



HAL
open science

Integrating digital traces into mixed methods designs

Yann Renisio, Amélie Beaumont, Jean-Samuel Beuscart, Samuel Coavoux, Philippe Coulangeon, Robin Cura, Brenda Le Bigot, Manuel Moussallam, Camille Roth, Thomas Louail

► To cite this version:

Yann Renisio, Amélie Beaumont, Jean-Samuel Beuscart, Samuel Coavoux, Philippe Coulangeon, et al.. Integrating digital traces into mixed methods designs: An application to the study of online music listening using survey, interview and stream history data collected from the same people. 2024. hal-04448365

HAL Id: hal-04448365

<https://hal.science/hal-04448365>

Preprint submitted on 9 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Integrating digital traces into mixed methods designs

An application to the study of online music listening

using survey, interview and stream history data collected from the same people

Yann Renisio,^{1,*} Amélie Beaumont,² Jean-Samuel Beuscart,³ Samuel Coavoux,⁴ Philippe Coulangeon,¹
Robin Cura,⁵ Brenda Le Bigot,⁶ Manuel Moussallam,⁷ Camille Roth,^{8,9} and Thomas Louail^{10,11,†}

¹*Centre for Research on Social Inequalities (CRIS), CNRS, Sciences Po, F-75007 Paris, France*

²*CESSP, Cresppa-CSU, F-75849 Paris, France*

³*Interdisciplinary Institute of Innovation (i3), Telecom Paris, F-91123 Palaiseau, France*

⁴*CREST, ENSAE, F-91764 Palaiseau, France*

⁵*Université Paris 1 Panthéon-Sorbonne, PRODIG, F-75005 Paris, France*

⁶*Université de Poitiers, CNRS, MIGRINTER, F-86000 Poitiers, France*

⁷*Deezer Research, F-75009 Paris, France*

⁸*CNRS, Computational Social Science Team, Centre Marc Bloch (CNRS/HU), Friedrichstrasse 191, D-10117 Berlin, Germany*

⁹*CAMS, Centre d'Analyse et de Mathématique Sociales (CNRS/EHESS), 54 bd Raspail, F-75006 Paris, France*

¹⁰*Laboratoire Géographie-cités, CNRS, Université Paris 1,*

Université Paris Cité, EHESS, F-93322 Aubervilliers, France

¹¹*Laboratoire PACTE (CNRS, UGA, Sciences Po Grenoble), F-38000 Grenoble, France*

We present a mixed methods research design that integrates individual-level traces of online content consumption with survey and interview data, all collected from the same sample of tens of thousands of people. We explain how the pairwise combination of these information sources solves methodological puzzles often encountered when measuring cultural practices and preferences. We provide a concrete illustration for the case of music listening on streaming platforms, and show that survey respondents and non-respondents do not stream the same music. We illustrate how the mixed data collected make it feasible to infer the social properties of non-respondents, and hence to assess bias in studies based exclusively on self-reported survey data. We provide empirical evidence that unlimited access to all kinds of recorded music on platforms does not blur the social boundaries between repertoires, and we show that artists have distinct audiences whose differences are more than generational or gender-based, but also encompass differences in educational attainment. Finally we describe how we used digital traces in individual interviews to foster spontaneous expressions of aesthetic judgments, which are known to be challenging to collect. We conclude by discussing potential applications of these results in the field of cultural sociology, and the feasibility of adopting similar experimental designs to investigate other social phenomena.

The use of mixed methods research (MMR) in scientific inquiry boasts a rich historical lineage (Maxwell 2016), and the large body of papers, journals, meta-reviews and handbooks focused on this matter provides useful typologies and debates to understand its strengths and challenges in the study of social phenomena (Bryman 2006, Clark and Creswell 2008, Greene et al. 1989, Leech and Onwuegbuzie 2009). Scholars in this field converge on two fundamental principles that help distinguish mixed from non mixed, and within MMR. To be considered mixed-methods, a research design must either use data of different types - generally according to their context of production: questionnaire vs interview vs observation - , or use different types of analysis of this data - generally differentiated according to their statistical or non statistical treatment. Among MMR designs, a secondary principle of differentiation is, when several sources of information are combined, whether these data are available for the same individuals (e.g. Tucci et al. 2021), or for groups deemed similar (e.g. Bourdieu 1979).

The case of multiple sources for the same individuals is generally the most fruitful, as it allows studying interactions of different social dimensions at the individual level, limiting issues of ecological fallacy (Robinson 1950). However, such data collection protocols are difficult to design for practical, technical, and ethical reasons (Leahey 2007, Pearce 2012), and most of the mixed methods studies that collect different types of data from the same people are of one of these two types of combination: 1) a questionnaire survey is conducted with a set of people (chosen randomly or not), followed by interviews with a subset of respondents ; 2) an ethnography of a group is conducted, and is complemented with interviews or a questionnaire survey to all or a fraction of its members. While the first design makes it possible to map out structuring differences and investigate more closely differences or similarities among socially distant individuals, an important limitation is that the data collected are exclusively of declarative nature, and information extracted from survey and interviews are unlikely to properly inform on what people actually do (Jerolmack and Khan 2014). On the other hand, ethnographic studies give access to the direct observation of people's actions, and can be contrasted or complemented with declarative

*Electronic address: yann.renisio@cnrs.fr

†Electronic address: thomas.louail@cnrs.fr

data (LaPiere 1934, Weller 1994), but it is more difficult to generalize their findings, and the social heterogeneity of the population studied is limited (Lamont and Swidler 2014).

The increase in the digital recording of a wide range of ordinary practices, via cell phones, mobile devices or social platforms (Lazer et al. 2009), represents a new opportunity for the study of human behavior, as it provides access to data of observational nature, at an unprecedented scale, both in terms of quantity (higher number of individuals from which the practices are observed, higher number of practices observed during longer time periods) and quality (fine grain information in the types and modalities of practices). By recording signals of actual practice rather than relying on self-reported statements, these data offer a nice trade-off between the two aforementioned approach: they constitute a more reliable source on actual behavior than declarations collected in surveys and interviews ; their scale and lower cost offer a powerful alternative to direct observation (Mohr et al. 2020).

The advantages of these new sources, heralded as an empirical revolution for social scientists (Burrows and Savage 2014, Salganik 2018), has initially led to streams of studies solely based on such materials (Bakshy et al. 2015, Lewis et al. 2012, Onnela et al. 2007). In parallel, other scholars have commented on the multiple limitations associated with digital traces of people’s activity, including noise, lack of information on the conditions of in which these data have been collected and subsequently treated/filtered, uncertain meaning that can be extracted from the data, the risks of generalization of the digital observation to non-digital similar practices, etc. (Bastin and Tubaro 2018, Boyd and Crawford 2012, Ollion and Boelaert 2015). Another type of limitation inherent to this sole observation of practice is that, while providing a broader and more reliable account of what individuals actually do, it does not inform on the variety of meanings associated to such activities, hence limiting the possibility to explain what people do only with digital recordings of what they did, and falling in the well documented traps of revealed preferences approaches (Grüne 2004, Moureau and Vidal 2009, Sen 1973).

The conjugation of the strengths of large scale observations of digital practices to those of traditional measures of attitudes and meaning-making such as surveys and interviews seems a promising future for a grounded explanation of individual activities. A recent trend of work is heading towards this direction (Eck et al. 2021, Parry et al. 2021, Stier et al. 2019), but such journey faces serious challenges. First, a significant hurdle arises from the divergence between the providers of these sources, typically situated on opposite sides of the public-private boundary. Most of individual digital traces are collected by private companies, whose R&D sectors tend to be academically trained and oriented towards goals that are more distant from social science and closer to fields like physics or computer science, that have a longer history

of partnership with the private sector. Conversely, the knowledge and time necessary to the design of large scale surveys including standardized measures of important socio-economic characteristics, as well as carefully crafted interview campaigns conducted by trained interviewers, is essentially held by the field of public academic social science, which scientific goals are generally orthogonal, if not opposite, to the private sector, and who interact rarely and be reluctant to collaborating with non-social sciences (Renisio 2017). A second reason, reproducing the previous one, but within social sciences, is that the skills necessary, on the one hand, to handle and analyse large scale platform data and, on the other hand, to design, and analyse surveys and to conduct interviews, are a structuring principle of division between our academic communities. From an early age, future social scientists are incentivized by a series of institutions to define themselves as either belonging to the "quants" or "quals" tribes (Renisio and Sinthon 2014). And, more often than not, research teams tend to reinforce this trend rather than act against it.

The case of music listening

The case of music consumption, in the field of study of cultural practices, offers a good illustration of the duality of the empirical needs of the social sciences (of observational nature, to know what people actually listen to; of declarative nature, to know who these people are, what they say and think about different kinds of music), and the rarity of their combination in empirical research. Two major research currents can be distinguished along the aforementioned lines.

The first, established for a long time, analyzes the social stratification of musical tastes based on information obtained mostly from questionnaires (Bennett et al. 2009, Bourdieu 1979, Coulangeon and Lemel 2007, Prieur et al. 2008, Sullivan and Katz-Gerro 2006) or more rarely interviews (e.g. Le Guern 2017). These sources provide detailed, self-reported information on individuals’ characteristics, listening practice and music taste, but are declarative only, which poses the dual problem of the reliability of retrospective accounts (Prior 2009) and the biases generated by the unequal social desirability of different repertoires (Krumpal 2013).

The second trend, driven by the emergence of streaming and online music consumption, studies the uses of platforms and algorithmic recommendation in particular, and relies upon users’ stream history data collected by these platforms. Previous studies have examined differences between users in terms of the amount of music streamed, or the number of unique songs and artists streamed within a common period (e.g. Louail and Barthelemy 2017, Zhang et al. 2013). Additionally, some studies have focused on the impact of recommender systems on the diversity of content streamed by users (e.g. Beuscart et al. 2019, Villermet et al. 2021). While this

research is much richer empirically than traditional studies in terms of actual data on practice, it nevertheless suffers from a lack of information on user characteristics and preferences. In particular, it is rare to be able to access dimensions such as socio-professional category and level of education, despite these being established by the literature as central in the stratification of cultural consumption.

