). For most users “it’s completely opaque to me how it all happens” (User 3). While not knowing how Spotify works “does concern me,” User 14 acknowledged “it doesn’t concern me enough to do the work and figure out the whys and the wherefores of it”. User 12 agrees, “if I really wanted to know, I could find out. I think that that’s how a lot of people feel” (User 12). As User 17 put it, “knowing the basics I think is enough for this platform.”

As a result, these users are happy to “put it on cruise control” and let the system “take the wheel” (User 5) allowing Spotify to “guide me where it goes” (User 16). This is, in part, because they believe not their habits but “other people's listening habits are really what drives it more than anything” (User 11). Users are “not so much into getting under the hood … we’re fine as long as it works. We’ve moved on from the guy in the garage, we don’t want to know how things are, we just want the car” (User 4). And the users are fine. As the survey revealed, 79% are satisfied with the recommendations they receive.

The acquiescence to the system is evident in user reactions when recommendations are poor or unsatisfactory. User 13 viewed this loss of recommendation accuracy as “a small perk that I had before, isn’t happening as frequently now.” User 11 suggested this ambivalence is “a willingness to lose” not necessarily a fault in the system.

This passive role of users is evident in how User 4 imagined how Spotify would describe itself: “this is who we are, take us, this is what we provide, this is what we do.”

Some users adopt a passive role while others resist the system. The effect of the system taking control is noted by User 16:

> I feel like lately my musical experience has been a little more passive and impersonal. In another phase of my life, I would have been more active and actually searched and looked through. But I think it’s less – I don’t know, it’s harder to do so when there’s so much. And I guess that’s the point of curation, and playlists, and recommending things.

Others expressed frustration. They wanted to “game their algorithm” (User 10) or take actions aimed at “confounding any attempt to box my taste in” (User 4). Another user would open a private session to hide their normal preferences from the algorithm (Users 15 & 19). User 13 admires a user who names their playlists after food remarking “what is the algorithm going to do with a title like onions and chives?” In attempts to influence the deciding algorithm some users “remove the outliers from your listening habits on Spotify” (User 12) and “be selective about what I use Spotify to listen to” (User 17). These users change their listening practices to contest the algorithm.

Users believe it is difficult to train the algorithm, so they are resigned to letting Spotify operate independently: “training the algorithm is a lot of effort” (User 3); “it takes a lot of time [to train the algorithm]” (User 10). User 15 doubted putting in the effort was “worth the exchange, to make better recommendations.” User 3 acknowledged the cost/benefit when the recommendations were good enough:

> Spending the time to do all of those actions that I laid out to get the absolutely pristine version of me telling it what I like through all the ways that I think that I can tell the black box what I like. I don’t need to do that because I’m getting out of it what I want.

User 9 was also skeptical: “that doesn’t make me think that giving them more is going to improve things that much”.

8.2.4 Spotify Surveils

Spotify surveils me by tracking everything I do

and making inferences from other data to make recommendations for me.

Users believe Spotify surveils them. User 3 outlined the extent of the data capture:

> It records everything that any user does, in any interaction. Every half-completed search, every click, every I spent 30 seconds on this page, I spent a minute on that page, whatever it is. They will have recorded absolutely every interaction any user has ever had with their interface.

Other users acknowledged the surveillance but are unclear how it occurs: "you are being surveilled without fully understanding how that is happening, or what is being done to surveil you" (User 13); "they're probably just collecting it all and filtering it out" (User 17). User 10 linked surveillance to system and user behaviours, noting that Spotify tracks "your behaviour immediately after being recommended a song. And I think compiles all that to determine how successful a recommendation is. And then if it is successful, it repeats that behaviour."

Most users accept this surveillance because "there's a surrender of personal information that it needs in order to make recommendations that you want. I think that's part of the deal. And that's a world that I've accepted" (User 20). It's "a bargain" (User 3) or "an exchange" (User 15): personal information for access to music. Users make an explicit decision: "I had to make a judgement call on whether I wanted their service or whether I wanted my privacy and for the moment the service won over the privacy" (User 19).

However, it is a bargain where privacy is viewed as inconsequential because it is "just music" (Users 7, 8, 12, and 16) and "just my music listening habits" (User 17). For User 11 "the worst thing that they can do is look for patterns in my musical habits. And that's data I'm willing to give up."

Surveillance is not just about user actions on the system. Nearly a third of users believe Spotify tracks their emotional states. "What I'm feeling when listening" was rated as a "very important" or "important" influence on recommendations by 32% of users. Users assume it tracks their location ("smart phones know where you are at all times" User 6) and as a result Spotify "definitely knows when I'm at work" (User 14). Surveillance of, and inferences from, these passive and implicit data signals are acknowledged if devalued: "It is pretty incidental what they get. I just know that they get it" (User 3).

User 13, who believes recommendations are made solely by algorithms, feels "a comfort thing" that it is algorithms, not people, who surveil them because algorithms have "certain limitations" and can know only so much. From this perspective, algorithmic surveillance preserves a level of privacy. Despite the level of surveillance, User 3 thinks Spotify overestimates its efficacy: "I think it thinks that it knows better than it does."

8.2.5 Spotify Withholds and Conceals

Spotify withholds and conceals information

about how recommendations are made and how I can influence them.

User 13 recognized that users and Spotify are "not exactly fully cooperating here because Spotify is still doing a lot that we don't necessarily know." Users believe Spotify withholds, conceals, and obscures information from them. This includes how users can shape recommendations, how effective those techniques are, what information Spotify has about users, and how Spotify promotes certain music and artists over others.

Users could describe few strategies to purposefully shape recommendations. They also expressed uncertainty about the effectiveness of the limited techniques available to do this. Using the "don't play this" option (e.g., disliking something with the "Don't like this song" or "Don't like this artist" indicator) was noted by User 3 as one such strategy, although User 20 cautioned that it "doesn't seem to influence algorithms too much." Similarly, User 2 didn't think removing a playlist as a strategy to shape recommendations "would really delete the information from Spotify." As User 20 noted, "Spotify is really good at adding them, but not so keen on removing songs." User 10 challenged Spotify to "give me a bigger vocabulary and then make it meaningful. Then prove to me that you've heard me" (User 10).

Many would like to "see behind the curtain and see what they have me pinned as" (User 11). One concern is simply transparency, "the right to know" (User 12) and another is reciprocity, "if someone's going to have all this information, I don't love it. But if they're going to, I want to benefit and see it too. I want to know" (User 2). User 13 wants "one giant list" of all the information Spotify has and a "switch beside each one" that authorizes Spotify to use it or not. User 15 is concerned that he "cannot review and edit and challenge and delete" his profile so they can correct "a mistake" or tell Spotify to "ignore this signal." This leads some to believe that Spotify "silos me into a particular style" or an "acoustic echo chamber" that doesn't facilitate "musical exploration" (User 16). This lead User 2 to wonder what is "missing" from their recommendations and why. When they didn't get recommendations to one of their favourite bands, they reacted skeptically "if they know me, they should know this."

Spotify, through its "algotorial" processes, manually alters weightings, inclusions in playlists, and promotions of specific artists or music without acknowledging this. While no longer the case, previously Spotify's apps signaled this to inform users. However, most users are aware this happens (Users 3, 11, and 14) as evidenced by the statement "Songs or artists that Spotify is promoting" being rated as a "very important" influence on the recommendations they receive by 47% of the users. One user invoked "payola" to describe these promotional processes (User 15) and another asked Spotify to be "more transparent and explicit" about "backend deals" (User 16). As a result, as User 18 noted, "when in doubt [Spotify will] give me the thing they're being paid to promote."

These interventions and promotions lead to certain biases in Spotify's recommendations. Users noted an "emphasis on recent listening" (User 14), "recordings deemed to be popular and/or profitable to Spotify" (User 9), and "newer material "(User 8). User 10 sees Spotify's use of age, gathered during initial registration on the system, as a barrier to diversity by emphasizing music from a specific era: "There's an age thing. It's very ageist."

However, despite concerns about Spotify withholding and concealing information, User 3 expressed the resignation of many when at the end of their interview they remarked:

> I'm going to spend the next little while just poking around at Spotify. See what else I can see that has been invisible to me. And then also probably put some of this aside and just go forth listening to my things that I listen to.

8.2.6 Spotify Empathizes

Spotify empathizes with me.

Users believe Spotify is a friend who understands them and empathizes with them. They describe the service as "my Spotify" (User 18) and "my algorithm" (User 8). It says, "good morning, good night, good afternoon, whatever, and then it pops up with six recommendations for you" (User 14). Spotify "knows my music taste better than most people I know" and their recommendations are "gifts" for which the user is "thankful"

(User 2). Spotify "knows me" (User 16) and is "someone [that] made me a mix tape" (User 14). User 8, a part-time exercise instructor who creates playlists for their work, described Spotify as "a business partner." When disappointing recommendations occur, Spotify "seems to lack imagination" (User 4).

User 3 believes that Spotify tries "very hard to know what you're feeling and use that as something to serve up music." The importance of feelings or emotional states as a data signal for recommendations was viewed as "very important" or "important" by 32% of the participants. The distinguishing statement in Factor 1 was about feelings and recommendations. Asked if Spotify, the algorithm, infers their feelings from what they are listening to, User 14 responded "Yeah, I think so." Spotify empathizes by matching a user's emotional state with relevant music ("I really do like it when it matches," User 2) or by attempting to adjust a user's mood ("if you start with a very depressing playlist maybe it does try to correct you in a more positive direction," User 17).

Spotify is "not Facebook" (User 14) and "it does feel more private" (User 19). Concerns regarding privacy, misinformation, polarization, and bullying on social media sites such as Facebook are contrasted with beliefs about Spotify as a safe, private, protected, apolitical, and largely inconsequential platform (Users 2, 14, 15, and 19). Regarding the latter, users believe it's "just music" (Users 7, 8, 12, and 16) and "just my music listening habits" (User 17). While many users declined to say they trusted Spotify, many expressed confidence and comfort with the system (Users 3, 4, 11, 14, and 19). User 2 doesn't trust Spotify "as a data company" but they do as "a friend," a musical confidant, "who knows you, what you like, and when they say, 'Hey listen to this band or try this restaurant' you can trust their judgement that they do know what you like."

8.2.7 Spotify Exploits

Spotify exploits my labour for the benefit of the company.

Users believe Spotify exploits them: “my choices, my preferences, are being harvested for their algorithm ... [and this is] the product people are paying for” (User 15). User 19 sees this as “doing some of their work” for which users “don’t get compensation.” Despite this “work,” users believe they are further exploited by receiving recommendations weighted towards music that is “profitable to Spotify” rather than personalized for them (User 9). When “in doubt” Spotify will “give me the thing they’re being paid to promote” (User 18). Unlike users who want options and means to contribute more information to Spotify (largely to improve their recommendations), User 3 noted that “I’m already doing labour in terms of data creation for them. I’m not going to volunteer to do more of it if there’s no benefit back to me.” Users of Spotify believe they are both customers and workers.

8.3 Classification of Spotify Folk Theories

DeVito and colleagues have classified folk theories according to their nature (abstract or operational) and their complexity (functional or structural) (DeVito, 2021; DeVito et al., 2017). Analyzing folk theories according to these classifications suggests ways for XAI to understand and address them. It aligns with classification as a prototypical machine learning task.

Abstract theories, which accounted for most of the theories elicited in DeVito et al., “do not include specific attempts to theorize how an algorithm might actually operate.

Instead, they rely on a more general sense that an algorithm is something that will, in turn, cause something to happen" (DeVito et al., 2017, p. 3169). In contrast, operational theories "demonstrate a specific understanding that there are some criteria by which an algorithm must make curation decisions" (DeVito et al., 2017, p. 3169). Researchers found "abstract and operational folk theories reflecting two notions of algorithms: as an other or interloper, and the algorithm as a process requiring decision criteria, respectively" (DeVito et al., 2017, p. 3171).

DeVito classifies the complexity of folk theories held by users as a hierarchy that moves from basic awareness and recognition of causal powers (Functional Theories) to the identification of "mechanistic fragments" (e.g., factors) and finally to the aggregation into "mechanistic ordering" (Structural Theories) (DeVito, 2021). In functional theories users have a basic understanding and believe that algorithms "do something." In structural theories users identify specific factors causing systems to make decisions and, at the most complex level, identify sets of factors acting through rankings, weightings, or other influences to make decisions.

Abstract theories reflect a functional level of algorithmic complexity whereas operational theories reflect a structural level of algorithmic complexity. The distinction between abstract and functional theories, and operational and structural theories was identified in how users perceive recommender systems. Users viewed the components of the system either as "single, constituting entities" or as "relations" among these entities (Bieringer et al., 2021, p. 8). In the former, users deconstruct the recommender system, focusing on specific factors, effects, and personification (i.e., an abstract/functional perspective). In the latter, users focused on the relationships among the various components (i.e., an operational/structural perspective). Users who view recommender systems from both perspectives had a better understanding of the system, positively assessed the quality of the system performance, and were more likely to be satisfied with the recommendations (Kulesza et al., 2012; Yeomans et al., 2019). With respect the folk theories of the XAI tools or mechanisms (i.e., not the machine learning models but the XAI systems), Vaughan & Wallach note that "which stakeholders require structural versus functional mental models of intelligibility tools in which settings remains an open question” (Vaughan & Wallach, 2021, p. 131).

8.3.1 Comparisons

Table 17 compares the folk theories of prior research with those derived from Spotify users.

<table>
<tr><th colspan="4">Table 17: Comparisons of Folk Theories of Socio-Technical Systems</th></tr>
<tr><th>Spotify Folk Theories</th><th>Ytre-Arne & Moe (2021)</th><th>Siles et al. (2020)</th><th>French & Hancock (2017)</th></tr>
<tr><td>Complies</td><td rowspan="2">Practical</td><td rowspan="3">Privileged Social Intermediary</td><td rowspan="3">Relational Assistant; Transparent Platform</td></tr>
<tr><td>Dialogues</td></tr>
<tr><td>Empathizes</td><td></td></tr>
<tr><td>Decides</td><td rowspan="2">Intangible; Reductive; Confining</td><td rowspan="3">Feedback Control System</td><td rowspan="2">Corporate Black Box</td></tr>
<tr><td>Withholds/Conceals</td></tr>
<tr><td>Surveils</td><td rowspan="2">Exploitive</td><td rowspan="2">Unwanted Observer</td></tr>
<tr><td>Exploits</td><td></td></tr>
</table>

While there is not a one-to-one equivalency between individual folk theories, the comparison illustrates common themes across this research. The themes are evident in three clusters of beliefs. In the first cluster, these systems are perceived as effective tools (“assistants” and “practical”) that respond to the objectives of the user (“dialogues” and “complies”) acting as “intermediaries” through clear processes (“transparent”). It is a partnership folk theory where agency is shared, and user benefit is central. In the second cluster, these systems are “black boxes” where user agency is limited (the system “decides”) by the “control” of the system (“withholds and conceals”). The result is “confining.” It is a folk theory of system power and user subservience. In the third cluster, beliefs about system control become more concerning with the perception that the system “exploits” users. Surveillance is “unwanted” and beyond what a user believes is necessary for effective recommendations. It is a folk theory about the lack of consumer protection.

One way these folk theories are different is the way they are expressed. Siles et al. and French & Hancock use noun statements while Ytre-Arne & Moe use adjectives. The Spotify folk theories are expressed as verbs. “Verbing”, as Dervin advocates (Dervin, 1993), provides a process view of folk theories that aligns with the actionable objectives of HCXAI.

8.3.2 Sentiments

Table 18 examines the sentiments of the folk theories from prior research with those derived from Spotify users.

Table 18: Sentiments of Folk Theories

Sentiment	**Spotify Folk Theories**	**Ytre-Arne & Moe (2021)**	**Siles et al. (2020)**	**French & Hancock (2017)**
Positive	Complies; Dialogues; Empathizes	Practical	Privileged Social Intermediary	Relational Assistant; Transparent Platform
Negative	Exploits; Withholds/Conceals	Intangible; Reductive;		Corporate Black Box;

		Confining; Exploitive		Unwanted Observer
Positive & Negative	Surveils; Decides		Feedback Control System	

These folk theories reflect both positive and negative sentiments. However, those that are both positive and negative warrant further attention. It is not that these folk theories are neutral. In different contexts they are viewed as positive or negative. Just as “Feedback Control System” is viewed as both “controlling” and “responsive” in Siles et al., so too are “Surveils” and “Decides” viewed in two ways. While “Surveils” may seem only negative, and many Spotify users described it as such (“I don’t like it”; “none of their business”), other users understood surveillance as a necessary or beneficial part of the “bargain” to obtain effective recommendations. This tension or contradiction is referred to as “the intimacy of surveillance” (Ruckenstein & Granroth, 2020). Similarly, “Decides,” which indicates autonomous control by the algorithm, and viewed largely negatively by Spotify users, was also seen positively by users who wanted to let Spotify “take the wheel” and have it on “cruise control.” The opposing views of “Surveils” and “Decides” reflect different user beliefs about agency and consent.

8.3.3 Nature and Complexity

Table 19 associates each of the folk theories elicited from Spotify users with the relevant DeVito classifications.

Table 19: Classification of Spotify Folk Theories	
Spotify Folk Theories	**Classification of Folk Theories: Nature / Complexity**
Decides	Abstract / Functional
Empathizes	Abstract / Functional
Complies	Operational / Structural
Dialogues	Operational / Structural
Withholds and Conceals	Operational / Structural
Surveils	Operational / Structural
Exploits	Operational / Structural

The classification of the Spotify user folk theories according to their nature and complexity suggests that these characteristics should be considered in XAI systems. They are central to the objective to “meet the user where they are in terms of understanding and literacy, regardless of how contradictory, sparse, or fragmented these understandings may be” (DeVito, 2021, p. 339:4).

According to DeVito, user behaviours differ between those who hold functional theories and those who hold structural theories (DeVito, 2021). In attempting to understand and adapt to changes in an algorithmic system, both functional theorizers and structural theorizers engage in a “continuous interplay between adaptive sensemaking and information foraging” (DeVito, 2021, p. 339:28). Sensemaking involves using the system differently to achieve the desired result while foraging involves finding explanatory material to understand the nature of the system changes.

However, functional theorizers found the “complexity of the information” a barrier to understanding while structural theorizers were deterred by “the organization of the information, including how easy it is to find relevant, and avoid irrelevant, information” (DeVito, 2021, p. 339:25). Structural theorists were able to incorporate new information into their theories, including a focus on the “why” of something (i.e., sensemaking) while functional theorists “simply react to effects” (DeVito, 2021, p. 339:30).