The RECORDS project is an attempt to connect these two lines, through a collaboration between social scientists and computer scientists working in public research institutions, with the research department of Deezer, a music streaming service that is one of the market leaders in France. The initial motivation was to bring together, in a framework that would provide the conditions that are necessary for public-private exploratory research (especially methodological experimentation), teams of people with different skills and a shared interest for the social dimension of music preferences. On streaming platforms users can navigate huge music catalogs that contain dozens of millions of tracks and artists, in fact almost all the recorded music listened to in France today. While these platforms have been around for some time now, and are nowadays the main source of access to recorded music, little is known (especially through numbers) on how much people are different when it comes to their *effective* listening practice, at least the part of it that is recorded in the logs that are collected by the platform they use to listen to recorded music.

Article outline. In the next section we present the GDPR-compliant, MMR design we implemented. It supported the collection of information of different nature (self-declared information in online surveys ; detailed information in extensive interviews ; observational, longitudinal data encoded in streaming logs) for a large sample of thousands of volunteer participants. We then present some of the results we obtained by analyzing pairwise combinations of the three aforementioned sources. Finally we discuss some methodological questions and perspectives opened by this work, in the field of cultural sociology and beyond.

Data collection

First we present each of the three data sources we collected, and then describe how we combined them in a MMR design. Although this design is rooted in a casual 'sequential explanatory design' (Clark and Creswell 2008), we have incorporated several elements that make it relevant for addressing current research issues in the field of cultural sociology — as will be demonstrated in the results section. Furthermore, its versatility makes it adaptable for studying a diverse array of social phenomena.

Platform data

The data provided by Deezer for this project are typical of the information collected by streaming platforms as part of their activity, independently of this research's purposes. It consists of three main sources: *(i) Catalog metadata:* This contains, in our case, the information about the approximately 90 million tracks made available to users. These are relational data that link songs, albums, and artist information together, along with the content of all the playlists available on the platform. *(ii) User data:* This includes the birth year and gender that users self-declared when registering on the platform, as well as the list of items users added to their 'favorites' collection by clicking on a heart-shaped icon. *(iii) Streaming history data:* This is the complete list of all timestamped stream events ever generated by the user account.

Users' streaming history data. In the context of music streaming, a 'stream' refers to the continuous, real-time delivery of audio content over the internet. When a user engages in music streaming on a platform, the music is sent to the listening device (typically a mobile app or desktop web browser) and played almost instantly. Legally, a stream is defined as a listening event lasting more than thirty seconds, as only these are considered for calculating the remuneration of the song's copyright holders. From a data collection perspective, a stream is a data record that includes a timestamp, a user account ID (allowing to connect with the user data), a track ID (allowing to connect with the catalog metadata), and various client-side information casually collected by websites and services (such as device type, IP address, etc.). The streaming history of a user comprises thousands of such timestamped streaming events, documenting all interactions of the user account with the platform catalog since its creation, spanning several years for many users. These raw data allow the reconstruction of higher-level information, including the complete list of songs, artists, albums, and playlists streamed by the user account, along with details such as the number of times each item was streamed, the total time spent streaming each item, and the channel/mode of access (e.g., search bar, personal or editorial playlist, recommendations by other users or algorithms). Additionally, the streaming history data indicate whether a song was listened to entirely, repeated, or skipped before completion. Contextual information, such as the type of internet connection used and a low-precision geographical location (derived from an IP-based geolocation third-party service), is also associated with each streaming event.

However, Deezer — as all other platform companies — does not collect systematically detailed information about their users' socio-demographic characteristics. For that purpose we designed an online survey and sent it to large samples of Deezer users.

Online survey data

Collecting precise socioeconomic information. In our MMR design, the primary objective of the online survey was to gather comprehensive self-declared information from Deezer users, complementing their individual streaming history data. This information encompassed variables such as age, gender, education level, household size, place of residence, marital status, employment status, occupation, income, as well as details about parents' country of birth and occupational categories when the respondent was 15. The survey, self-administered in nature, had a median response time of just under 15 minutes during the third wave, which is considered lengthy for an online self-administered survey. We hypothesized that introducing the socioeconomic module at the beginning of the survey would discourage some respondents from proceeding, so we chose to put it at the end, and to start the survey with music-related questions to enhance participants' engagement.

Collecting traditional declarations on cultural practices and music preferences. In addition to collecting detailed socioeconomic data, the survey aimed to gather declarative preferences and listening practice information, aligning with traditional quantitative studies in the sociology of cultural practices. This served a dual purpose: comparing our survey responses with other surveys and contrasting these answers with the respondents' streaming practices recorded by the platform. By the third wave, this information was divided into four modules.

The first module focused on the devices and internet services respondents used for listening to music, exploring the frequency and contexts of this activity both online and offline. The second module delved into the music genres respondents declared they listened to, their preferences among these genres, and their opinions on specific artists representative of different genres. To ensure simplicity and align with the typology presented on Deezer's platform, we adopted the nomenclature of 25 music genre labels featured on the Deezer homepage as a navigation tool. The third module centered on the use of Deezer, encompassing different navigation and recommendation/discovery features. Here we also inquired about the number of individuals regularly using the account, providing a basic yet incomplete criterion to identify accounts declared as shared in subsequent analyses of individual streaming practices. The fourth module focused on cultural and leisure practices more broadly, covering areas such as TV, radio, cultural outings, etc. This module replicated several questions from the 2018 edition of "Pratiques culturelles", a reference survey on the topic in France (Donnat 1998, 2009, Lombardo and Wolff 2020).

We ensured that survey responses related to listening habits and preferences could be easily matched with streaming data. Achieving this alignment is straightforward when surveying at the level of well-defined items, such as particular artists, bands, or songs, as the declared

appreciations can be unequivocally linked with indicators derived from the user's streaming history, like the total time spent streaming these artists. However, it becomes more uncertain at the level of genres or other vaguely defined categories. In the remainder of this paper, we navigate this challenge by concentrating on the alignment of survey responses with streaming data for well-defined items exclusively, namely artists. We develop the issues and potential improvements for genre-based analysis in the discussion section.

Deezer's international presence and the standardization of streaming data collected by platforms across countries makes it possible to internationalize the survey at little cost. To this end, the questionnaire has been translated into several languages, and tens of thousands of users will be solicited in Great Britain, Germany, Belgium, Switzerland and Netherlands. The socio-demographic block of the questionnaire has been standardized to the ISCO-08 reference nomenclature, and will enable a comparison on a European scale. However in the results section we will focus only on the 'french version' of the survey, that is the one distributed to users located in France.

Interview data

Individual interviews were designed to increase and enhance information on individuals for whom we already gathered streaming and survey data. The goal was to gain access to a less constrained description of the individual's perceptions of their music consumption.

We developed a modular interview framework that includes four modules. The first one explores the contexts and configurations of users' listening experiences, delving into details such as the equipment, usual locations and times for music listening, while also placing the streaming activity within the overall music listening activity of the interviewee. The second module examines music browsing and selection, with a particular focus on the use of algorithmic features, and on the users' creation of their own classifications, such as playlists and favorites. The third module aims to identify users' musical likes and dislikes, investigating their musical socialization through family, marriage, peer groups, or the workplace. In the final module, we asked interviewees to react to three songs that we selected beforehand, drawing on their streaming history data.

A first wave of thirty interviews took place between May and July 2021, conducted by seven researchers of our team, utilizing video conferencing due to the pandemic context. Respondents who had agreed in the questionnaire to be contacted for interviews were offered a gift voucher worth 30 euros as an incentive. The interviews, lasting an average of one and a half hours, were recorded and then coded by the interviewers using MaxQDA software. Transcriptions of the interviews were entrusted to professionals. The coding grid resulted from a collective

effort to finely theme the collected material and identify initial analytical avenues. The first ten interviews were coded by two team members blindly, a practice that revealed differing approaches to the material. We therefore decided that each interview would be coded by one person only, then reread by a second one to complete the initial coding. To ease the overall grasping of the interviews by the entire team, team members primarily coded interviews they had not conducted.

Simultaneously, each interview was summarized in a portrait created by the interviewer, outlining the main information about the interviewee and forming a separate corpus for an alternative understanding of the material beyond the more transversal coding. Interviews for the third wave are ongoing, aiming for a final number of one hundred interviews covering a broad social spectrum of users (see Figure 2).

Integrating components in a MMR design

The combination of these three sources of data can be presented in different ways. From a practical standpoint, the survey plays a central role: its initial question explicitly asks for the consent from the surveyed users to associate their responses with their streaming history data, while the final question inquired about the possibility of re-contacting them for an interview. From a technical perspective, the primary challenge was addressed by the Deezer R&D team, which provided us with a unique anonymized identifier (a hashed id) facilitating the linkage of platform data, survey responses, and interview transcripts, while preserving the anonymity of the respondents. Last but not least, from an ethical and legal standpoint, our team received support from both the CNRS and Deezer data protection officers to ensure the GDPR compliance of the data collection protocol.

At the core of this experimental MMR design lies an *explanatory sequential design* (Clark and Creswell 2008), in which we incorporated individual-level digital traces – that is, longitudinal observational data, integrated with declarative data. We also added several other ‘ingredients’ and the MMR design can be summarized by four salient properties: nested, sequential, iterative, and scalable, as illustrated by panels A and B of Figure 1.

Firstly, as shown on Figure 1B, our three sources and corresponding user samples are *nested* (Lieberman 2005): each of the interviewees is a user who responded to the online survey, and all these survey respondents are users whose streaming history data had been extracted and made available to the research team beforehand. All waves considered, there are about 100,000 users whose streaming history data were extracted and analyzed; 20,000 of them who responded to the survey; and 100 of them who took or will take part in an individual interview. Secondly, the MMR design is *sequential* (Leech and Onwuegbuzie 2009): chronologically, our protocol starts with the sampling of users to solicit from their

streaming history data ; then we obtain answers to our survey questionnaire among this sample ; and finally, we interview some of these respondents. Thirdly, our design is *iterative* (Pérez Bentancur and Tiscornia 2022): altogether, the basic data collection steps described in the aforementioned sequence constitute a wave, and all these steps are repeated and refined in subsequent waves based on previous analyses, as illustrated by Figure 1A. Finally, the design is *scalable*, and each wave expands at a larger scale.

Results

The three sources of data collected in our process complement each other and allow new results to emerge. We present their combination pairwise (as illustrated by Figure 1C), progressing from the most classical to the most original pairing experiments we conducted.

1. Surveys and interviews: A classic and powerful combination

The advantages and limits of the conjunct use of surveys and interviews stand as the most explored ones in the mixed methods literature (Harris and Brown 2019), including in the sociology of culture (Silva et al. 2009). This association helps gather both systematic information from surveys on large and potentially diverse or representative populations, complemented and nuanced by the richness of people’s self description of their own practices. While the benefits of this methodological synergy have been well-documented, this section aims to underscore two aspects that proved especially valuable for our research design. Firstly, the technique of “purposeful sampling” was employed to create a stratified sample from the pool of interview volunteers, with the goal of maximizing the social diversity of people interviewed (Palinkas et al. 2015). Conversely, interviews served to unveil the ambiguity or plurality of meanings inherent to questionnaire responses. This insight was instrumental in refining the survey questionnaire from one wave to another.

1.A. Ensuring social diversity among interviewees

The distribution of gender, educational levels, and age within our survey respondents sample exhibited uneven representation, characterized by a predominant presence of highly educated, middle-aged men (see Figure 2). This observation aligns with existing literature suggesting that practices such as online content consumption and online skills (Hargittai et al. 2019), music consumption (Katz-Gerro 1999), and participation in online surveys (Blasius and Brandt 2010) are practices where these specific social groups tend to be over-represented. This trend

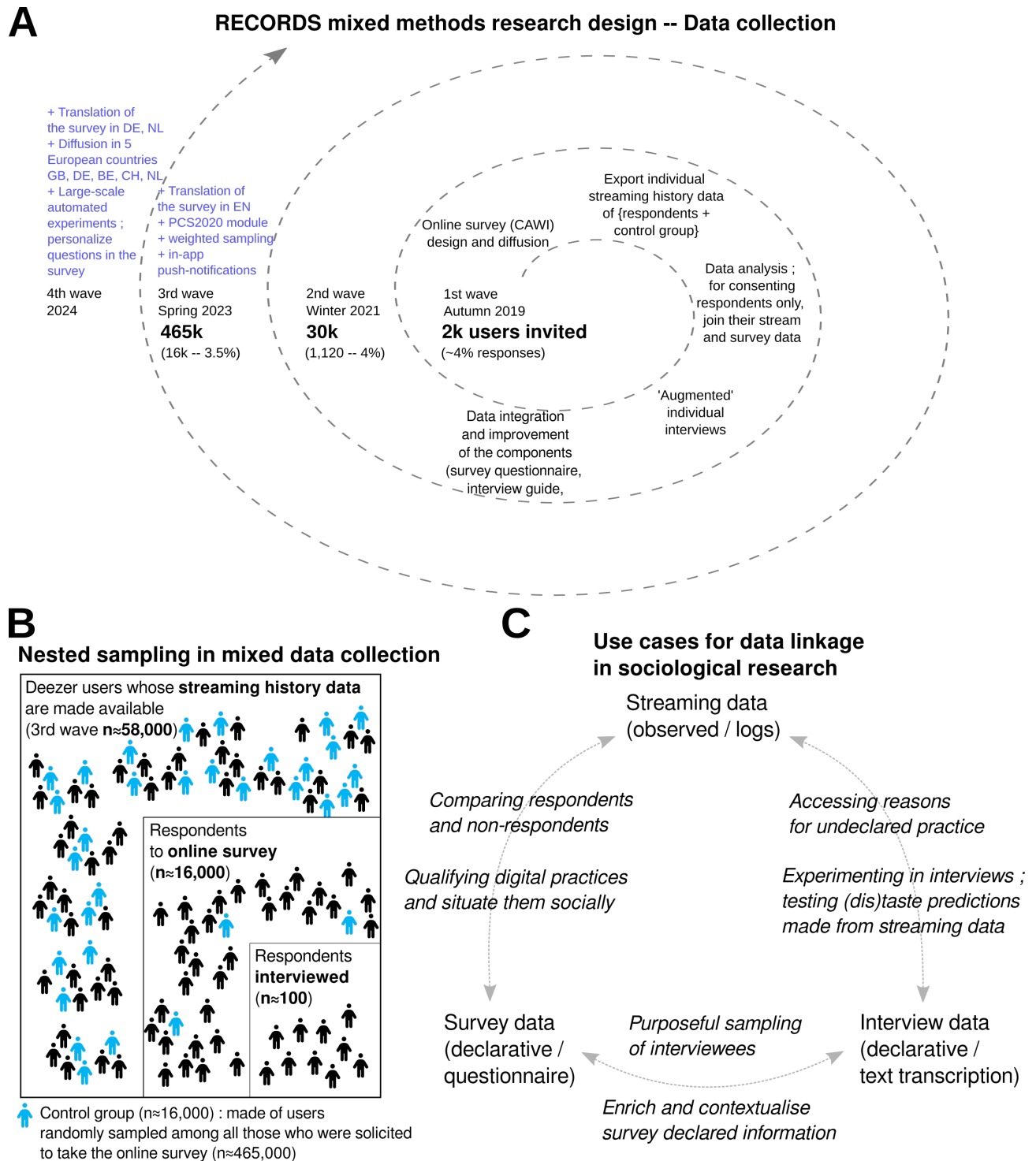


FIG. 1: Principles of data collection and article outline. (A) Data is collected in successive waves of increasing size, each wave consisting in the same sequence of operations. The number of Deezer users invited to take part in the survey is written in bold. This wave-based approach made it possible to improve the MMR design components and add new features along the way (in blue), and to scale up the size of the population invited to take the survey (B) Population and data are nested to access three different sources of information for the same population. (C) Some benefits of the pairing of the three sources.

is stronger in the sub-sample of individuals who volunteered for an interview. This subgroup consists primarily of highly educated men aged between 25 and 44 (a

demographic that closely mirrors the profile of our research group members). Notably, gender emerges as an influencing factor in interview agreement, increasing the

percentage of male participants from 60% among survey respondents to 66% among those willing to do an interview. Figure 2D highlights the intricate interplay of age, gender, and education levels, illustrating how these combined dimensions lead to sample deformation. Given the substantial impact of these three dimensions on cultural practices (Bryson 1997, Carter 2006, Lizardo and Skiles 2016, Noûs et al. 2021), our objective was to encompass a broad spectrum of combinations of these properties within our sample of interviewees.

To achieve this, rather than sending interview invitations to survey respondents randomly sampled among those who volunteered for an interview (as we did for our surveyed population), we defined 30 categories by intersecting the three variables (age, gender, education/degree). The goal was to conduct at least one interview per category in the first wave. Employing these criteria, we simulated several thousand potential samples. Subsequently, we filtered ten simulations that maximized profile diversity on secondary variables, including the total time spent streaming music over the previous year, the frequency of use of algorithmic recommendation features, the weight of jazz and classical music in the streaming history, the declared number of account users (one vs. more), and the type of municipality of residence (rural, urban, large urban). From this filtered pool, we collectively selected one simulated sample to proceed with.

Each interviewer then received an ordered list of individuals to contact within a specified group. For instance, among men aged 25 to 34 with lower education levels, the interviewer was asked to contact the first individual on the list. If there was no response one week after a first reminder (itself sent one week after the first interview invitation message), the next person on the list was contacted, and so forth. The assignment of these thirty {age, gender, degree level} groups to interviewers was randomized, with the exception of women with lower levels of education. To prevent introducing a gender difference alongside a class distinction, female interviewers within the team were specifically tasked with conducting interviews within these particular groups. We have effectively conducted one interview for each of the 30 groups identified in the second wave, forming the foundational pool from which excerpts are drawn for presentation in the following sections. The ongoing wave of interviews adopts a similar design.

1.B. Enriching the contextual information on practice

In turn, the interviews played a crucial role in identifying survey questions that may carry a different meaning than originally anticipated. For instance, Valentin and Laure (two of the interviewees) responded "no" to the survey question "Do you listen to music while working?". However, during the interviews they both described a consistent exposure to music while at work:

Valentin : I work in restocking at a supermarket with morning hours, which means that from noon onwards, the entire day is available.

Interviewer : During this work, do you sometimes listen to music or not?

Valentin : No. At work, no. I even think it's prohibited, just to stay alert. [...] Well, I do listen to music: there are speakers in the store. But there, clearly, it's not me who chooses.

Laure : Actually, in our department, we work in the operating room, and so, it's like in the car, it's the one who is driving or, in this case, the surgeon, who has the right and priority. So often, to maintain concentration in the room, it's the surgeon who somewhat chooses his playlist. But if it doesn't appeal to everyone, well, we make him feel that he also works as part of a team and that it's not acceptable. So, we try to find compromises.