Overall the barriers were “a lack of specificity for all, a lack of simplicity for functional theorizers, and an organizational challenge for structural theorizers” (DeVito, 2021, p. 339:32). Overcoming the barriers for functional theorizers requires repetition and elaboration, storytelling, and explanations that are compatible with user expertise. The key intervention for functional theorizers is algorithmic literacy with the goal to “boost users to structural theorization” (DeVito, 2021, p. 339:30). For structural theorizers, DeVito recommends “a platform-built exogenous foraging aid” that would make relevant information more accessible (DeVito, 2021, p. 339:32).

DeVito does not believe the move toward structural theorization is onerous. Research suggests that just being aware of some structural information is sufficient to induce a more complex algorithmic theory in users: “the threshold appears to be knowledge that multiple mechanistic fragments are part of the algorithmic decision-making process in question. In fact, it may not matter if knowledge of specific fragments is retained for long” (DeVito, 2021, p. 339:30).

Unlike the findings from DeVito et al., the nature and complexity of folk theories of the Spotify users in this study are predominately operational / structural. While Spotify users lack a detailed knowledge of machine learning, they do understand the relationships between and among data signals and the general processes of co-occurrence and matching, reflecting an awareness of “mechanistic fragments” and “mechanistic ordering.”

8.4 Are Explanations Important to Spotify Users?

Enhancing HCXAI using folk theories as guidance is the central focus of this study. However, the observations of Spotify users seem to identify only a tangential need for explanations. Are explanations relevant to recommender systems like Spotify and are explanations important to Spotify users?

Listening to music is an example of “passive entertainment“ categorized as “casual leisure” (Stebbins, 2009). Unlike serious leisure (Hartel et al., 2016), casual leisure is “an immediately, intrinsically rewarding, relatively short-lived pleasurable core activity, requiring little or no special training to enjoy it” (Stebbins, 2009, p. 622). Since XAI, like all explanations, is context dependent, some have argued that XAI cannot developed outside a work or task context where there are clear objectives upon which to base explanations (Cabour et al., 2021; Mueller et al., 2021). While listening to music is neither work nor a task, and the objectives are more difficult to identify and quantify, leisure pursuits are increasingly enabled by algorithmic decision-making that should be accountable to users.

Lim et al. provide a classification of XAI questions users might pose and the nature of the explanations sought (Lim et al., 2009). Most of Spotify’s explanations are of the “why” type: “Similar to [artists you’ve listened to],” “Because you listened to [a specific artist],” and “Trending now”. In the sole example of a “how to” explanation, Spotify subtitles the “Made for You” playlists with “get better recommendations the more you listen.” There are no “what if” explanations and no opportunity to ask for “why not” explanations.

Despite the limited and concise explanations provided by Spotify, system designers are aware of the importance of explanations to satisfaction and that different users require different explanations. The BART algorithm at the heart of Spotify’s recommendations (particularly with regard to exploit or explore decisions) attempts to "jointly optimize both item selection and explanation selection" (McInerney et al., 2018, p. 31).

However, for many in this study "it's just music" (User 16) and "just my music-listening habits" (User 17). The impact of the recommendations, satisfactory or not, in such a leisure pursuit is not especially consequential and requiring an explanation. Most don't recall seeing the explanations that Spotify provides ("I don't remember seeing explanations" User 15; "I've just never noticed" User 2). Where the explanations are noticed, the reactions are tepid at best: "It's not terrible, but often it's not very interesting" (User 9) and "it's nice that they do have that I suppose, but I don't know how much that influences my usage of it" (User 5). The explanations are used "a tiny bit" (User 11).

It is not just the textual annotations that provide explanations. The image that accompanies a recommended playlist is generated by Spotify and is usually a visual compilation of some of the artists represented in the playlist. For some this image is explanatory ("so that is an explanation I guess," User 2). For User 14 the image is "a reason for why it's being recommended" and is more important than the textual explanations: "it's up to the images to help me decide."

Where the explanations seem most helpful is in validating the accuracy of the algorithm's recommendations. User 12 likes to understand "why the person [or the algorithm] thinks that I would like this" and "if they are right". For User 4 "it's a bit of a challenge to see whether it actually is [i.e., "made for you"]. However, users were not interested in more detailed explanations: "No, I don't think it would [help]" (User 12), "they seem fairly sufficient" (User 4), and "I don't think it matters to me" (User 17). User 3 was concerned that a more detailed explanation would be intrusive: "I don't really care and I think it would be even creepier if it was more accurate ... I don't want my technology to be always telling me how accurate it is" (User 3).

Given these largely indifferent attitudes, do users care about explanations in a non-consequential context such as being recommended music? When asked directly about explanations, the answer from users is generally no. However, when asked about specific issues regarding Spotify's recommendation process, users raise many questions for which they want and, from a consumer perspective deserve, answers. For example, User 14 said "simple is fine" regarding explanations, but when reminded of all the data collected by Spotify and how Spotify may or may not use it, they responded differently: "I want a credit report style breakdown of everything, rather than just like a snippy thing where they want to make you feel comfortable why they're giving you this."

During the interviews users expressed concerns and had questions about whether Spotify sells their data, preferences certain artists, and is influenced by large music labels. They wanted to know where the metadata comes from, how their privacy is being protected or compromised, and whether Spotify is surveilling them. Users don't make a distinction between the data and the algorithm or machine learning model (the focus of much XAI) and the service using the data and the algorithm or model. For them, the data and the algorithm are intertwined in a more holistic view. Separating out and isolating explanations about the algorithms misses the service context in which the algorithms operate (DeVito, 2021). Users interact with the algorithm by using the service.

User 4 noted that while it concerns them that they don't know how the recommendations work "it doesn't concern me enough to do the work and figure out the whys and the wherefores of it". User 4 assumes that users, rather than the system, must "do the work" in order to reveal the "whys and wherefores." The tenets of HCXAI would suggest the opposite with the onus on the system to enable users "to understand, appropriately trust, and effectively manage" AI systems (DARPA, 2016). The folk theories of Spotify users indicate beliefs that implicate public policy concepts such as disclosure and consumer protection. The principles of HCXAI underscore the importance of these higher-level issues and regulatory concerns to XAI.

9 Folk Theories and the HCXAI Principles

9.1 Introduction

The elicited folk theories from Spotify users provide a unique view into how users of machine learning-based recommender systems believe they work. These beliefs describe not only how a user understands the system but how they must engage and interact with it. While some of the individual folk theories align with and support each other, taken collectively they do not form a unified whole. They do not aggregate to a singular theory of recommender systems. Rather these folk theories are separate vectors that sometimes align and at other times remain distinct. They contain contradictions and commonalities. However, as a window into the complex user beliefs that inform their interactions with Spotify, they offer insights into how HCXAI systems can more effectively provide machine learning explainability to the non-expert, lay public. Unlike previous research where folk theories are presented in descriptive terms (nouns and adjectives) (French & Hancock, 2017; Siles et al., 2020; Ytre-Arne & Moe, 2021), the Spotify folk theories in this study are active statements (verbs) aligning with the actionable viewpoint of the HCXAI principles.

The HCXAI principles were distilled from a survey of the XAI literature selecting "measures, architecture features, and explanation types that have been shown to be effective in some contexts, but they are likely to fail in others … [resulting in] the most central principles that should be observed when developing human-centered XAI" (Mueller et al., 2021). While termed "principles," the document more accurately serves as a set of "broad guidelines" (D. Wang et al., 2019) upon which HCXAI developers can base specific implementations.

In an important limitation, these principles were distilled from XAI research that included few user studies or direct user involvement. Using folk theories to inform HCXAI is a method to insert the user voice into these discussions.

XAI has been criticized for focusing narrowly on machine learning models and ignoring or minimizing the larger socio-technical context in which machine learning systems reside. The HCXAI principles move towards this more expansive view, particularly in their emphasis on building "explanatory systems, not explanations."

The folk theories elicited from Spotify users reinforce, challenge, and can augment the HCXAI principles. In many cases, the folk theories reinforce the principles. They affirm that explanations or explanatory systems should be guided by or be cognizant of user beliefs about machine learning systems. Understanding user expectations (i.e., their folk theories) have been shown to shape explanatory strategies (Riveiro & Thill, 2021). Challenges to the principles arise in areas where the folk theories provide a different perspective on XAI. In these cases, the principles can be adjusted to address these concerns. Finally, the folk theories raise issues and concerns that are not addressed in the principles and warrant inclusion. In many of these cases, the folk theories open wider areas of concern that draw on the socio-technical context of XAI (D. Singh et al., 2022).

9.2 Where Folk Theories Reinforce HCXAI Principles

9.2.1 Common Ground

HCXAI seeks to bridge understanding by seeking "common ground" between the machine learning system and the user. The "Dialogues" and "Empathizes" folk theories illustrate the relationship users have with the system. This exchange and interaction reflect both the openness to and the mechanism for finding common ground. Because “explanation is never a ‘one-off’”, the principles promote a “long-term interaction with users.” Spotify users who believe the system “Dialogues” and “Empathizes” with them are invested in a long-term relationship and would align with an explanatory system that remembers them, learns from them, and offers explanations when things change but not when the conditions remain the same. Spotify has implemented Reinforcement Learning models to build and utilize long-term user models of their interests and preferences (Stål, 2021). Emphasizing that Spotify and its users are in a dialogue would benefit the creation and sustaining of a user model and provide the “common ground” for explainability.

9.2.2 Self-Explanation

The principle of “active self-explanation” shifts the balance of power and agency toward the user. By giving the user more information and context, they are empowered to make their own assessments and explanations rather than only receiving an algorithmic explanation. This recognizes that folk theories go through an “exploration and elaborative phase” (Villareale & Zhu, 2021) where users are questioning and interrogating the system. The “supplementary data” and “situational data” (D. Wang et al., 2019) that enables self-explanation aligns with user desires for a “bigger vocabulary” with which to engage Spotify. Facilitating self-explanation responds to the “Withholds and Conceals” folk theory where users have questions and concerns but don’t trust Spotify as a “data company” to be fully forthright in providing explanations or answers.

9.2.3 Knowledge Transformation and Sensemaking

HCXAI makes “knowledge transformation and sensemaking” a central objective and prioritizes the need to “identify the user’s current understanding”. The “quantitative assessment” of a user’s understanding (Gentile et al., 2021) can be more onerous than a qualitative one. Knowing that users have specific and different theories about algorithmic agency (e.g., “Complies,” “Decides,” and “Dialogues,” i.e., “tell,” “receive,” and “discuss” respectively) communicates key user beliefs about the system. HCXAI can then “meet the user where they are” and tailor the explanations to those subjective perceptions. This principle implies a pedagogical role for HCXAI. Although the principles encourage the provision of learning tools (e.g., tutorials), they are less clear about a pedagogical strategy for HCXAI. This omission is discussed later.

9.2.4 Design for Failure

The principles note that trust is damaged when “AI fails in a way that a human would never fail.” Users are concerned that Spotify has “all that information” and “should know me” yet fails to provide what the user believes are obvious recommendations. This is experienced by users as a breakdown in the “Dialogues” folk theory where the system is no longer “listening” or they are no longer “training” the system effectively. Users can blame the system or themselves. The “Withholds and Conceals” folk theory attributes failure to a deliberate attempt by Spotify to “push or deter” songs or artists which would favour the interests of the company over the user. The impact of the “Withholds and Conceals” theory and the breakdown of the “Dialogues” theory reinforces the need for “design failure” (Villareale & Zhu, 2021) (i.e., designing HCXAI that acknowledges user or system failures) and the importance of the “self-explain” HCXAI principle as a means to encourage users to interrogate their own folk theories.

9.2.5 Explanations Are Not Always Necessary

The principle that explanations are “not always necessary” is clear from Spotify users who didn’t notice or didn’t care about the explanations the system provides. This is consistent with the “Decides” folk theory where users relinquish agency to the system (put it on “cruise control”) and don’t want or expect an explanation. Spotify’s explanations for specific user recommendations are concise and innocuous. Spotify’s overall explanation of the recommendation process is provided in a relatively obscure dropdown menu that was only added to the system in 2021. In both cases explanations are unobtrusive for those who don’t care or don’t care in the moment. They are examples of “hidden design” features for HCXAI (Ngo & Krämer, 2021).

9.2.6 Explanations as Verifications

The HCXAI principles acknowledge that an explanation can have “different consequences” and address different needs. The principles emphasize that different methods required to respond to those needs. For some Spotify users explanations (e.g., “Because you liked ….”) were a way to validate the accuracy of the recommendations (e.g., “to see if they were right”). The need for an explanation is not about “why” or “how” but rather for a set of confidence or performance metrics. The folk theories “Decides” and “Withholds and Conceals” reflect the concern about agency and the need for verification. The explanations provided by Spotify do not adequately address these sorts of questions and this is often true with other consumer recommender systems. Spotify users were interested in the performance of the recommender model not just for them but across the user community.

9.3 Where Folk Theories Challenge HCXAI Principles

9.3.1 Triggered Explanations

The HCXAI principles indicate that the need for an explanation is “triggered” by “surprise and violations.” However, the folk theories such as “Surveils”, and “Exploits” have both positive and negative connotations for users suggesting that the need for an explanation is less a “trigger” than an issue of a threshold or level of intensity that is context specific.

Recommender systems function more effectively when they know more about a user (experienced in some contexts as “surveillance”) and when they can use that information to enhance the experience of all users (experienced in some contexts as “exploitation”). Users are aware these are part of the “bargain” with the system.

Neither “Surveillance” nor “Exploitation” is a trigger event, and neither is a violation of expectations. Instead, they are contextual perspectives that diverge from the user’s conception of the “platform spirit” (DeVito, 2021). Spotify both surveils and exploits, and users are aware and accepting of this. However, in certain contexts and at certain times, concern reaches a temporary threshold or level of intensity where users require an explanation or a justification. The folk theories of Spotify users suggest a more nuanced view of what motivates the need for an explanation.

9.3.2 Multistakeholder Contexts

The HCXAI principles are human centered but who are the humans being prioritized? The focus is clearly on the end-user, particularly the non-expert, lay population but there are others involved who should be addressed in the principles. This is especially apparent in recommender systems which are multisided marketplaces with a diverse set of stakeholders. All the folk theories, “Complies,” “Dialogues,” “Decides,” “Surveils,” “Withholds and Conceals,” “Empathizes,” and “Exploits,” explicitly reference the presence and influence of Spotify but users also acknowledge that there are other stakeholders beyond Spotify whose interests also influence the system and, indirectly, its explanations (e.g., artists, music companies, music distributors, advertisers, and even other users). The explanatory system is informed by the needs, requirements, and preferences of these others. As a result, they inform the explanatory system, and mostly consequentially, the nature and extent of the information provided.

The “common ground” is more complex than the principles would suggest. The “explainer” involved is a network or aggregation of many, filtered through the explanatory system to the explainee. While the explanatory system is what the user encounters and engages with, it is a system guided and governed by the stakeholders. This economic and political perspective is absent in the HCXAI principles.

9.4 Where Folk Theories Augment HCXAI Principles

9.4.1 Consumer Protection

Are policy issues relevant to the HCXAI principles? Selbst & Barocas note that “questions about justifying a model are often just questions about policy in disguise” (Selbst & Barocas, 2018, p. 1133). The folk theories “Surveils,” “Exploits,” “Withholds and Conceals,” and “Decides” all raise issues of consumer protection: privacy, data protection, risk, and harm. While not all HCXAI implementations relate to consumer applications, the emphasis on the non-expert, lay population suggests a consumer focus. The HCXAI principles are silent on consumer protection issues.

The principles recognize that explanatory systems must be "accompanied by other things to succeed." A unique suggestion is to develop consumer-facing labels for data and algorithmic models analogous to the nutritional labels mandated by regulation for the food industry (Stoyanovich et al., 2020). Given that Spotify users enter into a contract with Spotify when they subscribe to the service and agree to Terms of Use, it is reasonable to position the HCXAI principles within a consumer protection framework.

9.4.2 Right to Explanation

While Mueller et al. references the EU General Data Protection Regulation (GDPR) and the "right to explanation," this right or policy does not appear in the HCXAI principles. The folk theories "Decides" and "Withholds and Conceals" both arise in part because users believe that Spotify is not completely forthcoming. "Decides" reflects the belief that Spotify's recommendation process is opaque. "Withholds and Conceals" amplifies this with the belief that Spotify deliberately hides information from users. Users believe Spotify should provide them with explanations although they described this as an expectation rather than a right. The principles should emphasize, if not a right, a user expectation and a provider obligation.

9.4.3 Manipulation

The folk theories "Surveils", "Exploits," "Empathizes," "Withholds and Conceals," and "Decides" all raise concerns about manipulation in the system and potentially in the explanatory system. The HCXAI principles indicate that explanations should or could be "persuasive" leading to "unjustified trust," but do not caution that they could also be manipulative, deceptive, or coercive. Perhaps this is assumed, but in consumer applications where the typical power imbalance favours the provider not the user, explanations can easily be deployed against the interests of the user. As a result of this possible manipulation in explanations, researchers have suggested that a “right to explanation” is “not a sufficient condition for ensuring fair, accountable, and transparent use of AI” (Schoeffer et al., 2022, p. 4). Perceptions of deception and manipulation are evident in the “Decides” and “Withholds and Conceals” folk theories. Given the opacity of the system, users harbour beliefs that Spotify is making recommendations that favour Spotify’s financial interests and not the preferences of the user. While not raised by Spotify users, “Withholds and Conceals” suggests the growing concern with “shadow banning” where users of recommender systems or social media believe some of their comments, selections, and preferences are deliberately not collected or are hidden by the system because they reflect undesirable choices or opinions (Savolainen, 2022).

However, the folk theory “Empathizes” suggests an emerging avenue for manipulation where the emotional state of users is captured or inferred and utilized to shape or direct their actions. The danger and consequences of such data capture is widely criticized (Crawford, 2021; Stark & Hoey, 2021).

The principles criticize “transparency” in explanations as insufficient and argue for “apparency” (i.e., systems and explanations that are “readily understood and not hidden”) but acknowledge that this is “still not enough.” The principles focus on obtaining and sustaining trust but not on the actions or explanations that would undermine that trust. Human-centered XAI should guard against manipulation and deception and the HCXAI principles should articulate this.

9.4.4 Explanatory Systems

The HCXAI principles emphasize explanatory systems over explanations. The folk theories that reflect concerns about Spotify (“Decides,” “Surveils,” “Withholds and Conceals,” and “Exploits”) challenge confidence and trust in the system. Users don’t trust Spotify as a “data company.” The lack of trust in Spotify could extend to the explanatory system and raises the following questions: Whose explanatory system is it? Where does the system reside? Does the user have any influence on the system?