The apparent discrepancy between the responses in the questionnaire and the descriptions provided in interviews seems to be tied to the interviewees' interpretation of the statement "listening to music" in the survey. It is possible that, given the collective nature of the listening activity they recount, they implicitly understood the survey statement as "listening to music individually." However, another dimension also characterizes these accounts: owing to their subordinate positions at work, they are subject to the musical preferences of some of their colleagues. This may explain why they do not associate this work context with the concept of "listening to music" as framed in the survey. This interpretation seems reinforced by the testimony of another interviewee, Fabienne, who answered "yes" to this question. Fabienne, holding a supervisory position in her workplace, sheds light on how she manages the broadcast of her preferred radio station (FIP) in the workshop she oversees:

Fabienne : So, I control it from my phone, the radio, yeah. But yeah, it's a speaker that we've set up, so what's annoying for... let's say... for many people I work with is that we're in a company where there's noise, [... And] music adds noise for some, and um... But for me, not at all! (laughter) [...] And when I like it, I turn it up! (laughter) [...] So it's true that some people endure it, but that's how it is, it's something, you know, in our profession... [...] Even if we don't have the same tastes. Even if music, sometimes, annoys some because, precisely, we're

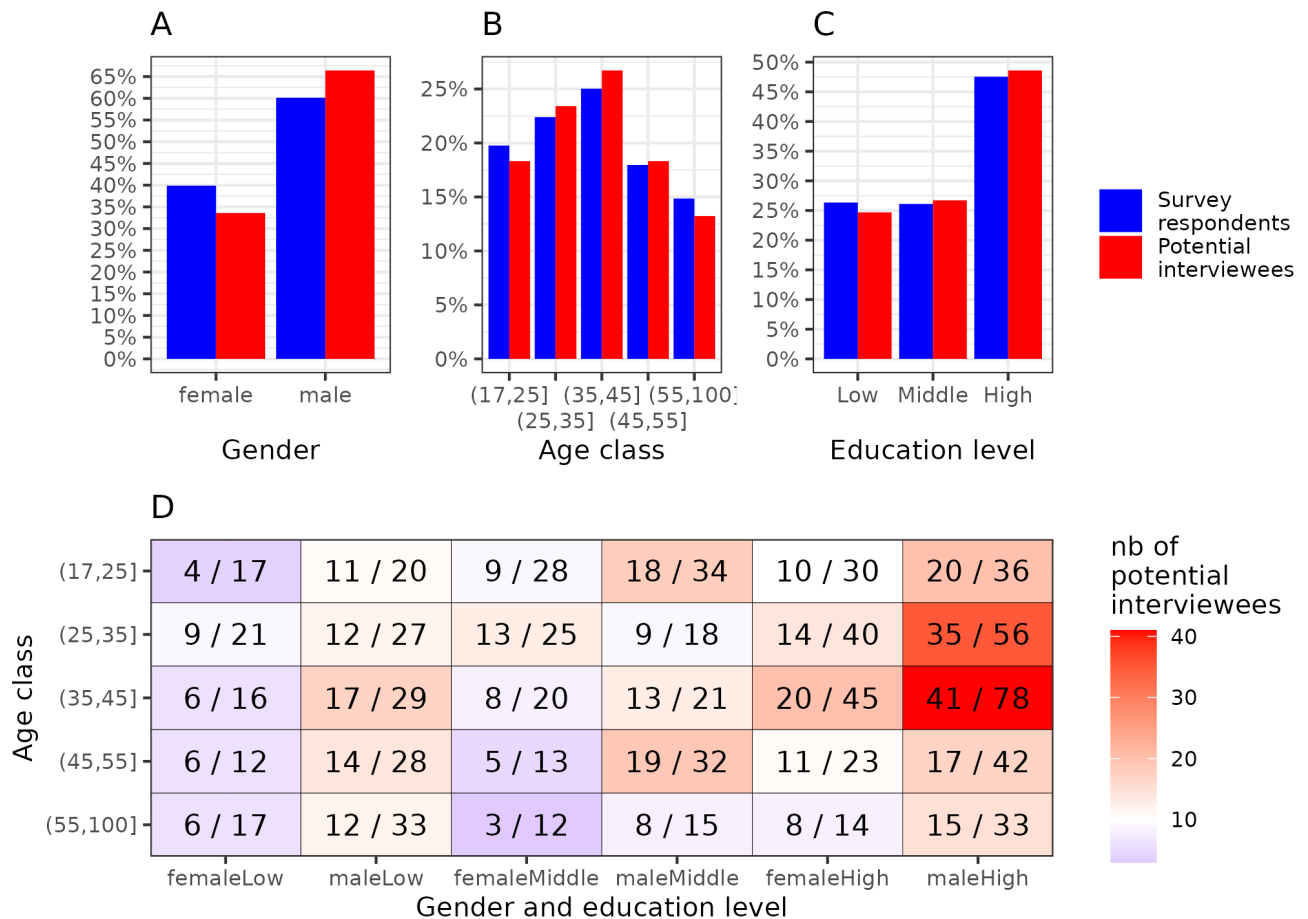


FIG. 2: Differences in social characteristics of online survey respondents, according to their willingness to take part to an interview. (A) Gender, (B) Age class, and (C) Education level distribution. (D) Combination of the three factors and number of interview volunteers over the total number of respondents in the group.

Field: 835 respondents to the second wave of online survey and who declared a gender as male or female, an age between 17 and 100, and a level of education.

Reading: (D) Out of the 78 men, aged between 34 and 45 with a high level of education, 41 gave their consent to be contacted for an interview.

Note: Level of education is divided into "Highschool degree or less" (low); "More than highschool, less than Masters degree" (middle), "Masters degree or higher" (high). For the youngest group, education level were replaced with parents' socio-economic status of both parents at age 15.

focused, we have to... we do meticulous work.
When it's too loud, it might bother some. So,
you have to be tolerant with music at work.

In this scenario, the interviewee asserts her musical preference over her colleagues, a decision that seems to generate tensions but remains non-negotiable for her. Tolerance, in this case, is expected to be exercised by her subordinates. While we cannot claim that her colleagues would have declared — had they been asked in a survey — that they do not listen to music at work, the contrast with the two previous accounts leans towards associating control of content and hierarchy, rather than collective contexts, in the disparity of meanings attributed to musical exposure.

A similar benefit of combining interviews and questionnaires surfaced when comparing the music actually streamed by survey participants with the music they declare listening in the survey (see the discussion). While analyzing the second wave of interviews and presenting our research in seminars, we observed that classical music was more frequently associated by interviewees and by colleagues with a diverse range of listening devices, beyond Deezer (including platforms or radios dedicated to that genre, physical records, etc.), more than any other music genre. To ensure that the differences we measured were not influenced by potential ambiguity in our survey, we introduced a supplementary question in the third wave's survey. After participants selected the genres they

listen to from a list of music genres, they were asked to select among them the ones they were also listening to on other media and platforms besides Deezer.

Although conventional, the triangulation of information gleaned from surveys and interviews has proven instrumental in acquiring crucial insights into our research subjects. Employing a snowball sample technique, as well as adopting a random sample design, would not have provided coverage of individuals with such diverse social profiles in interviews. Furthermore, the juxtaposition of data from questionnaires and interviews unveiled potential contradictions. These contradictions, in turn, prompted the generation of new hypotheses and facilitated the refinement of the data collection in subsequent waves.

2. Digital traces and surveys: linking observational accuracy to sociodemographic characteristics

A less common pairing of sources involves combining survey answers with digital traces. A first reason for integrating these two sources is to mitigate the limitations of each of them. The deficiency in measuring actual practices in surveys is offset by the wealth of observational data obtained from traces, while the absence of social qualifications in traces can be complemented with the socio-demographic data collected through surveys. In the following sections, we first show how having access to the streaming history data of a large sample of non-respondents made it possible to contrast their streaming practices and streamed content to those of respondents. Then we explain how we leveraged this mixed material to gauge the association between various social characteristics and music consumption.

2.A. Comparing respondents and non-respondents

Most questionnaire surveys rely on random sampling techniques to reach statistical representativity of the target population. However the assumptions of randomness are rarely met because of the heavy rate of solicited individuals that do not take the survey: if not answering a survey is not a random phenomenon, then the conditions required to generalize the results are not met. By construction, researchers know little if anything about the non-respondents. Collaborating with a digital platform to survey its users introduces a transformative shift similar to techniques from census or sampling based on registers. In our case, the availability of content consumption data and basic demographic information self-declared by all users when they register to the platform (birth year and gender, see Data collection section) makes it possible to measure how these factors (gender, age and music streamed) are distributed among respondents and non-respondents.

Figure 3 highlights remarkable differences in the music

streamed by both groups. To leave aside the effects of age (different generations listen to different artists and genres) and also the fluctuations in listening to the least-listened-to artists (which necessarily vary from one person to another), here we focused on the 300 artists listened to the most by each of the two groups, respondents and non-respondents, who declared to be between 35 and 44 years old (the largest age-class among respondents). We calculated which artists were over/under-represented in each group, by calculating for each artist i its rank r_i among the respondents (the most streamed artist has rank 1, the second most streamed rank 2, etc.) and its rank r'_i among the non-respondents. For each artist we calculated the rank ratio r_i/r'_i and plotted in blue (resp. red) those for which $r/r' < 1/2$ (resp. > 2). Looking at the artist names coloured in red and blue, a clear pattern emerges: among the 24 artists the most over-listened to by respondents (in blue), all are English or American rock artists or bands, that are now relatively old (almost all of them started recording music in the previous century). On the contrary, all of the 28 artist most over-streamed by non-respondents (in red) sing in French, and belong to the recent generations of French rap and RnB. Thanks to the coupling of streaming and survey data, we will see in the next section that the social characteristics of the people streaming these artists also tend to be different.

2.B. Qualifying practices and situating them socially

Recent works have suggested that the platformization of listening, along with access to vast catalogs of music at little or no cost, would shift the principles of social differentiation from what we consume to how we consume it [Webster \(2019\)](#). On the contrary the first results we obtained when analyzing the music actually streamed by approximately 15k survey respondents, over several years, confirm the statistical connections between individuals' social attributes and the music they listen to.

We illustrate this point in Figure 4 and table I. Fig. 4 maps the most listened to artists among our respondents aged 35 to 44 in a space defined horizontally by gender (percentage of female listeners among all the users that streamed the artist at least once), and vertically by education profile (percentage of listeners with at least a Master's degree). We observe that the top of the figure, corresponding to the artists streamed by the most highly educated users, is predominantly occupied by artists who interpret classical music (such as the Berliner Philharmoniker orchestra, Max Richter, Chilly Gonzales, etc.), while artists located in the bottom of the figure (who are mostly streamed by the respondents who have the lowest educational profile) are predominantly associated with French hip-hop music (La Fouine, Imen Es, Sniper, Hornet La Frappe, etc.). The variation is substantial, ranging from 20% to 50% of highly educated listeners among all the users that streamed these artists at least once. Similarly, we observe significant variations in the gender

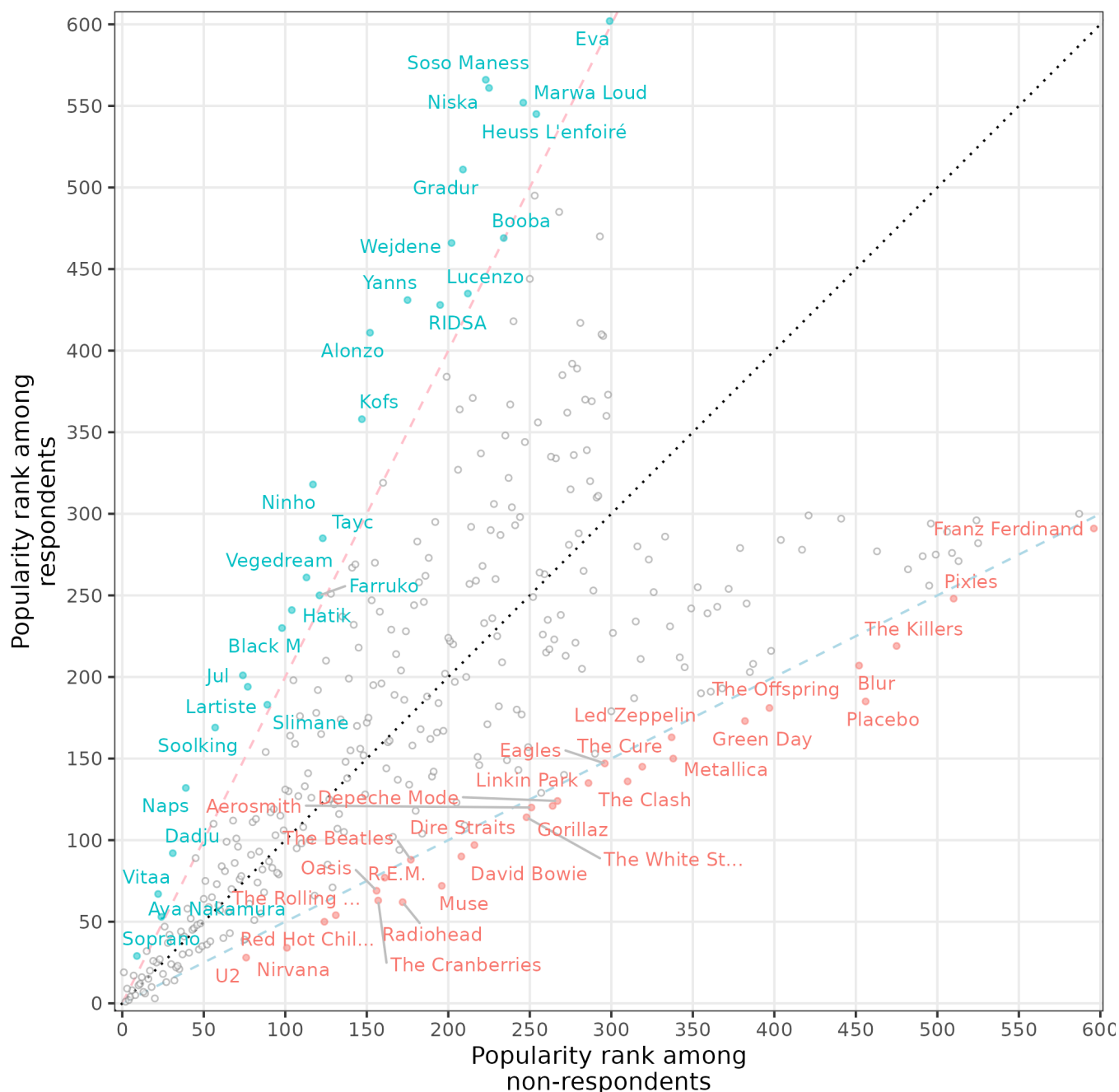


FIG. 3: Rank distribution of the respective top-300 artists of respondents and non-respondents belonging to a common age class (35-44).

Field: 3885 respondents and 7658 non-respondents of the third questionnaire wave, aged 35 to 44 y.o. in 2023.

Reading: The rank of the rapper Alonzo is below 200 among non-respondents users but above 400 among the respondents. The opposite is true for the rock artist Placebo.

note: Labels appear if the absolute rank difference >20 and the absolute rank ratio >2 .

composition of the artists' audiences, ranging from 25% to 60%. Among the artists who have the most masculine audiences (left of the plot), we find exclusively male artists, predominantly associated with genres like metal, rock, and hip hop. On the opposite side (mostly feminine

audiences, on the right), we observe a more balanced mix of male and female artists, with a higher prevalence of pop, RnB, and French (non-rap) genres. On this figure we also coloured in red and blue the artists previously labeled in Fig. 3, that is the top-artists that are over-

represented among respondents (in blue) and those over-represented among non-respondents (in red). The clear separation between these two groups of artists in a space defined by the social properties of the sole respondents strongly supports the hypothesis that the differences in the music streamed by these two populations inform on social differences. In other words, based on the streaming practices of our sample of solicited users and the social characteristics of the respondents in this sample, it appears that the population who did not respond is likely to be composed of more women and of less highly educated individuals.

To extend the analysis to different generations, we grouped the 15k survey respondents in age groups (25-34, 35-44, 45-54, and 54+) and determined for each group the top-100 artists that were the most streamed by each group during a common reference period of several years. We then ranked these artists according to the proportion of their listeners who declared in the survey they had a Master degree or higher diploma. Table I presents for each age group the top-10 and bottom-10 artists according to their rate of highly educated individuals among their listeners. First, the generational nature of music consumption is clearly visible, by the absence of common artists between the youngest and oldest cohorts, whereas adjacent cohorts consistently share some of their top or bottom-10 artists. Although classical music interpreters are not present at this popularity scale, a pattern akin to the previous figure persists: French hip hop is markedly over-represented among artists with the least highly educated audience, particularly in the two youngest cohorts. The artists in the top 10 of each age class predominantly enjoy critical acclaim in their respective genres (examples include Lana Del Rey, Billie Eilish, Amy Winehouse, Radiohead, or Nina Simone, representing one in each age class), while artists in the bottom 10 are often associated with “low-brow” listening in their genre (such as Booba, Kendji Girac, a-ha, or Slimane).

To our knowledge, these descriptions represent the first instance of integrating socio-demographic information from thousands of surveyed individuals with the systematic and automated recording of their actual consumption of cultural products, over a long period of time. In contrast to self-declared statements on practices or opinions extracted from a list of artists in scholarly surveys (like the French survey “Pratiques culturelles des Français”), our approach involves the fusion of survey responses and digital records of actual music consumption. Essentially, this unique combination enables, for the first time, the measurement of what scholars in the field of cultural practices have long sought to achieve — providing detailed insights into the relationship between social practices and socio-demographic data.

3. Digital traces and interviews: new pathways for eliciting and predicting judgments

The last pairwise combination of sources consists in using digital traces to ‘augment’ interviews. In contrast to the combination of questionnaires and interviews, integrating digital traces with interviews allows to introduce information into the conversation that the interviewee may not be aware of, or might prefer not to address. Because they have been passively recorded over long periods of time, streaming logs are considerably less subject to conscious control compared to survey responses. Our team of interviewers leveraged this opportunity in interviews through two primary methods. First, by using a data visualization application designed by one of the team members (Cura et al. 2022), we gained access to a detailed, visual and customizable summary of information on each interviewee’s streaming practice and streamed content. The app made this information easily searchable, and no prior technical knowledge was required to navigate the interviewees’ streaming data. This tool proved invaluable for both the preparation and the conduct of interviews. Secondly, these stream data were used to identify beforehand songs that we anticipated the interviewee might love or hate. This approach let us test the accuracy of our predictions, and helped provoke vehement reactions to certain pieces of music during the interviews.

3.A. Enhancing interviews with rich data on the interviewees’ practices

During the interviews, in more diverse ways than we initially anticipated, we used AMPLI (Cura et al. 2022) to enhance our interactions with the interviewees. Interviewers could query the streaming logs with various graphical filters, including time and place filters to identify listening habits at certain periods of times, or in certain locations. The use of videoconferencing facilitated the real-time use of the tool: we could engage in the conversation while searching for specific data, mentioning it when the application provided information that aligned with, or differed from, the interviewee’s ongoing discourse, or use it to propose respondents to play guessing games centered around their listening history. On several occasions, these use cases fostered the mention of significant biographical events.

Interviewer: We’ve been talking about world music and traditional music for a while now, but from your listening logs, which I was able to consult, I had the impression that you listened to a lot more electronic music...

Françoise: At one point, yes. But in fact, it’s a question of periods. I’ve been really ill