While some XAI techniques are post-hoc and independent from the system, in the consumer-facing applications the explanations (i.e., the explanatory systems) are embedded in the application. The APIs available from Spotify and other recommender systems allow access for limited data extraction but nothing sufficient to enable an external explanatory system. Given the focus on user goals, objectives, and contexts, the principles articulate explanatory systems where there is little or no input or influence from users. The principles outline what developers should provide to users through explanatory systems and not how developers and users are partners in the explanatory process.

The principles could, and perhaps should, preference explanatory systems that have some, or even complete, independence from the target service. Enabling an external, independent explanatory system or an explanatory agent would require technical protocols yet to be defined and policy requirements yet to be devised and imposed. However, such a system or agent could be tuned by the user to reflect their algorithmic literacy, experience with other systems, and preferences for types and extent of explanatory detail. One way to operationalize the “right to explanation” would be by requiring such access in a way similar to the idea of requiring that machine learning systems be “auditable” (Sandvig et al., 2014). A modest attempt at this is the eXplanatory Interactive Learning (XIL) module that lets a user challenge and improve an explanation as well as allow the system to query the user (Weber et al., 2022).

9.4.5 Pedagogy

In recommending that HCXAI be more than just explanations, the principles mention learning tools (e.g., tutorials) but not a pedagogical strategy to inform them as urged by others (Clancey & Hoffman, 2021; Kawakami et al., 2022). While it is obvious that

HCXAI is expected to explain, it is not as clear that it is expected to teach as well. Explanation is not instruction.

Many Spotify users hold the incorrect belief that personalized recommendations are made “solely by algorithms.” This is reflected partly in the folk theory that Spotify “Decides.” A priority for the “knowledge transformation and sensemaking” principle is “changing previous beliefs and preconceptions.” If a user’s incorrect belief as fundamental as “solely by algorithms” is to change, it will require more than explanations and the explanatory system envisioned by the principles. Preconceptions may be amendable to change but beliefs are more persistent.

DeVito argues that enhanced algorithmic literacy is the path to more complex and accurate folk theories of machine learning systems (DeVito, 2021). This is a pedagogical strategy which is consistent with the principles but which the principles fail to acknowledge and incorporate. While “learning objectives” are noted in the principles, these relate to those determined by the developers not those relevant to the user. Incorporating a pedagogical role requires a greater focus on teaching and learning principles.

9.4.6 Narratives: Play and Storytelling

Norman refers to folk theories as “storytelling” (Norman, 1983a), Siles et al. call them “vivid stories” (Siles et al., 2020), and Villareale & Zhu see play and folk theories linked (Villareale & Zhu, 2021). The folk theories “Decides” and “Empathizes” are classified as abstract and functional. The explanatory strategy for users who hold these beliefs is “repetition and elaboration, and the use of storytelling techniques” (DeVito, 2021, p. 339:32). Stories and play are narratives where users are engaged more deeply and informed less didactically. Narratives as part of an HCXAI explanatory system would advance accessibility through simplification and facilitate elaboration and repetition without being tiresome.

9.4.7 Anthropomorphizing

All the user folk theories anthropomorphize Spotify: it "Complies," "Dialogues," "Decides," "Surveils," "Withholds and Conceals," "Empathizes," and "Exploits." These indicate user beliefs that they are communicating with a system with human characteristics (Guzman, 2018). However, deliberate anthropomorphization in machine learning systems is widely critiqued as a barrier to user understanding and trust (Glikson & Woolley, 2020; Ngo & Krämer, 2021; D. Watson, 2019). While the HCXAI principles do not mention anthropomorphizing as a strategy or approach, the Spotify user folk theories suggest strongly that it is part of a human centered explanatory system strategy.

9.5 Summary of Folk Theories and HCXAI

The folk theories of Spotify users describe beliefs about agency, power, process, intent, and relationships. Applied to HCXAI, the folk theories support, challenge, and augment the principles of HCXAI. Taken collectively, the folk theories encourage HCXAI to take a broader view of XAI. The questions and concerns implicit in the folk theories indicate that users have explanatory issues that extend beyond the model veracity and authorization. The objective of HCXAI is to move towards a more user-centered, less technically focused XAI. This requires adopting principles that include policy implications, consumer protection issues, and concerns about intention and the possibility of manipulation.

10 Conclusion

Latanya Sweeney, Director of the Public Interest Tech Lab at Harvard, notes that "technology designers are the new policymakers; we didn't elect them but their decisions determine the rules we live by" (Sweeney, 2018). The principles of human-centered explainable AI (HCXAI) were motivated by the "need for use-inspired human-focused guidelines for XAI" (Mueller et al., 2021) that help these new "policymakers" be more responsive to the needs of users, particularly the non-expert, lay public. Folk theories describe the beliefs users hold about how machine learning systems work. They are a window into the way people and technology interact and communicate. Understanding folk theories provides insights into how XAI can be more effectively designed and deployed resulting in machine learning explainability that is more user focused.

The following research questions informed this study:

> What are the folk theories of users that explain how a recommender system works?
>
> Is there a relationship between the folk theories of users and the principles of HCXAI that would facilitate the development of more transparent and explainable recommender systems?

Using the Spotify music recommendation system as an example, users were surveyed and interviewed to elicit the folk theories of how personalized recommendations work in a machine learning system. Seven folk theories emerged: complies, dialogues, decides, surveils, withholds and conceals, empathizes, and exploits. These folk theories provide insights that support, challenge, and augment the HCXAI principles enhancing the guidelines for XAI system developers and designers. If utilized, these enhanced HCXAI guidelines offer a broader view of XAI and a user-centered focus that will improve machine learning explainability.

During their interview User 3 noted that “an algorithm doesn’t enjoy listening to music ... an algorithm doesn’t understand the emotional or cultural role that music plays in our lives ... even if an algorithm could make the best recommendations possible, what does that mean?” User 3 summarized their assessment of the Spotify recommender system saying “it’s precise, but it’s not accurate.” Algorithms, and the results they produce, instill complex beliefs in the users of those systems. Using folk theories to enhance HCXAI echoes Norman’s challenge for the field “It is the duty of machines and those who design them to understand people. It is not our duty to understand the arbitrary, meaningless dictates of machines.” (Norman, 2013, p. 6).

10.1 Contributions of the Study

The complexity and opacity of machine learning has compelled the need for explainability. Consumer services like Amazon, Facebook, TikTok, and Spotify have resulted in machine learning becoming ubiquitous in the everyday lives of the non-expert, lay public. In an age of “surveillance capitalism” (Zuboff, 2018), effective HCXAI is a critical component of a consumer protection strategy.

The elicited folk theories from Spotify users provide a unique view into how users of machine learning-based recommender systems believe they work. These beliefs describe not only how a user understands the system but how they must engage and interact with it. As a window into the complex user beliefs that inform their interactions with Spotify, the folk theories offer insights into how HCXAI systems can more effectively provide machine learning explainability to the non-expert, lay public.

While the folk theories in this study were used to enhance HCXAI principles, they can be applied beneficially to other areas. Caution is advised in generalizing beyond recommender systems like Spotify or machine learning systems with a consumer focus. However, the folk theories reflect beliefs applicable to system design (beyond XAI design), user training, policy development, and educational strategies for topics such as algorithmic literacy or general AI awareness. Policy development regarding consumer protection and a “right to explanation” is informed by the beliefs users hold about machine learning. In the case of algorithmic literacy, the accuracies, inaccuracies, or concerns evident in the folk theories allow for pedagogical design or interventions that “meet the user where they are."

10.2 Limitations

This study has several limitations which restrict the generalizability of the findings. The focus on a single recommender system (Spotify) as an example of recommender systems and machine learning allowed for more specific detail to emerge from users. However, investigating multiple systems might have resulted in a broader set of folk theories that would be more generalizable. In addition, unlike the consequential algorithmic decision-making systems where there is a material impact on users, Spotify’s recommendations are relatively benign. More consequential settings might have elicited different folk theories.

The sample size, 19 Spotify users, is small. While concepts and beliefs recurred during the survey and interviews, suggesting saturation, a larger sample might have resulted in a broader and more diverse set of folk theories. Finally, as with all research methods regarding the elicitation of folk theories, the research methodology used has known weaknesses. Surveys and interviews focus on reflective experience. Users were not observed using the system but rather remembering their use of the system.

Spotify, as with most recommender and machine learning systems, is rapidly evolving. Following the user survey and interviews in early 2021, Spotify has added new features and algorithmic processes. The observations of the users may no longer reflect the specific conditions of the current Spotify system.

10.3 Future Research

Future research could explore how XAI developers use the enhanced HCXAI principles informed by folk theories. This could include not only how the principles affect XAI system design but also how the resulting XAI systems are evaluated by users.

Replication studies have become increasingly important in machine learning (Haibe-Kains et al., 2020; Pineau, 2017). A reproducibility study of Spotify or a comparable, non-consequential consumer recommender system could validate the results of this study. A similar objective could be obtained with a multi-system study of consumer recommender systems.

The applicability of folk theories to instructional strategies regarding algorithmic literacy has been noted here and in prior research (DeVito, 2021). Designing a pedagogical framework using the elicited folk theories could evaluate their effectiveness in advancing algorithmic literacy.

References

Abdollahpouri, H., Burke, R., & Mobasher, B. (2017). Recommender systems as multistakeholder environments. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 347–348. https://doi.org/10.1145/3079628.3079657

Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B., & Malthouse, E. (2021). User-centered evaluation of popularity bias in recommender systems. *ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21)*. ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21), Utrecht, Netherlands. https://doi.org/10.1145/3450613.3456821

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable, and intelligible systems: An HCI research agenda. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 582:1-582:18. https://doi.org/10.1145/3173574.3174156

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Aggarwal, C. C. (2016). *Recommender systems*. Springer.

Akhtar-Danesh, N. (2017). A comparison between major factor extraction and factor rotation techniques in Q-methodology. *Open Journal of Applied Sciences*, *07*(04), 147–156. https://doi.org/10.4236/ojapps.2017.74013

Aldrich, S. E. (2011). Recommender systems in commercial use. *AI Magazine*, *32*(3), 28–34. https://doi.org/10.1609/aimag.v32i3.2368

Alharthi, H., Inkpen, D., & Szpakowicz, S. (2018). A survey of book recommender systems. *Journal of Intelligent Information Systems*, *51*(1), 139–160. https://doi.org/10.1007/s10844-017-0489-9

Alipour, K., Ray, A., Lin, X., Cogswell, M., Schulze, J. P., Yao, Y., & Burachas, G. T. (2021). Improving users' mental model with attention-directed counterfactual edits. *Applied AI Letters*, e47. https://doi.org/10.1002/ail2.47

Allen, R. B. (1997). Mental models and user models. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed., pp. 49–63). North Holland.

Aluri, R., & Riggs, D. E. (1988). Application of expert systems to libraries. *Advances in Library Automation and Networking*, *2*, 1–43.

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R., & Lalmas, M. (2020). Algorithmic effects on the diversity of consumption on Spotify. *Proceedings of The Web Conference 2020*, 2155–2165. https://doi.org/10.1145/3366423.3380281

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). "Just the way you are": Linking music listening on Spotify and personality: *Social Psychological and Personality Science*, *12*(4), 561–572. https://doi.org/10.1177/1948550620923228

Anderson, J. (2020). Understanding and interpreting algorithms: Toward a hermeneutics of algorithms. *Media, Culture & Society*, *42*(7–8), 1479–1494. https://doi.org/10.1177/0163443720919373

Aragon, C., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-centered data science*. MIT Press.

Armann-Keown, V., & Patterson, L. (2020). Content analysis in library and information research: An analysis of trends. *Library & Information Science Research*, *42*(4). https://doi.org/10.1016/j.lisr.2020.101048

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131

Association for Computing Machinery. (2017). *Statement on algorithmic transparency and accountability*. ACM. http://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf

Baer, J. (2019, December 13). The winding road to better machine learning infrastructure through Tensorflow Extended and Kubeflow. *Spotify Labs*. https://labs.spotify.com/2019/12/13/the-winding-road-to-better-machine-learning-infrastructure-through-tensorflow-extended-and-kubeflow/

Bailey, J., Steeves, V., Burkell, J., Shade, L. R., Ruparelia, R., & Regan, P. (2019). Getting at equality: Research methods informed by the lessons of intersectionality. *International Journal of Qualitative Methods*, *18*. https://doi.org/10.1177/1609406919846753

Banasick, S. (2019). KADE: A desktop application for Q methodology. *The Journal of Open Source Software*, *4*(36). https://doi.org/10.21105/joss.01360

Bates, M. J. (2005). An introduction to metatheories, theories, and models. In K. E. Fisher & S. Erdelez (Eds.), *Theories of information behavior* (pp. 1–24). American Society for Information Science and Technology.

Bechlivanidis, C., Lagnado, D. A., Zemla, J. C., & Sloman, S. (2017). Concreteness and abstraction in everyday explanation. *Psychonomic Bulletin & Review*, *12*, 1451–1464. https://doi.org/10.3758/s13423-017-1299-3

Beel, J., Langer, S., Genzmehr, M., & Nürnberger, A. (2013). Introducing Docear's research paper recommender system. *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '13*, 459. https://doi.org/10.1145/2467696.2467786

Berger-Tal, O., Nathan, J., Meron, E., & Saltz, D. (2014). The exploration-exploitation dilemma: A multidisciplinary framework. *PLoS One*, *9*(4). https://doi.org/10.1371/journal.pone.0095693

Bernardi, L., Mavridis, T., & Estevez, P. (2019). 150 successful machine learning models: 6 lessons learned at Booking.Com. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1743–1751. https://doi.org/10.1145/3292500.3330744

Besseny, A. (2020). Lost in Spotify: Folksonomy and wayfinding functions in Spotify's interface and companion apps. *Popular Communication*, *18*, 1–17. https://doi.org/10.1080/15405702.2019.1701674

Bibal, A., & Frénay, B. (2016). Interpretability of machine learning models and representations: An introduction. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 77–82. https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2016-141.pdf

Bieringer, L., Grosse, K., Backes, M., & Krombholz, K. (2021). Mental models of adversarial machine learning. *ArXiv:2105*. http://arxiv.org/abs/2105.03726

Biran, O., & Cotton, C. (2017). Explanation and justification in machine learning: A survey. *International Joint Conference on Artificial Intelligence. Workshop on Explainable Artificial Intelligence (XAI)*. International Joint Conference on Artificial Intelligence. Workshop on Explainable Artificial Intelligence (XAI), Melbourne. http://www.cs.columbia.edu/~orb/papers/xai_survey_paper_2017.pdf

Bobadilla, J., Ortega, F., Hernado, A., & Gutierrez, A. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems With Applications*, *41*(16), 7370–7389.

Bogost, I. (2016). *Play anything: The pleasure of limits, the uses of boredom, and the secret of games*. Basic Books.

Bolland, J. M. (1985). The search for structure: An alternative to the forced Q-sort technique. *Political Methodology*, *11*(1/2), 91–107.

Bonnefon, J.-F., & Rahwan, I. (2020). Machine thinking, fast and slow. *Trends in Cognitive Sciences*, *24*(12), 1019–1027. https://doi.org/10.1016/j.tics.2020.09.007

Bos, N., Glasgow, K., Gersh, J., Harbison, I., & Lyn Paul, C. (2019). Mental models of AI-based systems: User predictions and explanations of image classification results. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 184–188. https://doi.org/10.1177/1071181319631392

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.

Brown, S. R. (1971). The forced-free distinction in Q technique. *Journal of Educational Measurement*, *8*(4), 283–287. https://doi.org/10.1111/j.1745-3984.1971.tb00939.x

Brown, S. R. (1980). *Political subjectivity; Applications of Q methodology in political science*. Yale University Press.

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, *20*(1), 30–44. https://doi.org/10.1080/1369118X.2016.1154086

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.

Büchi, M., Fosch-Villaronga, E., Lutz, C., Tamò-Larrieux, A., & Velidi, S. (2021). Making sense of algorithmic profiling: User perceptions on Facebook. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118X.2021.1989011

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceeding of the ACM on Human-Computer Interaction*, *5*(CSCW1), 188:1-188:21. https://doi.org/10.1145/3449287

Bunn, J. (2020). Working in contexts for which transparency is important: A recordkeeping view of explainable artificial intelligence (XAI). *Records Management Journal*, *30*(2), 143–153. https://doi.org/10.1108/RMJ-08-2019-0038

Burkart, N., & Huber, M. F. (2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, *70*, 245–317. https://doi.org/10.1613/jair.1.12228

Cabour, G., Morales, A., Ledoux, É., & Bassetto, S. (2021). Towards an explanation space to align humans and explainable-AI teamwork. *ArXivcarroll*. http://arxiv.org/abs/2106.01503

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 report*. AI Now Institute.

Canada. (2019). *Algorithmic impact assessment*. https://canada-ca.github.io/aia-eia-js/

Canada. Office of the Privacy Commissioner of. (2020). *A regulatory framework for AI: Recommendations for PIPEDA reform*. https://www.priv.gc.ca/en/about-the-opc/what-we-do/consultations/completed-consultations/consultation-ai/reg-fw_202011/

Canadian Institutes of Health Research, Natural Sciences and engineering Research Council of Canada, & Social Sciences and Humanities Research Council. (2018). *Tri-Council policy statement: Ethical conduct for research involving humans*. http://pre.ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf

Carlsson, S., & Leijonhufvud, J. (2021). *The Spotify play*. Diversion Books.

Carroll, J. M., & Olson, J. R. (1988). Mental models in human-computer interaction. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 45–65). North-Holland. https://doi.org/10.1016/B978-0-444-70536-5.50007-5

Casey, B., Farhangi, A., & Vogl, R. (2019). Rethinking explainable machines: The GDPR's "Right to Explanation" debate and the rise of algorithmic audits in enterprise. *Berkeley Technology Law Journal*, *34*(1), 143–188. https://doi.org/10.15779/Z38M32N986

Castells, P., Hurley, N. J., & Vargas, S. (2015). Novelty and diversity in recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., pp. 881–918). Springer. 10.1007/978-0-387-85820-3_13

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Celma, O. (2010). *Music recommendation and discovery: The long tail, long fail, and long play in the digital music space*. Springer.

Chari, S., Seneviratne, O., Gruen, D. M., Foreman, M. A., Das, A. K., & McGuinness, D. L. (2020). Explanation ontology: A model of explanations for user-centered AI. In J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, & A. Plleres (Eds.), *The Semantic Web—ISWC 2020* (pp. 228–243). Springer.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed). Sage.

Charmaz, K., & Belgrave, L. L. (2012). Qualitative interviewing and grounded theory analysis. In J. Gubrium, J. Holstein, A. Marvasti, & K. McKinney (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (pp. 347–366). SAGE Publications. https://doi.org/10.4135/9781452218403.n25

Cho, J. (2018). Mental models and home virtual assistants (HVAs). *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3180286

Chodos, A. T. (2019). *Solving and dissolving musical affection: A critical study of Spotify and automated music recommendation in the 21st century* [PhD Dissertation, University of California San Diego]. https://escholarship.org/uc/item/2c27z9xk

Ciocca, S. (2017, October 10). *How does Spotify know you so well?* Medium. https://medium.com/s/story/spotifys-discover-weekly-how-machine-learning-finds-your-new-music-19a41ab76efe

Clancey, W. J. (1983). The epistemology of a rule-based expert system—A framework for explanation. *Artificial Intelligence*, *20*(3), 215–251. https://doi.org/10.1016/0004-3702(83)90008-5

Clancey, W. J., & Hoffman, R. R. (2021). Methods and standards for research on explainable artificial intelligence: Lessons from intelligent tutoring systems. *Applied AI Letters*. https://doi.org/10.1002/ail2.53

Colbjørnsen, T. (2018). My algorithm: User perceptions of algorithmic recommendations in cultural contexts. In A. L. Guzman (Ed.), *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (pp. 167–184). Peter Lang.

Cole, C., Lin, Y., Leide, J., Large, A., & Beheshti, J. (2007). A classification of mental models of undergraduates seeking information for a course essay in history and psychology: Preliminary investigations into aligning their mental models with online thesauri. *Journal of the American Society for Information Science and Technology*, *58*(13), 2092–2104. https://doi.org/10.1002/asi.20668

Collins, A., & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 243–265). Cambridge University Press.

Connaway, L. S., & Powell, R. R. (2010). *Basic research methods for librarians* (5th ed.). Libraries Unlimited.

Consumers International. (2019). *Artificial intelligence: Consumer experiences in new technology*. https://www.consumersinternational.org/media/261949/ai-consumerexperiencesinnewtech.pdf

Cordell, R. (2020). *Machine learning + libraries: A report on the state of the field*. Library of Congress. https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf

Cowen, A. S., Fang, X., Sauter, D., & Keltner, D. (2020). What music makes us feel. *Proceedings of the National Academy of Sciences*, *117*(4), 1924–1934. https://doi.org/10.1073/pnas.1910704117

Cox, A. M. (2021). *The impact of AI, machine learning, automation and robotics on the information professions*. CILIP. www.cilip.org.uk/resource/resmgr/cilip/research/tech_review/cilip_–_ai_report_-_final_lo.pdf

Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries. *Library Hi Tech*, *37*(3), 318–435. https://doi.org/10.1108/LHT-08-2018-0105

Cramer, H. (2010). *People's responses to autonomous and adaptive systems* [PhD, University of Amsterdam]. https://hdl.handle.net/11245/1.318609

Crawford, K. (2021). Time to regulate AI that interprets human emotions. *Nature*, *592*(7853), 167. https://doi.org/10.1038/d41586-021-00868-5

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE Publications.

Dai, Z., Chandar, P., Fazelnia, G., Carterette, B., & Lalmas, M. (2020). Model selection for production system via automated online experiments. *34th Conference on Neural Information Processing Systems*. https://proceedings.neurips.cc/paper/2020/file/0c72cb7ee1512f800abe27823a792d03-Paper.pdf

DARPA. (2016). *Explainable artificial intelligence (XAI)*. DARPA. http://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf

Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *ArXiv*. http://arxiv.org/abs/2006.11371

Davis, B., Glenski, M., Sealy, W., & Arendt, D. (2020). Measure utility, gain trust: Practical advice for XAI researcher. *ArXiv*. http://arxiv.org/abs/2009.12924

de Mul, J., & van den Berg, B. (2011). Remote control: Human autonomy in the age of computer-mediated agency. In M. Hildebrandt & A. Rouvroy (Eds.), *Law, human agency, and autonomic computing* (pp. 46–63). Routledge.

de Quirós, J. G., Baldassarri, S., Beltrán, J. R., Guiu, A., & Álvarez, P. (2019). An automatic emotion recognition system for annotating Spotify's songs. In H.

Panetto, C. Debruyne, M. Hepp, D. Lewis, C. A. Ardagna, & R. Meersman (Eds.), *On the Move to Meaningful Internet Systems: OTM 2019 Conferences* (pp. 345–362). Springer. https://doi.org/10.1007/978-3-030-33246-4_23

Deahl, D. (2017). *Spotify is testing "Sponsored Songs" in playlists*. The Verge. https://www.theverge.com/2017/6/19/15833880/spotify-sponsored-songs-playlists-test

Deahl, D. (2019). *Metadata is the biggest little problem plaguing the music industry*. The Verge. https://www.theverge.com/2019/5/29/18531476/music-industry-song-royalties-metadata-credit-problems

Dervin, B. (1993). Verbing communication: Mandate for disciplinary invention. *Journal of Communication*, *43*(3), 45–54. https://doi.org/10.1111/j.1460-2466.1993.tb01275.x

Dervin, B. (1994). Information - democracy: An examination of underlying assumptions. *Journal of the American Society for Information Science*, *45*(6), 369–385.

Dervin, B. (1999). On studying information seeking methodology: The implications of connecting metatheory to method. *Information Processing & Management*, *35*, 727–750.

DeVito, M. A. (2021). Adaptive folk theorization as a path to algorithmic literacy on changing platforms. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 339:1-339:38. https://doi.org/10.1145/3476080

DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., & Liu, S. (2018). How people form folk theories of social media feeds and what it means for how we study self-presentation. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3173694

DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). “Algorithms ruin everything”: #RIPTwitter, folk theories, and resistance to algorithmic change in social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3163–3174. https://doi.org/10.1145/3025453.3025659

Dhillon, I. S., Mallela, S., & Modha, D. S. (2003). Information-theoretic co-clustering. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’03*, 89–98. https://doi.org/10.1145/956750.956764

Diederich, J. (2018). Methods for the explanation of machine learning processes and results for non-experts. *PsyArXiv*. https://doi.org/10.31234/osf.io/54eub

Dieleman, S. (2014, August 5). *Recommending music on Spotify with deep learning*. https://benanne.github.io/2014/08/05/spotify-cnns.html

Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K. E., & Dugan, C. (2019). Explaining models: An empirical study of how explanations impact fairness judgment. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 275–285. https://doi.org/10.1145/3301275.3302310

Doherty, K., & Doherty, G. (2018). The construal of experience in HCI: Understanding self-reports. *International Journal of Human - Computer Studies*, *110*, 63–74. https://doi.org/10.1016/j.ijhcs.2017.10.006

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87. https://doi.org/10.1145/2347736.2347755

Domingos, P. (2015). *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *ArXiv*. https://arxiv.org/abs/1702.08608

Douglas, N. (2018, February 3). If your Discover Weekly playlist sucks, try this. *Lifehacker*. https://lifehacker.com/if-your-discover-weekly-playlist-sucks-try-this-1823436104

Dredge, S. (2020, July 30). Spotify CEO talks Covid-19, artist incomes and podcasting. *Musically, ferraro*. https://musically.com/2020/07/30/spotify-ceo-talks-covid-19-artist-incomes-and-podcasting-interview/

Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a “right to explanation” is probably not the remedy you are looking for. *Duke Law & Technology Review*, *16*, 18–84.

Edwards, L., & Veale, M. (2018). Enslaving the algorithm: From a 'right to an explanation' to a 'right to better decisions'? *IEEE Security & Privacy*, *16*(3), 46–54.

Ehsan, U., & Riedl, M. O. (2020). Human-centered explainable AI: Towards a reflective sociotechnical approach. *Proceedings of HCI International Conference on Human-Computer Interaction*. http://arxiv.org/abs/2002.01092

Ehsan, U., Wintersberger, P., Liao, Q. V., Watkins, E. A., Manger, C., Daume, H., Riener, A., & Riedl, M. O. (2022). *Human-centered explainable AI (HCXAI): Beyond opening the black-box of AI [Workshop proposal]*. ACM CHI Conference on Human Factors in Computing Systems, New Orleans. https://www.dropbox.com/s/w3io07tp0u1sjq0/HCXAI_CHI_2022_WS_Proposal.pdf

Elk, D. (2021, April 15). *Daniel Elk (Spotify) to Isedua Orbhabor (Access Now)*. https://www.accessnow.org/cms/assets/uploads/2021/04/Spotify-Letter-to-Access-Now-04-15-2021-.pdf

Eriksson, M. (2016). Close reading big data: The Echo Nest and the production of (rotten) music metadata. *First Monday*, *21*(7). https://doi.org/10.5210/fm.v21i7.6303

Eriksson, M., Fleischer, R., Joansson, A., Snickars, P., & Vonderau, P. (2019). *Spotify teardown: Inside the black box of streaming music*. MIT Press.

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016). First I "like" it, then I hide it: Folk theories of social feeds. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2371–2382. https://doi.org/10.1145/2858036.2858494

European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016*. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

FAT/ML. (n.d.). *Principles for accountable algorithms and a social impact statement for algorithms*. Retrieved August 16, 2017, from http://www.fatml.org/resources/principles-for-accountable-algorithms

Ferraro, A., Serra, X., & Bauer, C. (2021). What is fair? Exploring the artists' perspective on the fairness of music streaming platforms. In C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, & A. Piccinno (Eds.), *Proceedings of the 18th IFIP international conference on human-computer interaction* (pp. 562–584). http://arxiv.org/abs/2106.02415

Fleischer, R., & Snickars, P. (2017). Discovering Spotify. *Culture Unbound*, *9*(2), 130–145. https://doi.org/10.3384/cu.2000.1525.1792

Forsythe, D. (2001). *Studying those who study us: An anthropologist in the world of artificial intelligence* (D. J. Hess, Ed.). Stanford University Press.

Fransella, F. (2016). What is a personal construct? In D. A. Winter & N. Reed (Eds.), *The Wiley handbook of personal construct psychology* (pp. 1–8). John Wiley & Sons.

French, M., & Hancock, J. (2017). *What's the folk theory? Reasoning about cyber-social systems*. 67th Annual Conference of the International Communication Association. https://doi.org/10.2139/ssrn.2910571

Galuszka, P. (2015). Music aggregators and intermediation of the digital music market. *International Journal of Communication*, *9*, 254–273. https://doi.org/1932–8036/20150005

Gandy, O. H. (2011). Consumer protection in cyberspace. *TripleC*, *9*(2), 175–189. https://doi.org/10.31269/triplec.v9i2.267

Garcia-Gathright, J., St. Thomas, B., Hosey, C., Nazari, Z., & Diaz, F. (2018). Understanding and evaluating user satisfaction with music discovery. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 55–64. https://doi.org/10.1145/3209978.3210049

Garfinkel, A. (1981). *Forms of explanation: Structures of inquiry in social science*. Yale University Press.

Gasparini, A., & Kautonen, H. (2022). Understanding artificial intelligence in research libraries: An extensive literature review. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, *32*(1), Article 1. https://doi.org/10.53377/lq.10934

Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual Review of Anthropology*, *40*, 379–398. https://doi.org/10.1146/annurev-anthro-081309-145822

Gentile, D., Jamieson, G., & Donmez, B. (2021). Evaluating human understanding in XAI systems. *HCXAI '21: ACM CHI Workshop on Human-Centered Perspectives in Explainable AI*. https://www.dropbox.com/s/33riks0r050c6nx/HCXAI2021_paper_25.pdf

Gerlings, J., Shollo, A., & Constantiou, I. (2020). Reviewing the need for explainable artificial intelligence (xAI). *Proceedings of the Hawaii International Conference on System Sciences*. http://arxiv.org/abs/2012.01007

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academic of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, *35*(12), 61–70.

Goldschmitt, K. E., & Seaver, N. (2019). Shaping the stream: Techniques and troubles of algorithmic recommendation. In N. Cook, M. Ingalls, & D. Trippett (Eds.), *The Cambridge Companion to Music in Digital Culture* (pp. 63–81). Cambridge University Press.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision making and a "right to explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Graichen, L., Graichen, M., & Petrosjan, M. (2022). *How to facilitate mental model buidling and mitigate overtrust using HCXAI*. ACM CHI 2022 Workshop on

Human-Centered Explainable AI (HCXAI). https://www.dropbox.com/s/uptmz0d1xyx2ml6/HCXAI2022_paper_12.pdf

Griffey, J. (Ed.). (2019). Artificial intelligence and machine learning in libraries. *Library Technology Reports*, *55*(1). https://doi.org/10.5860/ltr.55n1

Grimes, G. M., Schuetzler, R. M., & Giboney, J. S. (2021). Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, *144*(113515), 1–10. https://doi.org/10.1016/j.dss.2021.113515

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Gunning, D., Vorm, E., Wang, J. Y., & Turek, M. (2021). DARPA's explainable AI (XAI) program: A retrospective. *Applied AI Letters*. https://doi.org/10.1002/ail2.61

Guzman, A. L. (Ed.). (2018). *Human-machine communication: Rethinking communication, technology, and ourselves*. Peter Lang.

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Shraddha, T., Kusko, R., & Sansone, S.-A. (2020). Transparency and reproducibility in artificial intelligence. *Nature*, *586*(7829), E14–E16. https://doi.org/10.1038/s41586-020-2766-y

Hamilton, K., Karahalios, K., Sandvig, C., & Eslami, M. (2014). A path to understanding the effects of algorithm awareness. *CHI '14 Extended Abstracts on Human*

Factors in Computing Systems, 631–642. https://doi.org/10.1145/2559206.2578883

Han, Z., Hansen, P., Xu, H., & Luo, R. (2020). Examining the classification and evolution of novice users' mental models of an academic database in the search task completion process. *Journal of Information Science, 46*(2), 205–225. https://doi.org/10.1177%2F0165551519828621

Hank, C., Jordan, M. W., & Wildemuth, B. M. (2009). Survey research. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 256–269). Libraries Unlimited.

Hansen, C., Mehrotra, R., Hansen, C., Brost, B., Maystre, L., & Lalmas, M. (2021). Shifting consumption towards diverse content on music streaming platforms. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 238–246. https://doi.org/10.1145/3437963.3441775

Hartel, J., Cox, A. M., & Griffin, B. L. (2016). Information activity in serious leisure. *Information Research, 21*(4). http://informationr.net/ir/21-4/paper728.html

He, W., Erdelez, S., Wang, F.-K., & Shyu, C.-R. (2008). The effects of conceptual description and search practice on users' mental models and information seeking in a case-based reasoning retrieval system. *Information Processing & Management*, *44,* 294–309. https://doi.org/10.1016/j.ipm.2007.03.008

Heath, A. (2015, September 14). Spotify has a secret “taste profile” on everyone, and they showed me mine. *Business Insider*. https://www.businessinsider.com/how-spotify-taste-profiles-work-2015-9

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2019). Metrics for explainable AI: Challenges and prospects. *ArXiv*. http://arxiv.org/abs/1812.04608

Hohman, F., Kahng, M., Pienta, R., & Chau, D. H. (2019). Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. *IEEE Transactions on Visualization and Computer Graphics*, *25*(8), 2674–2693. https://doi.org/10.1109/tvcg.2018.2843369

Hosey, C., Vujović, L., St. Thomas, B., Garcia-Gathright, J., & Thom, J. (2019). Just give me what I want: How people use and evaluate music search. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300529

Hove, S. E., & Anda, B. (2005). Experiences from conducting semi-structured interviews in empirical software engineering research. *11th IEEE International Software Metrics Symposium*, *2005*. https://doi.org/10.1109/METRICS.2005.24

Howells, G. (2020). Protecting consumer protection values in the fourth industrial revolution. *Journal of Consumer Policy*, *43*(1), 145–175. https://doi.org/10.1007/s10603-019-09430-3

Hsieh, C., & Hall, W. (1989). Survey of artificial intelligence and expert systems in library and information science literature. *Information Technology and Libraries*, *8*(2), 209.

Huang, S. A., Hancock, J., & Tong, S. T. (2022). Folk theories of online dating: Exploring people's beliefs about the online dating process and online dating algorithms. *Social Media + Society*, *8*(2). https://doi.org/10.1177/20563051221089561

Huland, S. (2021). *Identification of taste attributes from an audio signal* (United States Patent No. US 10,891,948 B2).

IEEE. (2019). *Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems*. IEEE. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf

Jackson, K. (2019). *Qualitative data analysis with NVivo* (Third edition.). Sage Publications.

Jacob, E. (2021, February 24). How much does Spotify pay per stream? What you'll earn per song, and how to get paid more for your music. *Business Insider*. https://www.businessinsider.com/how-much-does-spotify-pay-per-stream

Jacovi, A., Bastings, J., Gehrmann, S., Goldberg, Y., & Filippova, K. (2022). Diagnosing AI Explanation methods with folk concepts of behavior. *ArXiv*, 1–22.

Jannach, D., Resnick, P., Tuzhilin, A., & Zanker, M. (2016). Recommender systems—Beyond matrix completion. *Communications of the ACM*, *59*(11), 94–102. https://doi.org/10.1145/2891406

Jannach, D., Zanker, M., Felfering, A., & Friedrich, G. (2011). *Recommender systems: An introduction*. Cambridge University Press.

Jebara, T. (2020, January 16). For your ears only: Personalizing Spotify Home with machine learning. *Spotify Labs*. https://labs.spotify.com/2020/01/16/for-your-ears-only-personalizing-spotify-home-with-machine-learning/

Jehan, T., & DesRoches, D. (2014). *Analyzer documentation*. The Echo Nest. http://docs.echonest.com.s3-website-us-east-1.amazonaws.com/_static/AnalyzeDocumentation.pdf

Jehan, T., Lamere, P., & Whitman, B. (2010). Music retrieval from everything. *Proceedings of the International Conference on Multimedia Information Retrieval*, 245–246. https://doi.org/10.1145/1743384.1743428

Jennings, D. (2007). *Net, blogs and rock "n" roll: How digital discovery works and what it means for consumers*. Nicholas Brealey.

Jeyakumar, J. V., Noor, J., Cheng, Y.-H., Garcia, L., & Srivastava, M. (2020). How can I explain this to you? An empirical study of deep neural network explanation methods. *Advances in Neural Information Processing Systems*, *33*, 4211–4222. https://papers.nips.cc/paper/2020/file/2c29d89cc56cdb191c60db2f0bae796b-Paper.pdf

Johnson, D. (2020). *Machine learning, libraries, and cross-disciplinary research: Possibilities and provocations*. Hesburgh Libraries, University of Notre Dame. https://dx.doi.org/doi:10.7274/r0-wxg0-pe06

Kaminski, M. E. (2019). The right to explanation, explained. *Berkeley Technology Law Journal*, *34*(1), 189–218. https://doi.org/10.15779/Z38TD9N83H

Kaminski, M. E., & Ricci, F. (2012). Contextual music information retrieval and recommendations: State of the art and challenges. *Computer Science Review*, *6*(2–3), 89–119.