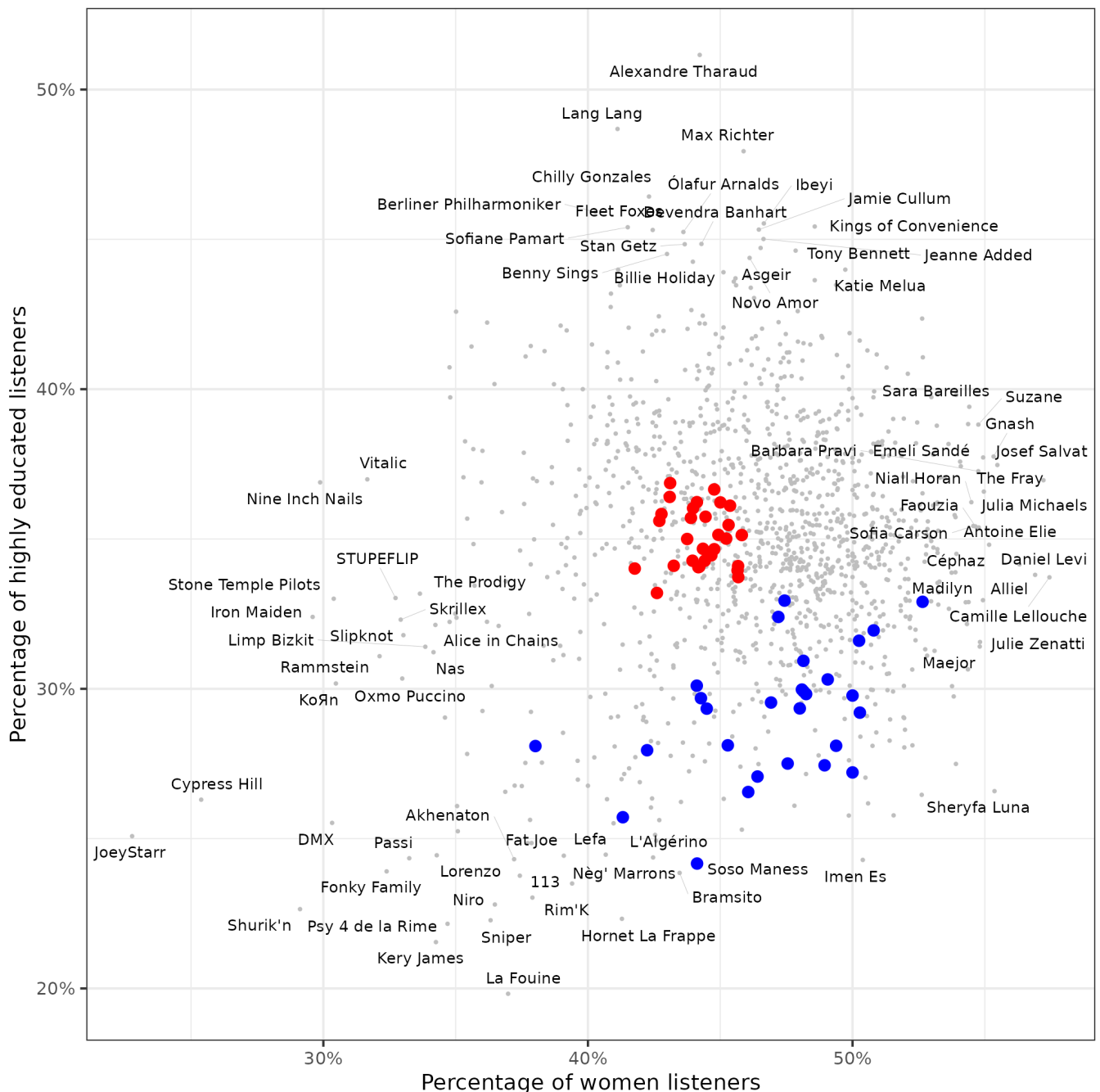


FIG. 4: Distribution of the most popular artists by the characteristics of their audience, in a given age class of respondents.

Field: 1483 respondents to the third wave's online survey, aged between 35 and 44 years old in 2023, who declared being the only user of their Deezer account, a gender as male or female, and a level of education. Artists are limited to the 1319 ones listened to by at least 300 respondents

Reading: Among our respondents aged 35-44, the hip hop artist Shurik'n has been listened to by more than 300 respondents, but his audience is composed of less than 30% of women, and less than 25% of highly educated individuals

Note: "Highly educated" is defined as having at least a Masters degree. Red and blue dots indicate respectively the labelled over/under represented artists among respondents compared to non-respondents in Fig. 3. Name of artists appear if they are among the top/bottom 20 most "feminine"/"highly educated".

for three years now, and I didn't listen to the same kind of music at all. I was really

fine... well, I'm still taking anti-depressants for example, but I really had moments when

rank	Age 25-34	Age 35-44	Age 45-54	Age > 54
1	Lana Del Rey	Juliette Armanet	Benjamin Biolay	Nina Simone
2	Taylor Swift	Lana Del Rey	Véronique Sanson	Neil Young
3	Billie Eilish	Amy Winehouse	M	Led Zeppelin
4	U2	Radiohead	The Beatles	Françoise Hardy
5	Ariana Grande	Billie Eilish	Daft Punk	Creedence Clearwater Revival
6	Glass Animals	Julien Doré	The Rolling Stones	Simon & Garfunkel
7	The Kid Laroi	The Beatles	Radiohead	Bob Dylan
8	Lizzo	M	Bruce Springsteen	Bernard Lavilliers
9	Daft Punk	Milky Chance	David Bowie	Serge Gainsbourg
10	Thérapie TAXI	Taylor Swift	Clara Luciani	Fleetwood Mac
...
91	Soolking	Major Lazer	GIMS	Vianney
92	Naps	Kungs	a-ha	George Michael
93	Kofs	Slimane	Maroon 5	Sia
94	Dadju	Eminem	Kendji Girac	Kate Bush
95	Tayc	Kendji Girac	Rosalía	Phil Collins
96	Booba	GIMS	Lil Nas X	The Weeknd
97	Soprano	Sean Paul	P!nk	GIMS
98	Alonzo	DJ Snake	Scorpions	Calogero
99	Jul	Soolking	Katy Perry	Louise Attaque
100	Ninho	Soprano	Calvin Harris	Kendji Girac

TABLE I: **Top and bottom 10 artists based on the proportion of highly educated listeners, among the 100 most-streamed artists of survey respondents, grouped in four distinct age classes.**

Field: 5205 respondents in the Wave 3 questionnaire, aged over 24, who reported being the sole users of their Deezer accounts, identifying as either male or female, and providing information on their level of education.

Reading: Lana Del Rey has the highest rate of highly educated listeners among the 100 most listened-to artists by respondents aged 25-34 and ranks second among respondents aged 35-44.

Note: Artists in bold are the ones who appear in several age classes

I wasn't fine at all, and I wasn't listening to world music, I was listening to other types of music. Well, extremely gothic music, quite violent stuff... well, not violent, but very aggressive, or which corresponded more to my depressive state and my anger too, you could say, which I don't have anymore [laughs]. [...]

Interviewer: When I looked at your number of streams per year, from 2018 onwards it explodes, and that's also when Electronic Dance Music becomes one of your most listened-to genres.

Françoise: Yes, and in 2018, well, it's not surprising, I didn't even remember: I had a burn-out. I'd been a caregiver for my father, my mother, and I was looking after my mother-in-law at the time. So I had a burn-out and ended up in a psychiatric hospital. I was listening to a lot of electronic music at the time, and I was terribly angry, really deeply angry. I've been in therapy ever since... [...] And now I listen to very little of it, or if I do, it just makes me happy, that's all. [...] I'd completely forgotten that I'd been listening to a lot of electronic music at the time. It's intuitive really.

This excerpt highlights the intricate connection between music, emotional states, and life events in Françoise's experience. It emphasizes the evolving and adaptive nature of music preferences, serving as coping mechanisms and expressions of inner emotions across different stages of her life. Such insights might have remained undiscovered by the interviewers without the analysis of listening patterns made beforehand by the interviewer through AMPLI.

Another use case emerged during the interview with Jean-Christophe, a 29-year-old metal fan and customs liaison officer. Notably, he did not mention other musical genres in the interview until the interviewer prompted him to guess his top 15 artists as they appeared on the application.

Interviewer: [After successfully mentioning most of his top artist, most of which are metal artists, the interviewer evokes one equally listened to, but from a different musical genre (French hip hop)] And Freeze Corleone.

J.-C.: [Expressing surprise] Uh, yeah, ok. Yeah, yeah. So... It's funny because it's not... In my head, it's not the same... It's

not in the same part of... Well, I don't know how to explain it. Uh... I don't listen to it the same way, actually. [...] Freeze Corleone, it's a bit like fast food. I like it, but basically, really basically, maybe not so much. Because... er... in fact, he's an artist I'm not sure I'll be listening to in ten years' time. For example. But right now, there are things I like, uh... uh... But after that, it's not what inspires me as a drummer. And I listen to it with certain people, with buddies, with uh... And yeah, in the car, I like it, it... But uh... it's not very... Musically, it's not developed, we'll say it like that. Even lyrically. But uh... Whereas Tool, well, they're a source of inspiration. And bands like Meshuggah, really instrumental bands. [...] Well, it's not that I'm ashamed, but musically, I am, a bit! Of... of listening to [Freeze Corleone]. (laughs) But then again, you should never be ashamed! I'm not very clear. I don't really know how to express myself, but uh... But for example, Tool, I'm going to recommend it to all the people who are likely to like this kind of music and this band. The same goes for Meshuggah or Converge, things like that. But Freeze Corleone, no, I'm not necessarily going to recommend it like that. [...]

Interviewer: .. And Alkpote?

J.-C.: (laughs) So we were talking about shame! (laughs) Now we're right in the thick of it! I'd never listen to it with my sisters or my mother, for example, in the car... (laughs) But uh... I recognize that there are some... It's entertaining! Well, that's what I wanted to say, that it's associated with entertainment. [...] For me, rap is also entertaining. Uh... Beyond the music, it makes me laugh. For me, it's comedy. It's so absurd and vulgar that I take it to the 5th degree and it makes us laugh... it makes me laugh uh...

Clearly, the interviewee wouldn't have mentioned these French rap artists if the interviewer had not brought them up during the interview. The expressed embarrassment and the multitude of justifications provided highlight an internalized judgment, akin to "I shouldn't be listening to this given who I am." This led him to explicitly articulate a socially marked hierarchy between music genres: he perceives French rap as the "fast food" of music, a genre that he feels "shame" listening to and wouldn't recommend, unlike the metal bands that inspire him and that he actively promotes. Consequently, by juxtaposing self-presentation and recorded traces of past listening, forms of social judgments emerge during the interview, generating explanations for these practices.

3.B. Predicting interviewees' (dis)tastes

In the last part of the interviews, we aimed to explore our ability to predict the musical preferences and aversions of interviewees based on their own streaming history data, and the streaming data of a random sample of Deezer users that were used to determine who were at the time the most streamed artists on Deezer in different music genres. Our overarching goal was to assess whether streamed content could serve as a reliable indicator of individual taste and distaste. Furthermore, we sought to determine whether, from the array of available signs in individual traces, we could effectively map the nuances of the types of music enjoyed or disliked by individuals. Questions such as whether recurrent listening of a particular track or artist consistently signaled taste, and whether the absence of listening, systematic skipping, or even "banning" of a track always indicated disgust, were central to our inquiry. We also wondered whether dislikes for unlistened-to genres were stronger than those for unlistened-to popular artists within a genre that was otherwise listened to a lot by the user. In essence, we aimed to establish a gradient from adored artists to disliked ones based solely on streaming history data.

With these questions in mind, for each interviewee we curated a personalized list of three tracks, anticipating positive, ambivalent, or negative reactions. Importantly, the interviewees were unaware of which tracks they would be listening to and commenting on. Overall, our highest success rates in accurately anticipating reactions occurred in cases of strong likes and dislikes. Specifically, this was observed when testing widely listened to and favored tracks, and very popular tracks in genres little or not at all listened to by the respondents, respectively. In the latter scenario, the reactions – sometimes vehement – often provided insightful revelations on the social principles fueling aesthetic judgments.