Kant, T. (2020). *Making it personal: Algorithmic personalization, identify, and everyday life*. Oxford University Press.

Karakayali, N., Kostem, B., & Galip, I. (2018). Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste. *Theory, Culture & Society*, *35*(2), 3–24. https://doi.org/10.1177/0263276417722391

Karizat, N., Delmonaco, D., Eslami, M., & Andalibi, N. (2021). Algorithmic folk theories and identity: How TikTok users co-produce knowledge of identity and engage in algorithmic resistance. *Proceedings of the ACM on Human-Computer Interaction*, *5*, 305:1-305:44. https://doi.org/10.1145/3476046

Kawakami, A., Arechiga, N., Guerdan, L., Lee, M., Cheng, Y., Carter, S., Sun, A., Zhu, H., Hu, A., Holstein, K., & Glazko, K. (2022). *Towards a learner-centered explainable AI*. ACM CHI 2022 Workshop on Human-Centered Explainable AI

(HCXAI). https://www.dropbox.com/s/wvvhi75m4rzvlmm/HCXAI2022_paper_30.pdf

Keane, M. T., Kenny, E. M., Delaney, E., & Smyth, B. (2021). If only we had better counterfactual explanations: Five key deficits to rectify in the evaluation of counterfactual XAI techniques. *ArXiv*. http://arxiv.org/abs/2103.01035

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, *31*, 457–501.

Kelly, G. (1955). *The psychology of personal constructs: A theory of personality*. Routledge. https://doi.org/10.4324/9780203405970

Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2020). A systematic review: Machine learning based recommendation systems for e-learning. *Education and Information Technologies*, *25*, 2635–2664. https://doi.org/10.1007/s10639-019-10063-9

Kim, D., Kim, K., Park, K.-H., Lee, J.-H., & Lee, K. M. (2007). A music recommendation system with a dynamic k-means clustering algorithm. *Sixth International Conference on Machine Learning and Applications (ICMLA 2007)*, 399–403. https://doi.org/10.1109/ICMLA.2007.97

Kim, T. W., Tong, Lu, Lee, K., Cheng, Z., Tang, Y., & Hooker, J. (2021). When is it permissible for artificial intelligence to lie? A trust-based approach. *ArXiv*. http://arxiv.org/abs/2103.05434

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. https://doi.org/10.1080/1369118X.2016.1154087

Kizilcec, R. F. (2016). *How much information? Effects of transparency on trust in an algorithmic interface*. 2390–2395. https://doi.org/dx.doi.org/10.1145/2858036.2858402

Knowles, B. (2022). *Explainable AI: Another sucessful failure*. ACM CHI 2022 Workshop on Human-Centered Explainable AI (HCXAI). https://www.dropbox.com/s/coiks0bk4eyy6xj/HCXAI2022_paper_04.pdf

Koch, J., & Fortes Rey, V. (2022). *Combing human-centred explainability and explainable AI*. ACM CHI 2022 Workshop on Human-Centered Explainable AI (HCXAI). https://www.dropbox.com/s/wqkox4ycnnh1q6c/HCXAI2022_paper_40.pdf

König, P. D., Wurster, S., & Siewert, M. B. (2022). Consumers are willing to pay a price for explainable, but not for green AI. Evidence from a choice-based conjoint analysis. *Big Data & Society*, *9*(1), 20539517211069630. https://doi.org/10.1177/20539517211069632

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30–37. https://doi.org/10.1109/MC.2009.263

Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., & Getoor, L. (2017). User Preferences for Hybrid Explanations. *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 84–88. https://doi.org/10.1145/3109859.3109915

Kozyrkov, C. (2018, November 16). Explainable AI won't deliver. Here's why. *Hacker Noon*. https://hackernoon.com/explainable-ai-wont-deliver-here-s-why-6738f54216be

Kudlow, P., Dziadyk, D. B., Rutledge, A., Shachak, A., & Eysenbach, G. (2019). The citation advantage of promoted articles in a cross-publisher distribution platform: A 12-month randomized controlled trial. *Journal of the Association for Information Science and Technology*, *71*, 1257–1274. https://doi.org/10.1002/asi.24330

Kühl, Ni., Lobana, J., & Meske, C. (2020). *Do you comply with AI? Personalized explanations of learning algorithms and their impact on employees' compliance behavior*. International conference on information systems, Munich. http://arxiv.org/abs/2002.08777

Kulesza, T., Stumpf, S., Burnett, M., & Kwan, I. (2012). Tell me more? The effects of mental model soundness on personalizing an intelligent agent. *Proceedings of the*

SIGCHI Conference on Human Factors in Computing Systems, 1–10. https://doi.org/10.1145/2207676.2207678

Kuleza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., & Wong, W.-K. (2013). Too much, too little, or just right? Ways Explanations impact end users' mental models. *IEEE Symposium on Visual Languages and Human Centric Computing*, 3–10. https://doi.org/10.1109/VLHCC.2013.6645235

Kumar, D. P., Sowmya, B. J., Chetan, & Srinivasa, K. G. (2016). A comparative study of classifiers for music genre classification based on feature extractors. *IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 190–194. https://doi.org/10.1109/DISCOVER.2016.7806258

Kuoppa, A. (2018). *Recommended by algorithms: Relevance, affordances and agency of music recommender systems* [Master's Thesis, University of Tampere]. http://urn.fi/URN:NBN:fi:uta-201806121995

Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. SAGE Publications.

Lalmas, M. (Director). (2019). *Spotify: Personalising the listening experience [Video]*. Turing Institute. https://youtu.be/IX8ZwVReQMs

Liao, Q. V., Gruen, D., & Miller, S. (2020). Questioning the AI: Informing design practices for explainable AI user experiences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376590

Liao, Q. V., Pribić, M., Han, J., Miller, S., & Sow, D. (2021). Question-driven design process for explainable AI user experiences. *ArXiv*. http://arxiv.org/abs/2104.03483

Liao, Q. V., & Varshney, K. R. (2022). *Human-centered explainable AI (XAI): From algorithms to user experiences*. http://arxiv.org/abs/2110.10790

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2119–2128. https://doi.org/10.1145/1518701.1519023

Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J., & Zhang, J. (2012). Expectation and purpose: Understanding users' mental models of mobile app privacy through crowdsourcing. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 501–510. https://doi.org/10.1145/2370216.2370290

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76–80.

Lippincott, S. (2020). *Mapping the current landscape of research library engagement with emerging technologies in research and learning*. Association of Research Libraries. https://www.arl.org/wp-content/uploads/2020/03/2020.03.25-emerging-technologies-landscape-summary.pdf

Lipton, Z. C. (2016). *The mythos of model interpretability*. ICML Workshop on Human Interpretability in Machine Learning. https://arxiv.org/abs/1606.03490

Lipton, Z. C. (2017). *The Doctor Just Won't Accept That!* NIPS 2017 Interpretable ML Symposium. https://doi.org/10.48550/arXiv.1711.08037

Liu, A., Zhang, Y., & Li, J. (2009). Personalized movie recommendation. *Proceedings of the 17th ACM International Conference on Multimedia*, 845–848. https://doi.org/10.1145/1631272.1631429

Liu, G. (2011). The application of intelligent agents in libraries: A survey. *Program: Electronic Library and Information Systems*, *45*(1), 78–97. https://doi.org/10.1108/00330331111107411

Lomborg, S., & Kapsch, P. H. (2019). Decoding algorithms. *Media, Culture & Society*, *42*(5), 745–761. https://doi.org/10.1177/0163443719855301

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Lonsdale, A. J., & North, A. C. (2011). Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology*, *102*(1), 108–134. https://doi.org/10.1348/000712610X506831

Luo, L., & Wildemuth, B. M. (2009). Semistructured interviews. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 232–241). Libraries Unlimited.

Makri, S., Blandford, A., Gow, J., Rimmer, J., Warwick, C., & Buchanan, G. (2007). A library or just another information resource? A case study of users' mental models of traditional and digital libraries. *Journal of the American Society for Information Science and Technology*, *58*(3), 433–445. https://doi.org/10.1002/asi.20510

Maravić, I. (2016a, February 25). Spotify's event delivery—The road to the cloud (Part I). *Spotify Labs*. https://labs.spotify.com/2016/02/25/spotifys-event-delivery-the-road-to-the-cloud-part-i/

Maravić, I. (2016b, March 3). Spotify's event delivery—The road to the cloud (Part II). *Spotify Labs*. https://labs.spotify.com/2016/03/03/spotifys-event-delivery-the-road-to-the-cloud-part-ii/

Maravić, I. (2016c, March 10). Spotify's event delivery—The road to the cloud (Part III). *Spotify Labs*. https://labs.spotify.com/2016/03/10/spotifys-event-delivery-the-road-to-the-cloud-part-iii/

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon.

McDonald, G. (2021). *Every noise at once*. https://everynoise.com

McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction*, *3*, 1–23. https://doi.org/10.1145/3359174

McFee, B., Bertin-Mahieux, T., Ellis, D., & Lanckriet, G. (2012). The million song dataset challenge. *International Conference on World Wide Web*, 909–916. https://doi.org/10.1145/2187980.2188222

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin Press.

McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., & Mehrotra, R. (2018). Explore, exploit, and explain: Personalizing explainable recommendations with bandits. *Proceedings of the 12th ACM Conference on Recommender Systems*, 31–39. https://doi.org/10.1145/3240323.3240354

Mehrotra, R., Bhattacharya, P., & Lalmas, M. (2020). Inferring the causal impact of new track releases on music recommendation platforms through counterfactual predictions. *Fourteenth ACM Conference on Recommender Systems*, 687–691. https://doi.org/10.1145/3383313.3418491

Mehrotra, R., Lalmas, M., Kenney, D., Lim-Meng, T., & Hashemian, G. (2019). Jointly leveraging intent and interaction signals to predict user satisfaction with slate recommendations. *The World Wide Web Conference*, 1256–1267. https://doi.org/10.1145/3308558.3313613

Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., & Diaz, F. (2018). Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. *Proceedings of the 27th ACM*

International Conference on Information and Knowledge Management, 2243–2251. https://doi.org/10.1145/3269206.3272027

Mehrotra, R., Shah, C., & Carterette, B. (2020). Investigating listeners' responses to divergent recommendations. *Fourteenth ACM Conference on Recommender Systems*, 692–696. https://doi.org/10.1145/3383313.3418482

Michell, G., & Dewdney, P. (1998). Mental models theory: Applications for library and information science. *Journal of Education for Library and Information Science*, *39*(4), 275–281. https://doi.org/10.2307/40324303

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Mittal, N., Nayak, R., Govil, M. C., & Jain, K. C. (2010). Recommender system framework using clustering and collaborative filtering. *3rd International Conference on Emerging Trends in Engineering and Technology*, 555–558. https://doi.org/10.1109/ICETET.2010.121

Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems*, *11*(3–4), 24:1-24:45. https://doi.org/10.1145/3387166

Molinder, N. (2018, January 30). Why building more rights databases won't solve the music industry metadata problem. *Hypebot*.

https://www.hypebot.com/hypebot/2018/01/why-building-more-rights-databases-wont-solve-the-music-industry-metadata-problem.html

Molnar, C. (2018). *Interpretable machine learning: A guide to making black box models explainable*. Leanpub. http://leanpub.com/interpretable-machine-learning

Moray, N. (1996). A taxonomy and theory of mental models. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *40*(4), 164–168. https://doi.org/10.1177/154193129604000404

Morgan, D. L. (2018). Themes, theories, and models. *Qualitative Health Research*, *28*(3), 339–345. https://doi.org/10.1177/1049732317750127

Morris, J. W. (2015). Curation by code: Infomediaries and the data mining of taste. *European Journal of Cultural Studies*, *18*(4–5), 446–463. https://doi.org/10.1177/1367549415577387

Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *ArXiv:1902.01876 [Cs]*. http://arxiv.org/abs/1902.01876

Mueller, S. T., Veinott, E. S., Hoffman, R. R., Klein, G., Alam, L., Mamun, T., & Clancey, W. J. (2021). *Principles of explanation in human-AI systems*. Explainable Agency in Artificial Intelligence Workshop. AAAI 2021. http://arxiv.org/abs/2102.04972

Naak, A., Hage, H., & Aïmeur, E. (2008). Papyres: A research paper management system. *2008 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, E-Commerce and E-Services*, 201–208. https://doi.org/10.1109/CECandEEE.2008.132

Nam, J., Choi, K., Lee, J., Chou, S.-Y., & Yang, Y.-H. (2019). Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from Bach. *IEEE Signal Processing Magazine*, *36*(1), 41–51. https://doi.org/10.1109/MSP.2018.2874383

National Research Council. Committee on Human Factors. (1987). *Mental models in human-computer interaction: Research issues about what the user of software knowsNRC*. National Academy Press.

Ngo, T., & Krämer, N. (2021). Exploring folk theories of algorithmic news curation for explainable design. *Behaviour & Information Technology*. https://doi.org/10.1080/0144929X.2021.1987522

Norman, D. A. (1983a). Design rules based on analyses of human error. *Communications of the ACM*, *26*(4), 254–258.

Norman, D. A. (1983b). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Psychology Press.

Norman, D. A. (2013). *The design of everyday things* (Rev. Ed.). Basic Books.

North, A. C., & Hargreaves, D. J. (1996). Situational influences on reported musical preference. *Psychomusicology*, *15*(1–2), 30–45. https://doi.org/10.1037/h0094081

Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, *27*(3), 393–444. https://doi.org/10.1007/s11257-017-9195-0

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Ontario. (2021). *Consultation: Ontario's trustworthy artificial intelligence (AI) framework*. http://www.ontario.ca/page/ontarios-trustworthy-artificial-intelligence-ai-framework-consultations

Oribhabor, I., Brody, J., Null, E., & Leufer, D. (2021, April 2). *Isedua Oribhabor (Access Now) to Daniel Elk (Spotify)*. https://www.accessnow.org/cms/assets/uploads/2021/04/Access-Now-letter-to-Spotify_2-April-2021-.pdf

Østerlund, C., Jarrahi, M. H., Willis, M., Boyd, K., & Wolf, C. T. (2021). Artificial intelligence and the world of work, a co-constitutive relationship. *Journal of the Association for Information Science and Technology*, *72*(1), 128–135. https://doi.org/10.1002/asi.24388

Padilla, T. (2019). *Responsible operations. Data science, machine learning, and AI in libraries*. OCLC Research. https://doi.org/10.25333/xk7z-9g97

Páez, A. (2019). The pragmatic turn in explainable artificial intelligence (XAI). *Minds and Machines*, *29*, 441–459. https://doi.org/10.1007/s11023-019-09502-w

Palacio, S., Lucieri, A., Munir, M., Hees, J., Ahmed, S., & Dengel, A. (2021). XAI handbook: Towards a unified framework for explainable AI. *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*. http://arxiv.org/abs/2105.06677

Park, M., Thom, J., Mennicken, S., Cramer, H., & Macy, M. (2020). Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, *3*, 230–236. https://doi.org/10.1038/s41562-018-0508-z

Park, O.-C., & Gittelman, S. (1995). Dynamic characteristics of mental models and dynamic visual displays. *Instructional Science*, *23*(5–6), 303–320. https://doi.org/10.1007/BF00896876

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Payne, S. J. (1991). A descriptive study of mental models. *Behaviour & Information Technology*, *10*(1), 3–21. https://doi.org/10.1080/01449299108924268

Payne, S. J. (2003). Users' mental models: The very ideas. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks toward a multidisciplinary science* (pp. 135–156). Morgan Kaufmann.

Pelly, L. (2017). The problem with Muzak: Spotify's bid to remodel an industry. *The Baffler*, *37*. https://thebaffler.com/salvos/the-problem-with-muzak-pelly

Pichl, M., Zangerle, E., & Specht, G. (2017). Understanding user-curated playlists on Spotify: A machine learning approach. *International Journal of Multimedia Data Engineering and Management*, *8*(4), 44–59. https://doi.org/10.4018/IJMDEM.2017100103

Pineau, J. (2017, October 6). *Reproducibility challenge*. http://www.cs.mcgill.ca/~jpineau/ICLR2018-ReproducibilityChallenge.html

Popper, B. (2015). Tastemaker: How Spotify's Discover Weekly cracked human curation at internet scale. *The Verge*. https://www.theverge.com/2015/9/30/9416579/spotify-discover-weekly-online-music-curation-interview

Porcaro, L., & Gomez, E. (2019). 20 years of playlists: A statistical analysis on popularity and diversity. *20th International Society for Music Informaiton Retrieval Conference*, 130–136. http://hdl.handle.net/10230/41906

Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, *97*, 205–227. https://doi.org/10.1016/j.eswa.2017.12.020

Prasad, G., Nie, Y., Bansal, M., Jia, R., Kiela, D., & Williams, A. (2020). To what extent do human explanations of model behavior align with actual model behavior? *ArXiv*. http://arxiv.org/abs/2012.13354

Prey, R., Valle, M. E. D., & Zwerwer, L. (2020). Platform pop: Disentangling Spotify's intermediary role in the music industry. *Information, Communication & Society*, 1–19. https://doi.org/10.1080/1369118X.2020.1761859

Price, B. (2002). Laddered questions and qualitative data research interviews. *Journal of Advanced Nursing*, *37*(3), 273–281. https://doi.org/10.1046/j.1365-2648.2002.02086.x

Proferes, N. (2017). Information flow solipsism in an exploratory study of beliefs about twitter. *Social Media + Society*, *3*(1), 1–17. https://doi.org/10.1177/2056305117698493

Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173–182. https://doi.org/10.1145/2702123.2702174

Rahma, A., Mardiatno, D., & Rahmawati Hizbaron, D. (2020). Q methodology to determine distinguishing and consensus factors (a case study of university students' ecoliteracy on disaster risk reduction). *Geosciences and Environmental Sciences Symposium (ICST 2020)*. https://doi.org/10.1051/e3sconf/202020001003

Ramon, Y., Vermeire, T., Toubia, O., Martens, D., & Evgeniou, T. (2021). Understanding consumer preferences for explanations generated by XAI algorithms. *ArXiv:2107.02624 [Cs]*. http://arxiv.org/abs/2107.02624

Rana, P., Jain, N., & Mittal, U. (2020). An introduction to basic concepts on recommender systems. In S. N. Mohanty, J. M. Chatterjee, S. Jain, & A. A. Elngar (Eds.), *Recommender systems with machine learning and artificial intelligence* (pp. 1–25). Scrivener Publishing. https://doi.org/10.1002/9781119711582.ch1

Raskin, J. D. (2016). Personal construct psychology in relation to an integrative constructivism. In D. A. Winter & N. Reed (Eds.), *The Wiley handbook of personal construct psychology* (pp. 34–44). John Wiley & Sons.