Predicting dislikes: reactions to listening to a musical extract chosen, on the basis of individual listening histories, to displease

Alain, 66, former customer service manager, on "Djadja" by Aya Nakamura: "Immediate destruction. This is something I wouldn't call artistic creativity. It's the opportunistic navel-gazing of a society focused on online social networks and all that goes with it. This charming Djadja certainly pleases a lot of people and undoubtedly makes a lot of people laugh on stage, but not me. No, but when you see the image she gives in public or through these things. It's all right, it's the big family of TV shows that show just how low human stupidity has sunk. It's a horror. But I can understand why some people like

it. But it's about satisfying the greatest number at the expense of quality. There's no quality there."

Benoit, 46, engineer, on "La Moulaga" by Jul and Heuss l'enfoiré: "Well, that's a total turn-off. Vococode, vocode or I don't know... Vocoding, ah no, it's really not possible, it's... it's crap, well it's... Ah I... Yeah really. Really bad, yeah, it's...Ah no, that sound there... that doctored sound of the voice, frankly, it repels me, but really. And it's a turn-off... I mean, it really... it makes my hair stand on end. Really, it's... It's hell... musical hell."

Nico, 35, care assistant in an EPHAD, on Gambi's "Popopop": "So that's not my audience! (laughs) Well, it's not my... it's not too... [...] A bit too much street rap, uh... A bit too much of what's being done at the moment, you know. [...] that kind of sound, there are dozens of them doing the same thing. [...] and anyway, it's not... it's not targeted for me. I think it really belongs to a community that's not mine. [...] people... how can I put it? People from... the suburbs, you know."

Caroline, 39, nurse, on Tryo's "L'hymne de nos campagnes": "Tryo, why I hate it, because I think it's a bit of a lie, this group, under the guise of committed lyrics, a bit hippy, it's still very commercial, and then a bit light... I mean, I find that the lyrics aren't really... I find that the lyrics are maybe a bit too simple and without a second degree. Because I've listened to much worse things, but with a second degree, so... But here, no, Tryo doesn't even have a second degree, so, no, I don't like it."

Jean-Christophe, 29, customs liaison officer, on Louane's "Jour 1": "Ok. Uh... Well, that's not my style at all, it annoys me more than anything else. Uh... After all, I have a minimum of respect for the artist, uh... all that implies, obviously, but uh... I don't like it. Even as background music, um... I really have to be imposed on, whether it's on the Leclerc radio or... (smile) There's no choice for me to listen to that. [...] Well, it's obvious, [the target audience] tends to be teenagers. Clearly uh... uh... and then rather feminine, obviously. Even if... there's the housewife under 50 who'll... who'll have to listen...! (laughs)"

These arguments encompass a spectrum of critiques, ranging from the perceived lack of "artistic quality" and a decline in cultural standards (as expressed through terms like "stupidity," "too simple," and "commercial") to disapproval of specific musical elements (such as the "doctored sound"). These critiques occasionally converge into explicit references to contexts or populations linked to social class judgments (evidenced by mentions like "Leclerc" — a supermarket chain that advertizes on its low prices, "the greatest number," "People from the suburbs," and "the housewife under 50"). Even judgments on specific musical features, like the vocoder, can be associated with a more popular audience. These method of elicitation of reactions to the listening of precise music pieces (rather than traditional analysis of opinions about debatable and more abstract typologies – see the discussion) could provide substantial insights on the thesis that "dislikes are used to construct boundaries primarily around markers of disadvantaged class status" (Lizardo and Skiles 2016, p. 15).

Discussion

We have presented an original MMR design that sustains a collaborative, public-private research on online music consumption. This project has enabled the collection of mixed, diverse data from the same individuals, at an unprecedented scale. We have highlighted some traditional and more innovative use cases of the pairwise combinations of our three data sources – interview data, survey data, and digital activity traces collected by a music streaming platform. Our main contribution emerged from the combination of individual streaming logs, of observational nature, with traditional data collected in surveys and interviews, of declarative nature. Use cases include the identification of differences in content consumed between respondents and non-respondents ; the measure of the relationship between the social characteristics of respondents and their actual, everyday listening practices ; and the algorithmically driven elicitation of life events as well as aesthetic and social judgements. Another wave of data collection, scheduled for 2024, will attempt to test on a larger scale some exploratory hypotheses and heuristics that have emerged along the way.

Contributions to sociological puzzles

This design makes it possible to address four general questions in sociology.

The social characterization of actual practices on a large scale. The joint use of social characteristics and digital traces is a promising venue for the multidimensional analysis of the relationships between social position variables and behavior (see Figure 4). This canvas could

also nourish studies on the intersecting effects between users' social characteristics and platforms' recommendations (editorial and algorithmic). Differences in listening practices and music listened to between respondents and non-respondents raise important warnings regarding the generalizability of conclusions drawn from random sample surveys on cultural practices (see Figure 3), especially if they are based on self-administered questionnaires, or when the response rate is low or non-communicated. In addition to the problems of inaccuracy of self-reports as a proxy of actual practice, it seems that the music streamed by individuals who are likely to participate in surveys is not representative of their age class and gender group.

The empirical building of categories from the ground up. The aforementioned case of classical music illustrates a more general problem in sociology, that is defining and evaluating content categories (e.g. typologies of occupations or academic disciplines). In this example the strong over-declaration of classical music by users, once compared with their actual streaming practices, led us to suspect a tendency of users to listen to that genre on other media than Deezer. After modifying the survey questionnaire to account for this possibility, we discovered that classical music is indeed the genre that lends itself most to listening on other media and technical configurations among respondents. However, even if we restrict our analysis to the responses of people who declared that they listen to classical music mainly on Deezer, it remains the music genre that is the most over-reported. The main theories in sociology of culture (structural homology, omnivorism, snobivorism, etc.) are generally assessed on the ground of survey data analyses, that relate social position indicators with the self-declared listening and preferences for music genres. Such simplifications are problematic, as genres are very volatile categories, encompassing very heterogeneous products, with no consensus in the criteria or procedure to attribute one or several genre labels to a given song (Lena and Peterson 2008, Savage and Gayo 2011). On the other hand, genre categories, institutionalized in various ways and used by many to define their own consumption or productions, do capture social differences. For instance, our analysis of artists consumption suggests significant social differences in the audiences profiles for artists strongly associated with the genres of French rap or American rock. The challenge of producing large scale, robust, and reproducible categorization of music in line with existing typologies is then crucial. It is indeed necessary to connect current and past work, and ensure more cumulative results. Our team has been working on the elaboration of multiple complementary typologies that can be translated into well-defined, algorithmic procedures. For example, a given genre can be operationalized (i) as the content broadcasted by professional radio stations dedicated to this genre; (ii) as the songs that are over-represented in personal or editorial playlists named after this genre, or (iii) as the artists mentioned on pages dedicated to the genre in encyclopedic platforms or music databases.

Many other solutions could be implemented, and we are convinced that important theoretical improvement will emerge, if researchers use multiple competing typologies and test the sensitivity of their results to small fluctuations in the definition of these genre categories. In addition, a database of exogenous indicators of cultural legitimacy is being constituted, by systematically collecting reviews from several websites in the cultural and music press, as well as from sites aggregating Internet users' ratings.

The algorithmic unveiling of preferences through experimental settings. Our preliminary experimentation on reactions to music exposure in interviews has highlighted the potential of such elicitations for getting to "gut reactions" that reveal entrenched social judgements. This kind of experiment could be developed in the study of various phenomena. We plan to systematize this process in the next wave's online survey. We will include a series of musical excerpts selected algorithmically, on the basis of the user's streaming history, and will record their appreciations both explicitly (via open-ended and scaled questions) and indirectly (by measuring, for example, the duration before they stop listening to the extract). In addition, the preliminary results we presented (see Figure 4) suggest a relationship between individuals' social properties and the music they listen to. If such relationship were to be strong and statistically robust, an important research avenue would be to test the accuracy of predictions of the social properties of any user (including non-respondents) from their cultural consumption data (e.g. Hinds and Joinson 2018, Krismayer et al. 2019).

The detection of social norms. More generally, a notable advance provided by this design is the detection of empirical interstices where different sources of data highlight what seems to be contradictions between what people declare doing, what we observe them doing, and how they describe these activities. These diverging information, often conceived in the literature as biases that MMR designs can help correct, seem in fact to convey important sociological meaning. In some cases at least, we observed that these discrepancies between observations, declarations, and descriptions constitute an invaluable locus to investigate social norms, as these situations underscore tensions between internalized hierarchies and actual behavior.

Considerations on private-public research partnerships

Finally we highlight how other topics could benefit from a similar research design, while underlining the challenges of this kind of collaboration.

The successful implementation of this MMR design depended on many factors. Three of them were (i) setting up a partnership with a company whose research department was willing to run an extensive sociological survey

on its users ; (ii) obtaining a budget, including incentives to motivate people to participate in interviews, especially those who are the less genuinely disposed to do so ; and (iii) an interdisciplinary collaboration framework, over the course of several years. The collective nature of this work is what has enabled us to develop an original contribution to the MMR literature. None of the participants had the time, knowledge or know-how to carry out all the steps of this research on their own. While still rare in sociology, this research mode was more akin to that routinely used in other sciences. We believe that the study of social phenomena would benefit from giving greater prominence to these long-term collaborations, which are more costly than individual research, both in terms of the material resources, but also and above all in terms of the time devoted to research by tenured professors and researchers. At a time when private companies invest massive resources to gather information on human behavior and possibly influence our representations and actions, we feel the urgency of more social scientists with stable working conditions collaborating to share the public good of scientific knowledge.

On the other hand, an important stake to point out relates to the conditions and consequences of collaborations with private companies. In such configurations, public institutions and scholars should proceed with care. This kind of collaboration allows the company to collect precise, declarative personal and social data, and it is likely that they could not afford to collect them on their own, were they not supported in doing so by legitimate public research institutions. A possible use case of such mixed data could be to support the training of large scale statistical models to relate people's social properties with their music consumption. Knowing that such tools could be incorporated in algorithmic recommendation, should sociologists refrain from collecting such data in collaboration with a company? Another point of view is to consider that collecting such data is necessary for better describing, understanding and ultimately producing public information about the social stratification of cultural consumption.