Ray, A., Cogswell, M., Lin, X., Alipour, K., Divakaran, A., Yao, Y., & Burachas, G. (2021). Generating and evaluating explanations of attended and error-inducing input regions for VQA models. *Applied AI Letters*. https://doi.org/10.1002/ail2.51

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 175–186. https://doi.org/10.1145/192844.192905

Reynolds, T. J., & Gutman, J. (1988). Laddering theory method, analysis, and interpretation. *Journal of Advertising Research*, *28*(1), 11–31.

Ribera, M., & Lapedriza, A. (2019). Can we do better explanations? A proposal of user-centered explainable AI. *Joint Proceedings of the ACM IUI 2019 Workshops*. IUI Workshop, New York ; http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-12.pdf

Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2015). *Recommender systems handbook* (2nd ed.). Springer. 10.1007/978-0-387-85820-3_13

Ridley, M. (2019). Explainable artificial intelligence. *Research Library Issues*, *299*, 28–46. https://doi.org/10.29242/rli.299.3

Ridley, M. (2022). Explainable AI (XAI): Adoption and advocacy. *Information Technology and Libraries*, *41*(2), 1–17. https://doi.org/doi.org/10.6017/ital.v41i2.14683

Riveiro, M., & Thill, S. (2021). “That’s (not) the output I expected!” On the role of end uer expectations in creating explanations of AI systems. *Artificial Intelligence*, *298*, 1–27. https://doi.org/10.1016/j.artint.2021.103507

Roe, K. (1985). Swedish youth and music: Listening patterns and motivations. *Communication Research*, *12*(3), 353–362. https://doi.org/10.1177/009365085012003007

Rogers, N. (2020, November 17). Why is an obscure b-side Pavement’s top song on Spotify? It’s complicated. *Stereogum*. https://www.stereogum.com/2105993/pavement-harness-your-hopes-spotify/columns/sounding-board/

Rossi, F. (2016). *Artificial intelligence: Potential benefits and ethical challenges* (PE 571.380). European Parliament. Policy Department C: Citizen's Rights and Constitutional Affairs. http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_EN.pdf

Rouse, W. B., & Morris, J. R. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*, 349–363.

Ruckenstein, M., & Granroth, J. (2020). Algorithms, advertising and the intimacy of surveillance. *Journal of Cultural Economy*, *13*(1), 12–24. https://doi.org/10.1080/17530350.2019.1574866

Rüping, S. (2006). *Learning interpretable models* [PhD Dissertation, University of Dortmund]. http://www.stefan-rueping.de/publications/rueping_2006b.pdf

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods*, *15*(1), 85–109. https://doi.org/10.1177/1525822X02239569

Saarikallio, S., & Erkkilä, J. (2007). The role of music in adolescents' mood regulation. *Psychology of Music*, *35*(1), 88–109. https://doi.org/10.1177/0305735607068889

Saeed, W., & Omlin, C. (2021). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. *ArXiv*. http://arxiv.org/abs/2111.06420

Saffer, D. (2010). *Designing for interaction: Creating innovative applications and devices* (2nd ed.). New Riders.

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications.

Samek, W., & Muller, K.-R. (2019). Towards explainable artificial intelligence. In W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, & K.-R. Muller (Eds.), *Explainable AI: Interpreting, explaining and visualizing deep learning* (pp. 5–22). Springer International Publishing.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing algorithms: Research methods for detecting discrimination on internet platforms*. Annual Meeting of the International Communication Association, Seattle, WA. http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf

Sasse, M. A. (1991). How to t(r)ap users' mental models. In M. J. Tauber & D. Ackermann (Eds.), *Human Factors in Information Technology* (Vol. 2, pp. 59–79). North-Holland. https://doi.org/10.1016/B978-0-444-88602-6.50007-4

Sasse, M. A. (1997). *Eliciting and describing users' models of computer systems* [PhD Dissertation, University of Birmingham]. https://etheses.bham.ac.uk/id/eprint/7356/

Savolainen, L. (2022). The shadow banning controversy: Perceived governance and algorithmic folklore. *Media, Culture & Society*. https://doi.org/10.1177/01634437221077174

Schmettow, M., & Sommer, J. (2016). Linking card sorting to browsing performance—Are congruent municipal websites more efficient to use? *Behaviour & Information Technology*, *35*(6), 452–470. https://doi.org/10.1080/0144929X.2016.1157207

Schneider, J., Handali, J., Vlachos, M., & Meske, C. (2020). Deceptive AI explanations: Creation and detection. *ArXiv:2001.07641 [Cs, Stat]*. http://arxiv.org/abs/2001.07641

Schoeffer, J., De-Arteaga, M., & Kuehl, N. (2022). *On the relationship between explanations, fairness perceptions, and decisions*. ACM CHI 2022 Workshop on Human-Centered Explainable AI (HCXAI). http://arxiv.org/abs/2204.13156

Schrills, T., & Franke, T. (2020). How to answer why—Evaluating the explanations of AI through mental model analysis. *ArXiv*. http://arxiv.org/abs/2002.02526

Schwartz, E. H. (2021, May 7). Musicians demand Spotify not develop emotional speech recognition patent. *Voicebot.Ai*. https://voicebot.ai/2021/05/07/musicians-demand-spotify-not-develop-emotional-speech-recognition-patent/

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Proceedings of the 28th International Conference on Neural Information Processing Systems, V. 2*, 2503–2511.

Seadle, M. (2003). Mental models for personal digital assistants (PDAs). *Library Hi Tech*, *21*(4), 390–392. https://doi.org/10.1108/07378830310516215

Seaver, N. (2019). Captivating algorithms: Recommender systems as traps. *Journal of Material Culture*, *24*(4), 421–436. https://doi.org/10.1177/1359183518820366

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review*, *87*(3). https://ir.lawnet.fordham.edu/flr/vol87/iss3/11

Sheh, R., & Monteath, I. (2018). Defining explainable AI for requirements analysis. *KI - Kunstliche Intelligenz*, *32*(4), 261–266. https://doi.org/10.1007/s13218-018-0559-3

Shen, H., & Huang, T.-H. (2021). *Explaining the road not taken*. ACM CHI Workshop of Operationalizing Human-Centered Perspectives in Explainable AI. http://arxiv.org/abs/2103.14973

Shneiderman, B. (2022). *Human-centered AI*. Oxford University Press.

Siles, I., Segura-Castillo, A., Solís, R., & Sancho, M. (2020). Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South: *Big Data & Society*. https://doi.org/10.1177/2053951720923377

Simon, H. A. (1992). What is an “explanation” of behavior? *Psychological Science*, *3*(3), 150–161. https://doi.org/10.1111/j.1467-9280.1992.tb00017.x

Simonite, T. (2018, December 12). Google's AI guru wants computers to think more like brains. *Wired*. https://www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/

Singh, D., Slupczynski, M., Pillai, A. G., & Pandian, V. (2022). *Grounding explainability within the context of global south in XAI*. ACM CHI 2022 Workshop on Human-Centered Explainable AI (HCXAI). http://arxiv.org/abs/2205.00931

Singh, R., Dourish, P., Howe, P., Miller, T., Sonenberg, L., Velloso, E., & Vetere, F. (2021). Directive explanations for actionable explainability in machine learning applications. *ArXiv*. http://arxiv.org/abs/2102.02671

Singhal, A., Sinha, P., & Pant, R. (2017). Use of deep learning in modern recommendation system: A summary of recent works. *International Journal of Computer Applications, 180*(7), 17–22.

Slack, D., Hilgard, S., Singh, S., & Lakkaraju, H. (2021). Feature attributions and counterfactual explanations can be manipulated. *ArXiv*. http://arxiv.org/abs/2106.12563

Slaney, M., & White, W. (2006). Measuring playlist diversity for recommendation systems. *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, 77–82. https://doi.org/10.1145/1178723.1178735

Slattery, P. (2019). The cheater's guide to Spotify. *OneZero*. https://onezero.medium.com/cheaters-guide-to-spotify/home

Smith, J., Sonboli, N., Fiesler, C., & Burke, R. (2020). Exploring user opinions of fairness in recommender systems. *ArXiv*. http://arxiv.org/abs/2003.06461

Smith, L. C. (1976). Artificial intelligence in information retrieval systems. *Information Processing and Management*, *12*(3), 189–222. https://doi.org/10.1016/0306-4573(76)90005-4

Smith, L. C. (1989). Artificial intelligence: Relationships to research in library and information science. *Journal of Education for Library and Information Science*, *30*(1), 55–56. https://doi.org/10.2307/40323500

Smith-Renner, A., Fan, R., Birchfield, M., Wu, T., Boyd-Graber, J., Weld, D. S., & Findlater, L. (2020). No explainability without accountability: An empirical study of explanations and feedback in interactive ML. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. http://doi.org/10.1145/3313831.3376624

Snead, K. C., Magal, S. R., Christensen, L. F., & Ndede-Amadi, A. A. (2015). Attribution theory: A theoretical framework for understanding information systems success. *Systemic Practice and Action Research*, *28*(3), 273–288. https://doi.org/10.1007/s11213-014-9328-x

Snickars, P. (2017). More of the same: On Spotify radio. *Culture Unbound*, *9*(2), 184–211.

Söderström, G. (Host). (2021a). *Spotify: A product story [Audio podcast]*. https://open.spotify.com/show/3L9tzrt0CthF6hNkxYIeSB

Söderström, G. (Host). (2021b, March 31). *Human vs machine [Audio podcast episode]. In Spotify: A product story* [No. 4]. https://open.spotify.com/episode/0T3nb0PcpvqA4o1BbbQWpp

Sokol, K., & Flach, P. (2020). One explanation does not fit all: The promise of interactive explanations for machine learning transparency. *Künstliche Intelligenz*, *34*(2), 235–250. https://doi.org/doi.org/10.1007/s13218-020-00637-y

Sonboli, N., Smith, J. J., Berenfus, F. C., Burke, R., & Fiesler, C. (2021). *Fairness and transparency in recommendation: The users' perspective*. UMAP '21, Utrecht, Netherlands. https://doi.org/10.1145/3450613.3456835

Spotify. (2018). *Prospectus. Form F-1 registration statement (IPO)*. United State Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/1639920/000119312518063434/d494294df1.htm

Spotify. (2020a). *Company info*. For the Record. https://newsroom.spotify.com/company-info/

Spotify. (2020b). *Privacy policy*. https://www.spotify.com/ca-en/legal/privacy-policy/

Spotify. (2020c, April 15). Reach for the top: How Spotify built Shortcuts in just six months. *Spotify Labs*. https://labs.spotify.com/2020/04/15/reach-for-the-top-how-spotify-built-shortcuts-in-just-six-months/

Spotify. (2021a). *Annual report*. https://s22.q4cdn.com/540910603/files/doc_financials/2021/q4/0307a021-254e-43c5-aeac-8242b0ea3ade.pdf

Spotify. (2021b). *Loud & clear*. https://loudandclear.byspotify.com/

Spotify. (2021c). *Spotify [Mobile application software]*. https://www.spotify.com/uk/download

Spotify (Director). (2021d, February 22). *Spotify stream on [Video]*. YouTube. https://youtu.be/Vvo-2MrSgFE

Spradley, J. P. (1979). *The Ethnographic interview*. Holt, Rinehart and Winston.

Staggers, N., & Norcio, A. F. (1993). Mental models: Concepts for human-computer interaction research. *International Journal of Man-Machine Studies*, *38*(4), 587–605. https://doi.org/10.1006/imms.1993.1028

Stål, O. (2021). *How Spotify uses ML to create the future of personalization*. TransformX. https://youtu.be/n16LOyba-SE

Stark, L., & Hoey, J. (2021). The ethics of emotion in artificial intelligence systems. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 782–793. https://doi.org/10.1145/3442188.3445939

Stebbins, R. A. (2009). Leisure and its relationship to library and: Information science: Bridging the gap. *Library Trends*, *57*(4), 618–631.

Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., & Basilico, J. (2021). Deep learning for recommender systems: A Netflix case study. *AI Magazine*, *42*(3), 7–18. https://doi.org/10.1609/aimag.v42i3.18140

Stephenson, W. (1954). *The study of behavior: A technique and its methology*. University of Chicago Press.

Stoeckli, E., Dremel, C., Uebernickel, F., & Brenner, W. (2020). How affordances of chatbots cross the chasm between social and traditional enterprise systems. *Electronic Markets*, *30*(2), 369–403. https://doi.org/10.1007/s12525-019-00359-6

Stop Spotify Surveillance. (2021, May 4). *Coalition letter to Spotify*. https://www.stopspotifysurveillance.org/letter.pdf

Stoyanovich, J., Van Bavel, J. J., & West, T. V. (2020). The imperative of interpretable machines. *Nature Machine Intelligence*, *2*(April), 197–199. https://doi.org/10.1038/s42256-020-0171-8

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Sun, H. (2018). *Digital revolution tamed: The case of the recording industry*. Springer.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

Swartout, W. (1983). XPLAIN: A system for creating and explaining expert consulting programs. *Artificial Intelligence*, *21*, 285–325.

Swartout, W., Paris, C., & Moore, J. (1991). Design for explainable expert systems. *IEEE Expert-Intelligent Systems & Their Applications*, *6*(3), 58–64. https://doi.org/10.1109/64.87686

Sweeney, L. (2018). *How to save democracy and the world*. Association of Computing Machinery (ACM) Conference on Fairness, Accountability, and Transparency, New York.

Talja, S., Tuominen, K., & Savolainen, R. (2005). "Isms" in information science: Constructivism, collectivism and constructionism. *Journal of Documentation*, *61*(1), 79–101. https://doi.org/10.1108/00220410510578023

Tarus, J., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: A review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, *50*(1), 21–48. https://doi.org/10.1007/s10462-017-9539-5

Taylor, J. E. T., & Taylor, G. W. (2021). Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychonomic Bulletin & Review*, *28*, 454–475. https://doi.org/10.3758/s13423-020-01825-5

Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig & W. Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 17–36). SAGE Publications. 10.4135/9781526405555.n2

Thoma, M., Ryf, S., Mohiyeddini, C., Ehlert, U., & Nater, U. M. (2012). Emotion regulation through listening to music in everyday situations. *Cognition and Emotion*, *26*(3), 550–560. https://doi.org/10.1080/02699931.2011.595390

Thomas, P., Billerbeck, B., Craswell, N., & White, R. W. (2019). Investigating searchers' mental models to inform search explanations. *ACM Transactions on Information Systems*, *38*(1), 10:1-10:25. https://doi.org/10.1145/3371390

Toff, B., & Nielsen, R. K. (2018). "I Just Google It": Folk theories of distributed discovery. *Journal of Communication*, *68*(3), 636–657. https://doi.org/10.1093/joc/jqy009

Tullio, J., Dey, A., Chalecki, J., & Fogarty, J. (2007). *How it works: A field study of non-technical users interacting with an intelligent system*. 31–40. https://doi.org/10.1145/1240624.1240630

Turek, M. (2016). *Explainable artificial intelligence (XAI)*. DARPA. https://www.darpa.mil/program/explainable-artificial-intelligence

Tuzhilin, A., & Koren, Y. (Eds.). (2008). *Proceedings of the 2nd KDD workshop on large-scale recommender systems and the Netflix prize competition*. Association for Computing Machinery.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302. https://doi.org/10.1109/TSA.2002.800560

van den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. *Proceeding of the 26th International Conference on Neural Information Processing Systems*, 2643–2651.

VanScoy, A. (2021). Using Q methodology to understand conflicting conceptualizations of reference and information service. *Library & Information Science Research*, *43*(3). https://doi.org/10.1016/j.lisr.2021.101107

Vaughan, J. W., & Wallach, H. (2021). A human-centered agenda for intelligent machine learning. In M. Pelillo & T. Scantabmurlo (Eds.), *Machines we trust: Perspectives on dependable AI* (pp. 123–138). MIT Press.

Verma, S., Lahiri, A., Dickerson, J. P., & Lee, S.-I. (2021). *Pitfalls of explainable ML: An industry perspective*. JOURNE Workshop. Fifth Conference on Machine Learning and Systems. http://arxiv.org/abs/2106.07758

Vermeire, T., Laugel, T., Renard, X., Martens, D., & Detyniecki, M. (2021). How to choose an explainability method? Towards a methodical implementation of XAI in practice. *ArXiv*. http://arxiv.org/abs/2107.04427

Vermorel, J., & Mohri, M. (2005). Multi-armed bandit algorithms and empirical evaluation. In J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge, & L. Torgo (Eds.), *European Conference on Machine Learning* (pp. 437–448). Springer Berlin Heidelberg.

Villareale, J., & Zhu, J. (2021). Understanding mental models of AI through player-AI interaction. *HCXAI '21: ACM CHI Workshop on Human-Centered Perspectives in Explainable AI*. http://arxiv.org/abs/2103.16168

Villermet, Q., Poiroux, J., Moussallam, M., Louail, T., & Roth, C. (2021). Follow the guides: Disentangling human and algorithmic curation in online music consumption. *Fifteenth ACM Conference on Recommender Systems*, 380–389. https://doi.org/10.1145/3460231.3474269

Vilone, G., & Longo, L. (2020). Explainable artificial Intelligence: A systematic review. *ArXiv*. http://arxiv.org/abs/2006.00093

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, *31*(2), 842–887.

Wallace, N. (2017, January 25). *EU's right to explanation: A harmful restriction on artificial intelligence*. TechZone. http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm#

Wallace, N., & Castro, D. (2018). *The impact of the EU's new data protection regulation on AI*. Centre for Data Innovation. http://www2.datainnovation.org/2018-impact-gdpr-ai.pdf

Waller, I., & Anderson, A. (2019). Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. *The World Wide*

Web Conference on - WWW '19, 1954–1964. https://doi.org/10.1145/3308558.3313729

Walmsley, J. (2021). Artificial intelligence and the value of transparency. *AI & Society, 36*(2), 585–595. https://doi.org/10.1007/s00146-020-01066-z

Wang, C., & Blei, D. (2011). *Collaborative topic modeling for recommending scientific articles*. 448–456. https://doi.org/10.1145/2020408.2020480

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300831

Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machines*, *29*, 417–440. https://doi.org/doi.org/10.1007/s11023-019-09506-6

Watson, J. E. (2019, October 1). Reflecting on Spotify's recommender system. *SongData*. https://songdata.ca/2019/10/01/reflecting-on-spotifys-recommender-system/

Watts, S., & Stenner, P. (2012). *Doing Q methodology research: Theory, method and interpretation*. SAGE Publications.

Weber, L., Lapuschkin, S., Binder, A., & Samek, W. (2022). Beyond explaining: Opportunities and challenges of XAI-based model improvement. *ArXiv*. http://arxiv.org/abs/2203.08008

Webster, J., Gibbins, N., Halford, S., & Hracs, B. J. (2016). Towards a theoretical approach for analysing music recommender systems as sociotechnical cultural intermediaries. *Proceedings of the 8th ACM Conference on Web Science*, 137–145. https://doi.org/10.1145/2908131.2908148

Werner, A. (2020). Organizing music, organizing gender: Algorithmic culture and Spotify recommendations. *Popular Communication*, *18*(1), 78–90. https://doi.org/10.1080/15405702.2020.1715980

Whitman, B. (2012, December 11). How music recommendation works—And doesn't work. *Variogram*. https://notes.variogr.am/2012/12/11/how-music-recommendation-works-and-doesnt-work/

Wolf, C., & Ringland, K. (2019). Designing accessible, explainable AI (XAI) experiences. *ACM SIGACCESS: Accessibility and Computing*, *125*. https://doi.org/10.1145/3386296.3386302

Woźnica, K., Pękala, K., Baniecki, H., Kretowicz, W., Sienkiewicz, E., & Biecek, P. (2021). Do not explain without context: Addressing the blind spot of model explanations. *ArXiv*. http://arxiv.org/abs/2105.13787

Algorithmic accountability act, United State Senate (2022) (testimony of Ron Wyden). https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202022%20Bill%20Text.pdf

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use characteristics, and impact. *MIS Quarterly*, *31*(1), 137–209.

Xie, B., Zhou, J., & Wang, H. (2017). How influential are mental models on interaction performance? Exploring the gap between users' and designers' mental models through a new quantitative method. *Advances in Human-Computer Interaction*, *2017*, 1–14. https://doi.org/10.1155/2017/3683546

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32*(4), 403–414. https://doi.org/10.1002/bdm.2118

Young, I. (2008). *Mental models: Aligning design strategies with human behavior*. Rosenfield Media.

Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society*, *43*(5), 807–824. https://doi.org/10.1177/0163443720972314

YuMin, S., Yuan, Z., & JinYao, Y. (2018). Neural network based movie rating prediction. *Proceedings of the 2018 International Conference on Big Data and Computing*, 33–37. https://doi.org/10.1145/3220199.3220204

Zehlike, M., Yang, K., & Stoyanovich, J. (2022a). Fairness in ranking, Part I: Score-based ranking. *ACM Computing Surveys*. https://doi.org/10.1145/3533379

Zehlike, M., Yang, K., & Stoyanovich, J. (2022b). Fairness in ranking, Part II: Learning-to-rank and recommender systems. *ACM Computing Surveys*. https://doi.org/10.1145/3533380

Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). Transparency in algorithmic and human decision-making: Is there a double standard? *Philosophy & Technology*, *32*(4), 661–683. https://doi.org/10.1007/s13347-018-0330-6

Zhang, B., Kreitz, G., Isaksson, M., Ubillos, J., Urdaneta, G., Pouwelse, J. A., & Epema, D. (2013). Understanding user behavior in Spotify. *INFOCOM*, 220–224. https://doi.org/10.1109/INFCOM.2013.6566767

Zuboff, S. (2018). *The age of surveillance capitalism*. PublicAffairs.

Appendices

Appendix 1: Principles of HCXAI (Mueller et al. 2021)

The property of being an explanation is not a property of statements, visualizations, or examples. Explaining is a process by which an explainee and explainer achieve common ground. XAI Designers should recognize that an explanation is not merely an artifact delivered from an algorithm, but must be understood by a user to be effective.

Work matters. It is impractical to develop a useful and usable explanation system outside of a work context. Explanation relates the tool to the user knowledge in the context of the goals and tasks, and whether a particular algorithm or visualization will support that work cannot be known in the abstract. Human-centered design principles suggest involving the user early and throughout the system development process (Greenbaum and Kyng 1991; Deal and Hoffman 2010; Hoffman et al. 2010; Clancey 2020).

The importance of active self-explanation. The "spoon feeding" paradigm is oblivious to the fact that irrespective of whatever material people receive by way of explanation, they still engage in a motivated attempt to make sense of the AI and the explanatory material. Developers should recognize this and focus explanatory systems on information that empowers users to self-explain, rather than simply delivering an output of an algorithms that is intended to serve as explanatory.

Build explanatory systems, not explanations. Rarely does the initial explanation coming directly from the AI algorithms provide a useful explanation, let alone an ideal one. It must be accompanied by other things (instructions, tutorial activities, comparisons, exploratory interfaces, user models, etc.) to succeed.

Combined methods are necessary. Much work on XAI involves testing a particular concept or algorithm in isolation. But when designing an explanatory system, multiple kinds of information can complement one another. For example, both global and local explanations may be justifiable and reinforce one another. Showing examples that establish a pattern may play a different role than contrasting examples which establish a critical causal relationship. Using examples of related heatmaps may be more powerful than either examples or heatmaps in isolation. And it is crucial to keep in mind that actual work contexts involve multiple systems, so the user is continuously challenged with their own "system integration" task.

An explanation can have many different consequences. Often, developers create and test explanations to determine whether they work. However, different explanations can have very different effects. This includes differing effects on qualitative assessments (satisfaction, trust), versus knowledge measures and performance measures. The explanations should be tuned to the goal, keeping in mind the fact that people may be led to trust and rely upon an AI system simply being given more and more information about it, whether or not that information leads to better or deeper understanding.

Measurement matters. Because explanations can have their impact on a number of ways, and so can be assessed along many dimensions (goodness, satisfaction and trust, knowledge/understanding, and performance). Designers should identify what consequences the explanation should have in order to develop an appropriate measurement and assessment approach.

Knowledge and understanding are central. Much of the research on XAI focuses on algorithmic visualizations, which distracts from the fact that the focus of explanation is on developing a better understanding of the system. Understanding leads to appropriate trust and appropriate reliance, and therefore overall better performance with the system.

Context matters: Users, timing, goals. An explanation is not a beacon revealing the truth. The best explanation depends on context: who the user is, what their goal is, when they need an explanation, and how its effectiveness is measured. Developers should consider use cases, user models, timeliness, and attention and distraction limitations for their explanations.

The power of differences and contrast. A central lesson of XAI is the utility of contrast, comparison, and counterfactuals in understanding the boundary conditions of a system. A useful exercise is to first develop learning objectives for an explanatory system, and identify the contrasts necessary to support those objectives.

Explanation is not just about transparency. If something is transparent, you cannot see it. This word is widely misused. What is needed are systems whose workings are *apparent*, that is, readily understood and not hidden (the "black box metaphor"). Especially in the context of fairness in algorithmic decision making, many have advocated “transparency” as an approach to explanation. This can never be enough, because a user may still not understand how a system works even if its algorithms are somehow apparent---observable and visible. Other methods (contrast, global explanation and local justification, examples, explorable interfaces that permit hypotheticals, etc.) will be necessary in most situations to harness apparency and develop understanding. In "real world" work contexts, people always feel some mixture of justified trust, unjustified trust and justified and unjustified mistrust. These attitudes are in constant flux and rarely develop in a smooth progression toward some ideal and stable point. Trust can come and go in a flash. When the AI fails in a way that a human would never fail, reliance can collapse.

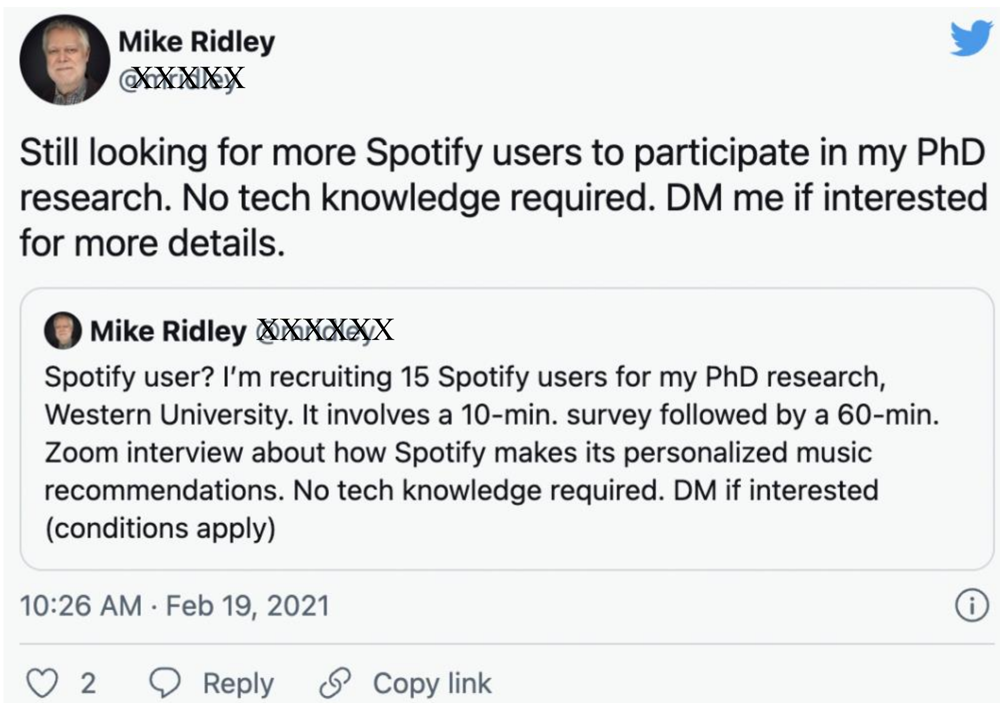
The need for explanation is “triggered.” Too often, XAI systems deliver an explanation regardless of whether one is needed. However, explanations are not always necessary. In normal human reasoning, explanation is triggered by states such as surprise and violations of expectation. Advances in XAI will come when systems begin to understand situations that are likely to engender surprise and violate user expectations.

Explanation is knowledge transformation and sensemaking. The achievement of an understanding is not just the learning or incorporation of information; it involves changing previous beliefs and preconceptions. The acquisition of knowledge involves both assimilation and accommodation, to use Piagetian terminology. The power of explanation is that it can activate fast “System 2” learning modes that quickly reconfigure knowledge with minimal feedback, bypassing the slower “System 1” trial-and-error feedback-based learning often used to understand a system.[4] XAI systems should harness this by attempting to identify the user’s current understanding (so that it can better predict how to transform this knowledge), and support the information that will help make these transformations.

Explanation is never a “one-off.” Especially for AI systems that learn or are applied in dynamic contexts, users often need repeated explanations and re-explanations. How has the algorithm changed? Are these new data valid? XAI systems might benefit from considering the long-term interaction with users, even in simple ways like recognizing that once learned, an explanation may not need to be given again unless something has changed.

[4] Mueller et al. have System 1 and System 2 backwards; System 1 = “fast”; System 2 = “slow.”

Appendix 2: Spotify User Recruitment Tweets



Appendix 3: Spotify User Survey and Interview Letter of Information and Consent

Western FIMS
Faculty of Information & Media Studies

Library & Information Science

Letter of Information and Consent

STUDY: User Theories of the Spotify Music Streaming and Recommender System

STUDENT RESEARCHER: Michael Ridley, Ph.D. Candidate

PRINCIPAL INVESTIGATOR: Dr. Jacquelyn Burkell

Faculty of Information and Media Studies
Western University

PURPOSE OF THE STUDY

The study is to determine how users of Spotify, a music streaming and recommendation system, think the system makes its personalized recommendations.

INCLUSION CRITERIA

You are eligible to participate in this study if you meet the following criteria: You use the Spotify music recommendation system, reside in Canada or the United States, speak English, and are 18 years or older.

PROCEDURES

You are asked to participate in a survey and a follow-up interview using Zoom (or another communications system). The survey should take approximately 5-10 minutes. The follow-up interview should take approximately 60 minutes. Should you choose to participate you may choose to abstain from any questions you wish not to answer. During the survey and interview you will be asked to answer a series of questions regarding your use of Spotify and how you think it makes its recommendations.

POTENTIAL RISKS AND DISCOMFORTS

Participants will be asked to answer based on their own experiences and views on decisions made about their person and may experience some discomfort. You are not required to answer and questions for any reason, including discomfort.

POTENTIAL BENEFITS TO PARTICIPANTS AND/OR TO SOCIETY

Participants may gain insight into what information is used by recommender systems. We will gain insight into how people think recommender systems work. This information could be useful when helping to explain recommender systems for the purposes of transparency, accountability, and trust.

COMPENSATION FOR PARTICIPATION

You will be compensated $25.

CONFIDENTIALITY

Some study data may be made available to journals and/or other researchers. Only your anonymized data will be shared. Any other information that is obtained in connection with this study, and that can be identified with you, will remain confidential and will be disclosed only with your permission.

Only the student researcher, Michael Ridley, and the thesis supervisor, Jacquelyn Burkell will have access to the survey data.

The survey and interview data and other study data will not have any information identifying you and your name will not be used anywhere in the study or documents resulting from the study. All identifiers will be removed from study data.

The interview will be audio recorded. Only the student researcher, Michael Ridley, the thesis supervisor, Jacquelyn Burkell, and Transcript Heroes (a transcribing service) will have access to the audio recording. Once the interview has been transcribed the audio recording will be destroyed.

The transcription of the interview discussion and other study data will not have any information identifying you and your name will not be used anywhere in the study or documents resulting from the study. All identifiers will be removed from study data. A study number will be assigned to each participant and linked via a master list to your name and any other identifying information. The master list will be kept and stored separate from the other study data.

Michael Ridley and Jacquelyn Burkell will have access to all of the research data. Study data will be stored on the Western University server and may also be stored on encrypted, password protected devices.

No one else will have access to research data except for representatives of The University of Western Ontario Non-Medical Research Ethics Board who may require access to your study-related records to monitor the conduct of the research. The data will be destroyed after 7 years.

PARTICIPATION AND WITHDRAWAL

Your participation in this study is voluntary. You may decide not to be in this study. Even if you consent to participate you have the right to not answer individual questions or to withdraw from the study at any time. We will provide you any new information that may affect your decision to stay in the study.

You do not waive any legal right by acknowledging this consent form.

You may withdraw from the study at any time until findings have been submitted for publication. Any related data that has not been anonymized and aggregated will be removed and destroyed.

De-identified personal quotes may be used in the findings of this study. Please indicate in the following question if you consent to the use of your de-identified quotes.

SUBSEQUENT USE OF DATA

Data may be used in subsequent studies, in publications, and in presentations.

WHOM TO CONTACT FOR QUESTIONS?

Michael Ridley, Student Researcher

Dr. Jacquelyn Burkell, Principal Investigator

If you have any questions about your rights as a research participant or the conduct of this study, you may contact The Office of Human Research Ethics.

Western University, Faculty of Information and Media Studies,

With this link you can download a PDF copy of this Letter of Information.

Submission of this survey serves as consent to participate in this research (both this survey and the follow-up interview).

Appendix 4: REB Approval - Folk Theories of the Spotify Recommender System

Date: 9 February 2021

To: Dr. Jacquelyn Burkell

Project ID: 116652

Study Title: Folk Theories of Spotify's Recommender System

Short Title: Folk Theories of the Spotify Recommender System

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: March 5 2021

Date Approval Issued: 09/Feb/2021

REB Approval Expiry Date: 09/Feb/2022

Dear Dr. Jacquelyn Burkell

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. **All other required institutional approvals and mandated training must also be obtained prior to the conduct of the study.**

Documents Approved:

Document Name	Document Type	Document Date	Document Version
Spotify Users Debrief Nov 1 2020	Debriefing document	01/Nov/2020	1
Spotify Users Twitter Recruitment Nov 1 2020	Recruitment Materials	01/Nov/2020	1
Spotify Users Interview Guide Nov 1 2020	Interview Guide	01/Nov/2020	1
User Survey-Interview LOI-C Jan 15 2021	Implied Consent/Assent	15/Jan/2021	1
User Survey LOI-C Jan 15 2021	Online Survey	15/Jan/2021	2

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Kelly Patterson , Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix 5: Spotify User Survey

Spotify User Survey

Interest. Which of the following best describes your interest in music?

○ Passionate about music with extensive knowledge (1)

○ Keen about music but balanced with other interests (2)

○ Music is important but other things are far more important (3)

○ Engage with music but are generally indifferent (4)

Listener .Type Would you describe yourself as a “specialist” (listens to mostly the same artists and genres) or a “generalist” (listens to a wide variety of artists and genres)?

○ Specialist (1)

○ Generalist (2)

Version. Do you subscribe to Spotify (pay version) or use the free (ad-supported) version?

○ Paid (subscription) version (1)

○ Free (ad-supported) version (2)

Length. How long have you been using Spotify?

◯ Less than 1 year (1)

◯ 1 to 5 years (2)

◯ More than 5 years (3)

Often. How often do you listen to Spotify?

◯ Every day (1)

◯ Most days (2)

◯ At least weekly (3)

◯ Less often than weekly (4)

Mode. How do you primarily listen to Spotify?

◯ On a laptop or desktop computer? (1)

◯ On a smartphone or mobile device? (2)

◯ On a smart assistant (e.g., Alexa, Google Home)? (3)

◯ Other (4) ____

Satisfied. Are you generally satisfied with Spotify's personalized music recommendations to you?

◯ Yes (1)

◯ No (2)

How Made. How do you think Spotify's personalized music recommendations are made?

- ○ Solely by algorithms (1)
- ○ Primarily by algorithms and partly by humans (2)
- ○ Primarily by humans and partly by algorithms (3)
- ○ Solely by humans (4)
- ○ Don't Know (5)

Recos Made. How does Spotify use information to determine the personalized music recommendations for you?

____

____

____

____

____

Recos Influence. What could you do to shape the personalized music recommendations you receive from Spotify?

____

____

____

____

____

Factors. To what extent do you think the following influence Spotify's music recommendations for you?

	Very Important (1)	Important (2)	Somewhat Important (3)	Not Important (4)
Marking something a "like" (i.e., "heart") (1)				
What I listen to (2)				
How long I listen to a song or playlist (3)				
How many times I listen to a song, artist or playlist (4)				
What other people are listening to (5)				
What my friends are listening to (6)				
Songs that are similar to other songs I “liked” or listened to (7)				

Playlists I've created (8)	○	○	○	○
Playlists other users have created (9)	○	○	○	○
What people my age listen to (10)	○	○	○	○
What people in my location (city/country) listen to (11)	○	○	○	○
What people with my level of education listen to (12)	○	○	○	○
Where I am while listening (13)	○	○	○	○
What I'm doing while listening (14)	○	○	○	○
What I'm feeling while listening (15)	○	○	○	○

The time of day I'm listening (16)	○	○	○	○
The day of the week I'm listening (17)	○	○	○	○
The season of the year I'm listening (18)	○	○	○	○
Songs or artists that Spotify is promoting (19)	○	○	○	○
Posts about Spotify I make to Facebook or other social media (20)	○	○	○	○
Comments from other people about music on Facebook or other social media (21)	○	○	○	○

Reviews of music in magazines, blogs, videos, news sources (22)	○	○	○	○

Thank you for completing this survey. If we have not already scheduled a follow-up interview with you, we will do so shortly.

The next section will provide a summary of your responses for future reference.

Please download the PDF version of this or make a copy for subsequent reference and use.

Appendix 6: Factor Loadings

Part. No.	Q-sort	Factor Group	Factor 1	F1	Factor 2	F2	Factor 3	F3	Factor 4	F4	Factor 5	F5	Factor 6	F6
15	User16	F1-1	0.8061	*	0.1006		-0.0953		0.1247		0.0638		0.363	
5	User6	F1-2	0.6995	*	0.185		0.0883		0.4201		-0.0871		0.2453	
2	User3	F1-3	0.6741	*	0.6219		0.024		0.1243		0.1606		0.0959	
10	User11	F1-4	0.6015	*	0.3365		0.1528		0.2078		-0.0061		0.2662	
12	User13	F1-5	0.5997	*	0.3621		-0.0274		0.3568		0.0052		0.4014	
6	User7	F1-6	0.5448	*	0.2849		0.0938		0.388		-0.0428		0.3349	
13	User14	F1-7	0.5389	*	0.1497		0.0565		0.4095		-0.0905		0.0378	
14	User15	F2-1	0.2092		0.8469	*	-0.0535		0.1322		-0.0751		0.3234	
11	User12	F2-2	0.2		0.5653	*	0.0673		0.1545		-0.0022		0.2663	
3	User4	F4-1	0.1849		0.0598		-0.0712		0.7857	*	0.0393		0.0735	
16	User17	F4-2	0.2602		0.1268		0.0921		0.5481	*	-0.0317		0.4105	
7	User8	F4-3	0.3229		0.2448		0.116		0.5164	*	-0.0283		0.2386	
1	User2	F6-1	0.1716		0.16		-0.0258		0.1197		0.0246		0.6758	*
19	User20	F6-2	0.4981		0.1662		0.062		0.4133		-0.0755		0.6344	*
18	User19	F6-3	0.4265		0.4473		0.1043		0.2799		0.0114		0.6321	*
8	User9	F6-4	0.1055		0.1731		0.0094		0.0281		-0.0187		0.6063	*
4	User5	F6-5	0.3306		0.3735		0.1416		0.3994		-0.0177		0.5766	*
9	User10	F7-1	0.0559		0.2356		-0.0251		0.2125		-0.0084		0.2304	
17	User18	F7-2	0.3894		0.4283		0.1725		0.4663		0.0309		0.2423	

Appendix 7: Spotify User Interview Guide

Spotify Users Interview Guide

Introduction:

As a result of the online survey that precedes the interview, participants will have read the Letter of Information and indicated consent.

> Welcome and thank the participant.
>
> Explain the purpose of the interview: to explore how users of the Spotify music streaming service think the system makes personalized music recommendations.
>
> Explain why they were selected: because they are Spotify users.
>
> Ask if there are any questions arising from the survey or the Letter of Information.

Discussion Guide:

The questions are intended as a guide. The interviewer may alter or ask additional questions with the goal of learning more about what the participant thinks about the nature and process of Spotify recommendations.

Questions:

1. a) In the survey you completed you indicated that you were [satisfied/not satisfied] with Spotify's recommendations. Can you talk about why this is the case?

1. b) In the survey you completed you indicated that you thought Spotify's recommendations were made by [algos, algos & humans, humans & algos] or [didn't know]. Can you talk about this more?

2. In the survey you provided a response to the question about how Spotify uses information to determine the personalized music recommendations for you. Let me remind you of your response.

Can you elaborate on this?

In order to focus participant responses on elements of machine learning, follow-up questions will directly or indirectly reference the three key ML functions: representation, evaluation, and optimization. Moving from the general to the specific, follow-up questions will seek a deeper understanding of concepts raised by the participant. Counterfactual or contrastive questions will broaden the conversation by probing areas unexplored by the participant.

a) Representation

What kind of data does Spotify learn from?

How does Spotify learn? For example, humans learn by a variety of methods: repetition, trial and error/reward, exploration (play, curiosity), from others, from self-direction

How much data is Spotify trained on?

b) Evaluation

What criteria does Spotify use to make recommendations?

What constitutes success in Spotify recommendations?

Why does Spotify consider a song or artist a good match for you?

c) Optimization

How does Spotify determine the best recommendation for you? What information is being given priority?

Spotify has over 285 million listeners (some free, some paid), music by millions of artists, is own by the 3 largest music publishing and distribution companies in the world, and, for the free version, is supported by advertisers. Given these different stakeholders, whose interests does it prioritize? How does it do that?

3. In the survey you provided a response to the question about how what you could do to shape the personized music recommendations you receive from Spotify. Let me remind you of your response.

Can you elaborate on this?

In order to focus participant responses on elements of machine learning, follow-up questions will directly or indirectly reference the three key ML functions: representation, evaluation, and optimization. Moving from the general to the specific, follow-up questions will seek a deeper understanding of concepts raised by the participant. Counterfactual or contrastive questions will broaden the conversation by probing areas unexplored by the participant.

a) Representation

What data are Spotify not using?

What are the limitations of the data Spotify uses?

How do you retract or remove information from Spotify about you and your interests?

b) Evaluation

Spotify creates and maintains a Taste Profile about you. What do you think is in the Taste Profile? What does Spotify use it for? Do you think your Taste Profile is accurate? Why/Why not?

Your musical interests and tastes shift, over time, sometimes frequently. How does Spotify recognize this and respond to it?

c) Optimization

How often does the system make mistakes?

How accurate are Spotify's recommendations?

Are the Spotify explanations of why you are being recommended playlists etc. useful? If so, why; if not, why not? What would a good explanation be like?

4. In the survey you provided a response to the question about the extent to which 22 different factors might influence Spotify's music recommendations for you.

You indicated that factors [x, y, x] were "Very Important". Can you elaborate about why you think these are the most significant factors?

You indicated that factors [x, y, x] were "Not Important". Can you elaborate about why you think these are not significant factors?

5. Is there anything else about the Spotify music recommendation system that you would like to add?

Conclusion

Thank participant.

Ask for any questions or concerns.

Send transcript for review of participant (to correct factual errors or omissions)

Email compensation ($25 e-gift card).

Appendix 8: Transcription Confidentiality - Spotify User Interviews

Transcript Heroes
Transcription Confidentiality Agreement 2021

THIS AGREEMENT (the "Agreement") is entered into on this date of February 18th, 2021 by and between Michael Ridley ("the Discloser" or the "Disclosing Party") and Transcript Heroes Transcription Services Inc. (the "Recipient" or the "Received Party").

The Receiving Party desires to provide transcription services to the Disclosing Party. During the provision of services, the Disclosing Party may share certain information with the Receiving Party. Therefore, in consideration of the mutual promises and covenants contained in this Agreement the parties agree as follows:

1. Definition of Confidential Information.
(a) For purposes of this Agreement, "Confidential Information" means any data or information that is proprietary to the Disclosing Party and not generally known to the public, whether in tangible or intangible form, whenever and however disclosed, including, but not limited to: (i) information contained in audio and video recordings, (ii) transcriptions of audio and video recordings; and (iii) any other information that should reasonably be recognized as confidential information of the Disclosing Party.

(b) Notwithstanding anything in the foregoing to the contrary, Confidential Information shall not include information which: (i) was known by the Receiving Party prior to receiving the Confidential Information from the Disclosing Party; (ii) becomes rightfully known to the Receiving Party from a third-party source not known (after diligent inquiry) by the Receiving Party to be under an obligation to Disclosing Party to maintain confidentiality; (iii) is or becomes publicly available through no fault of or failure to act by the Receiving Party in breach of this Agreement; (iv) is required to be disclosed in a judicial or administrative proceeding, or is otherwise requested or required to be disclosed by law or regulation.

2. Disclosure of Confidential Information.
In accordance with seeking transcription services the Disclosing Party may disclose Confidential Information to the Receiving Party. The Receiving Party will:

(a) limit disclosure of any Confidential Information to its officers, employees, or agents (collectively "Representatives") who have a need to know such Confidential Information in order to provide the transcription services to which this Agreement relates, and only for that purpose;

(b) advise its Representatives of the very private and very confidential nature of the Confidential Information and of the obligations set forth in this Agreement and

Appendix 9: NVivo Codebook - Spotify User Interviews

Name	Description	Files	References
1. Good Quotes	Memorable quotes for use in highlighting key issues or observations	14	34
AI Algorithms General	General comments about AI, algorithms, or machine learning. Not for comments related to Spotify or RS	4	8
Explanations	General comments about explanations (not specific to Spotify). This node can be aggregated usefully.	17	46
Desiderata re Explanations	What users would like in explanations from Spotify. What is missing? More detail?	1	1
Need Desire for Explanations	Whether users feel a need or a desire for explanation. Are explanations in Spotify important/useful?	9	13
Spotify Explanations	Specific comments about the explanations that Spotify provides. Also includes users who didn't notice any explanations	17	32
How Made	This is used to aggregate the "how made" comments. Do not code into this.	19	126
Algorithms	Comments about the use of algorithms (machines, AI) in making Spotify recommendations	19	53
Algorithms vs Humans	Comments that compare, contrast, or in any way debate the role of algorithms and humans in Spotify's recommendations.	9	21
Humans	Comments about the involvement of humans in making Spotify recommendations	16	35
Not Knowing	Comments about not knowing how Spotify's recommendations are made and whether this is a concern for them.	6	17
Reco Concepts	General comments about recommendation concepts; specific comments in the child	19	300

Name	Description	Files	References
	notes. Useful to aggregate all.		
Agency	Comments about agency or control; either human, machine/algorithm or both	5	13
Categorization	Comments about how Spotify classifies, tags, codes, catalogues or otherwise identifies the information about music, artists, and users. Include discussions of metadata.	14	36
Cold Start	Comments regarding new users of Spotify (or any RS) and the lack of information upon which to make recommendations	5	6
Diversity	All aspects of diversity (or lack thereof) in Spotify recommendations. Diversity means different artists, genres, songs than the user normally receives. Diversity and novelty are the same.	11	15
Feedback	Comments about feedback by the user to influence recommendations.	13	48
History	Comments about a user's historical listening patterns.	3	4
Learning	Comments about how Spotify (the algorithms) "learn" about users, music, artists. How does Spotify know things?	10	20
Listening	Comments about listening (frequency, dwell time, etc.) and the effect on recommendations.	7	11
Popularity Bias	Comments about the popular bias problem (popular items get recommended more which in turn makes them more popular; unvirtuous cycle. Also a place for user who think they are given only "popular" or mainstream recommendations.	9	10
Privacy	Comments about privacy or surveillance.	15	39
Randomness	Comments about randomness in recommendations.	3	8

Name	Description	Files	References
Recency	Comments about the recency bias in Spotify (i.e., recent listening, liking, or play listing disproportionately influences future recommendations).	11	24
Related Associated	References to how data elements are "related" or "associated" with other data elements. It is the nature of the relationship being captured. Also includes relationships among musicians/groups.	5	12
Serendipity	Comments about serendipity (as opposed to randomness)	1	1
Similarity	Comments about similarity, similar, sameness, like-ness. Include adjacency and proximity.	14	34
Statistical Math Factors	Comments that capture the statistical or mathematical foundations of RS/AI.	7	10
Trust	Trust in Spotify's recommendations. Trust in Spotify as a corporation goes under the Spotify node.	5	9
XYZ	Code here if a specific comment with no relevant child node. Review later.	0	0
Reco Information	General Comments about information used in making recommendations. Specific data elements added to relevant child notes.	19	337
Affect	Comments about affect or emotional states as data elements for recommendations.	13	45
Artist	Comments about artists as data elements for recommendations.	3	7
Audio ID	Comments about music characteristics (e.g., tone, beats, rhythm - known as audio ID, audio DNA, or audio characteristics) as data elements for recommendations.	10	19
Demographics	Comments about demographics as data elements for recommendations.	10	15

Name	Description	Files	References
External Media	Comments about external media (e.g., Facebook, social media, blogs, music journalism) as data elements for recommendations.	14	35
Genre	Comments about genres as data elements for recommendations.	8	13
I Do Active	Comments about data elements for recommendations derived from the active and explicit actions of users (e.g., listening, marking “likes”, creating playlists, etc.). These are action on the Spotify system. Actions/characteristics not on the system go under (I Do Passive)	19	69
I Do Passive	Comments about data elements for recommendations derived from the passive and implicit characteristic of users (e.g., where I am, what I’m feeling, what I’m doing etc.).	7	10
Others Do	Comments about the active and passive data signals of others (including friends) that influence recommendations.	14	31
Spatial	Comments about spatial characteristics as data elements for recommendations (where I am, what country/city I live in, what I’m doing).	7	12
Spotify	Comments about actions by Spotify as data elements for recommendations (e.g., artists, song, genres promoted by Spotify).	13	26
Temporal	Comments about temporal characteristics as data elements for recommendations (e.g., time of day, day of week, season of year).	15	24
Weighting	Comments about how these various data elements are weighted in terms of priority, importance, or other factors.	6	11
XYZ	Code here if a specific comment with no relevant child node. Review later.	0	0
Reco Shaping	General comments about how a user might shape (change, adjust) the recommendations they receive. Also used to	11	33

Name	Description	Files	References
	aggregate later.		
Alter Delete	Specific comments on how a user might alter or delete data elements in their Spotify profile that are influencing recommendations.	4	7
Manipulating Gaming	Specific comments on how a user might "game" or manipulate data elements to confuse, obstruct or otherwise fool Spotify's recommendation process.	4	6
Methods	Specific comments and methods about how a user might shape (change, adjust) the recommendations they receive.	9	20
Reco Type	Classification of comments to reflect particular recommender system technologies. Use specific child notes; aggregate with this node.	0	0
Collaboration	Comments that reflect or define a collaboration-based recommender system	5	10
Content		9	13
Hybrid	Comments that reflect or define a hybrid recommender system	0	0
Knowledge	Comments that reflect or define a knowledge-based recommender system	0	0
Model	Comments that reflect or define a model-based recommender system	4	6
Rules	Comments that reflect or define a rules-based recommender system	1	1
Social	Comments that reflect or define a social-based recommender system	1	2
Satisfaction	Use only to aggregate child nodes.	19	56
Poor	Comments of poor results from recommendations or Spotify services.	7	18
Satisfied No	Expressions of dissatisfaction with Spotify's recommendations.	4	8

Name	Description	Files	References
Satisfied Yes	Expressions of satisfaction with Spotify's recommendations.	15	30
Spotify	General comments about Spotify. Also used to aggregate child nodes.	16	72
Consulted During Interview	Comments that indicate the user looked at Spotify during the interview.	8	9
Multistakeholder	Comments about the multistakeholder nature of Spotify (users, artists, record labels, producers) and which of these is preferenced.	6	13
Personification	Comments that personify the Spotify recommender system.	9	18
Spotify as Corporation	Comments about Spotify as a corporation.	6	7
Spotify vs Social Media	Comments comparing and contrasting Spotify with other platforms or social media (e.g., Facebook, etc.).	8	13
Survey Information	Special node containing each users' answer to the question in the survey: "How does Spotify use information to make personal music recommendations."	19	19
Survey Shape	Special node containing each users' answer to the question in the survey: "What could you do to shape the personalized music recommendations you receive from Spotify?"	19	19
Users	Use this to aggregate child notes	18	123
Attitudes	General user attitudes to Spotify, recommendations, music, etc. Attitudes as opposed to specific concerns or observations that might be better attached to another node.	15	63
Objectives Goals	Comments about user objectives or goals in using Spotify and its recommendations.	13	31

Name	Description	Files	References
Profile	Comments from users about the profile Spotify maintains on them to determine their interests and preferences (i.e., their “tastes”).	11	29

Curriculum Vitae

Name	Michael Ridley
Post-secondary Education & Degrees	University of Guelph Guelph, Ontario, Canada 1975 B.A. (Hon.)
	University of Toronto Toronto, Ontario, Canada 1979 M.L.S.
	University of New Brunswick Fredericton, New Brunswick, Canada 1980 M.A.
	University of Toronto Toronto, Ontario, Canada 2014 M.Ed.
	University of Western Ontario London, Ontario, Canada 2023 Ph.D.
Honours & Awards	Ron MacDonald Distinguished Service Award Canadian Research Knowledge Network 2015
	Outstanding Alumni Award Faculty of Information, University of Toronto

2013

Miles Blackwell Award for Outstanding Academic Librarian
Canadian Library Association
2010

Larry Moore OLA Distinguished Service Award
Ontario Library Association
2007

Related Work Experience

Librarian Emeritus
University of Guelph
2019 –

Instructor, First-Year Seminar Program
2004-2018

Chief Information Officer (CIO) and Chief Librarian
University of Guelph
2004-2012

Chief Librarian
University of Guelph
1995-2003

Associate University Librarian
University of Waterloo
1991-1995

Publications

Ridley, M. (2022). Machine information behaviour. In S. Hervieux & A. Wheatley (Eds.), *The rise of AI: Implications and applications of artificial intelligence in academic libraries* (pp. 175–188). Association of College and University Libraries.

Ridley, M. (2022). Explainable artificial intelligence: Adoption and advocacy. *Information Technology and Libraries*, 41(2). https://doi.org/10.6017/ital.v41i2.14683

Ridley, M. (2022). Can information behaviour inform machine learning? *arXiv*. https://doi.org/10.48550/arXiv.2205.00538

Ridley, M. & Pawlick-Potts, D. (2021). Algorithmic literacy and the role for libraries. *Information Technology and Libraries*, 40(2). https://doi.org/10.6017/ital.v40i2.12963

Ridley, M. (2019). Explainable artificial intelligence. *Research Library Issues* (299). https://doi.org/10.29242/rli.299.3

Ridley, M. (2019) The autonomous turn in information behaviour. *Information Research* 24(1).