Beyond music listening and preferences, a research design that combines the integration of longitudinal digital traces with surveys and interviews ; a nested-sampling of the different groups ; and a wave-based iterative approach, could be reproduced in different fields of study. In recent years, individual ordinary practices have been increasingly commodified, through internet platforms that collect people's digital data. Examples of social practices that could be investigated with such a MMR approach include eating habits, with the development of online food shopping (Jenneson et al. 2022); online dating (Bergström 2022, Lewis 2013); visiting health-care professionals (platforms for online medical appointments), or education pathways via centralized selection platforms (Frouillou et al. 2020).

Acknowledgements

The authors thank Noé Latreille de Fozières who conducted several of the individual interviews carried out for the project; and Marion Baranes, Boris Beaude, Laurent Beauguitte, Thomas Bouabça, Paul Chapron, Hadrien Commenges, Clément Dargent, Marion Maisonobe, Jérémie Poiroux and Clémence Saillard for their input when designing the first wave's version of the survey questionnaire. For discussions that have undoubtedly contributed to the development of this research, we also thank the participants and organizers of the research seminars series where we've been invited to present this work at an earlier stage: the sociology departments of Sciences Po Paris and Grenoble; the French Ministry of Culture's Department of Studies, Foresight, Statistics and Documentation (DEPS).

Author Contributions

All authors designed the study and the online survey questionnaire. MM, RC and TL implemented the questionnaire. MM coordinated the distribution of the online survey and provided the anonymous streaming data. AB, BLB, JSB, RC, SC and YR designed and conducted the interviews. JSB coordinated the interviews. RC designed and implemented the visualization app used during the interviews. TL and YR analyzed the streaming history and survey data, and prepared the figures. AB, TL and YR drafted the manuscript. All authors commented and approved the final version of the manuscript. CR, JSB, MM, PC and TL coordinated the study.

Competing Interests

The authors declare that they have no competing interests.

Funding and licence

This paper has been realized in the framework of the 'RECORDS' grant (ANR-2019-CE38-0013) funded by the ANR (French National Agency of Research). A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

- E. Bakshy, S. Messing, and L. A. Adamic. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239):1130–1132, June 2015. doi: [10.1126/science.aaa1160](https://doi.org/10.1126/science.aaa1160).
- G. Bastin and P. Tubaro. Le moment big data des sciences sociales. *Revue française de sociologie*, 59(3):375–394, 2018. ISSN 0035-2969. doi: [10.3917/rfs.593.0375](https://doi.org/10.3917/rfs.593.0375).
- T. Bennett, M. Savage, E. Silva, A. Warde, M. Gayo-Cal, and D. Wright. *Culture, Class, Distinction*. Taylor & Francis, Londres, 2009.
- M. Bergström. *The New Laws of Love: Online Dating and the Privatization of Intimacy*. Medford: Polity Press, 2022. ISBN 978-1-5095-4352-6.
- J.-S. Beuscart, S. Coavoux, and S. Maillard. Les algorithmes de recommandation musicale et l’autonomie de l’auditeur. *Réseaux*, 1(213):17–47, Mar. 2019. ISSN 0751-7971. doi: [10.3917/res.213.0017](https://doi.org/10.3917/res.213.0017).
- J. Blasius and M. Brandt. Representativeness in Online Surveys through Stratified Samples. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 107(1):5–21, July 2010. ISSN 0759-1063. doi: [10.1177/0759106310369964](https://doi.org/10.1177/0759106310369964).
- P. Bourdieu. *La Distinction. Critique Sociale Du Jugement*. Minuit, Paris, 1979. ISBN 2-7073-0275-9.
- D. Boyd and K. Crawford. Critical Questions for Big Data. *Information, Communication & Society*, 15(5):662–679, June 2012. ISSN 1369-118X. doi: [10.1080/1369118X.2012.678878](https://doi.org/10.1080/1369118X.2012.678878).
- A. Bryman. Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1):97–113, Feb. 2006. ISSN 1468-7941. doi: [10.1177/1468794106058877](https://doi.org/10.1177/1468794106058877).
- B. Bryson. What about the univores? Musical dislikes and group-based identity construction among Americans with low levels of education. *Poetics*, 25(2):141–156, Nov. 1997. ISSN 0304-422X. doi: [10.1016/S0304-422X\(97\)00008-9](https://doi.org/10.1016/S0304-422X(97)00008-9).
- R. Burrows and M. Savage. After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, 1(1):2053951714540280, Apr. 2014. ISSN 2053-9517. doi: [10.1177/2053951714540280](https://doi.org/10.1177/2053951714540280).
- P. L. Carter. Straddling Boundaries: Identity, Culture, and School. *Sociology of Education*, 79(4):304–328, Oct. 2006. ISSN 0038-0407. doi: [10.1177/003804070607900402](https://doi.org/10.1177/003804070607900402).
- V. L. P. Clark and J. W. Creswell. *The Mixed Methods Reader*. SAGE, 2008. ISBN 978-1-4129-5144-9.
- P. Coulangeon and Y. Lemel. Is “Distinction” really outdated? Questioning the meaning of the omnivorization of musical taste in contemporary France. *Poetics*, 35(2-3):93–111, 2007. ISSN 0304422X. doi: [10.1016/j.poetic.2007.03.006](https://doi.org/10.1016/j.poetic.2007.03.006).
- R. Cura, A. Beaumont, J.-S. Beuscart, S. Coavoux, N. Latreille de Fozières, B. Le Bigot, Y. Renisio, M. Moussallam, and T. Louail. Uplifting Interviews in Social Science with Individual Data Visualization: The case of Music Listening. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, CHI EA ’22, pages 1–9, New York, NY, USA, Apr. 2022. Association for Computing Machinery. ISBN 978-1-4503-9156-6. doi: [10.1145/3491101.3503553](https://doi.org/10.1145/3491101.3503553).
- O. Donnat. *Les pratiques culturelles des Français: enquête 1997*. Ministère de la culture et de la communication, Département des études et de la prospective la Documentation française, Paris, 1998. ISBN 978-2-11-003991-0.
- O. Donnat. *Les Pratiques Culturelles Des Français à l’ère Numérique: Enquête 2008*. La Découverte ; Ministère de la culture et de la communication, Paris, 2009. ISBN 978-2-7071-5800-0 978-2-11-097539-3.
- A. Eck, A. L. C. Cazar, M. Callegaro, and P. Biemer. “Big Data Meets Survey Science”. *Social Science Computer Review*, 39(4):484–488, Aug. 2021. ISSN 0894-4393. doi: [10.1177/0894439319883393](https://doi.org/10.1177/0894439319883393).
- L. Frouillou, C. Pin, and A. van Zanten. Les plateformes APB et Parcoursup au service de l’égalité des chances ? L’évolution des procédures et des normes d’accès à l’enseignement supérieur en France. *L’Année sociologique*, 70(2):337–363, 2020. ISSN 0066-2399. doi: [10.3917/anso.202.0337](https://doi.org/10.3917/anso.202.0337).
- J. C. Greene, V. J. Caracelli, and W. F. Graham. Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, 11(3):255–274, Sept. 1989. ISSN 0162-3737. doi: [10.3102/01623737011003255](https://doi.org/10.3102/01623737011003255).
- T. Grüne. The Problems of Testing Preference Axioms with Revealed Preference Theory. *Analyse & Kritik*, 26(2):382–397, Nov. 2004. ISSN 2365-9858. doi: [10.1515/auk-2004-0204](https://doi.org/10.1515/auk-2004-0204).
- E. Hargittai, A. M. Piper, and M. R. Morris. From internet access to internet skills: Digital inequality among older adults. *Universal Access in the Information Society*, 18(4):881–890, Nov. 2019. ISSN 1615-5289. doi: [10.1007/s10209-018-0617-5](https://doi.org/10.1007/s10209-018-0617-5).
- L. R. Harris and G. T. L. Brown. Mixing interview and questionnaire methods: Practical problems in aligning data. *Practical Assessment, Research, and Evaluation*, 15(1), Nov. 2019. ISSN 1531-7714. doi: [10.7275/959j-ky83](https://doi.org/10.7275/959j-ky83).
- J. Hinds and A. N. Joinson. What demographic attributes do our digital footprints reveal? A systematic review. *PLOS ONE*, 13(11):e0207112, Nov. 2018. ISSN 1932-6203. doi: [10.1371/journal.pone.0207112](https://doi.org/10.1371/journal.pone.0207112).
- V. L. Jenneson, F. Pontin, D. C. Greenwood, G. P. Clarke, and M. A. Morris. A systematic review of supermarket automated electronic sales data for population dietary surveillance. *Nutrition Reviews*, 80(6):1711–1722, June 2022. ISSN 0029-6643. doi: [10.1093/nutrit/nuab089](https://doi.org/10.1093/nutrit/nuab089).
- C. Jerolmack and S. Khan. Talk Is Cheap Ethnography and the Attitudinal Fallacy. *Sociological Methods & Research*, pages 178–209, 2014. ISSN 0049-1241, 1552-8294. doi: [10.1177/0049124114523396](https://doi.org/10.1177/0049124114523396).
- T. Katz-Gerro. Cultural Consumption and Social Stratification: Leisure Activities, Musical Tastes, and Social Location. *Sociological Perspectives*, 42(4):627–646, Dec. 1999. ISSN 0731-1214. doi: [10.2307/1389577](https://doi.org/10.2307/1389577).
- T. Krismayer, M. Schedl, P. Knees, and R. Rabiser. Predicting user demographics from music listening information. *Multimedia Tools and Applications*, 78(3):2897–2920, Feb. 2019. ISSN 1573-7721. doi: [10.1007/s11042-018-5980-y](https://doi.org/10.1007/s11042-018-5980-y).

- I. Krumpal. Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4): 2025–2047, June 2013. ISSN 1573-7845. doi: [10.1007/s11135-011-9640-9](https://doi.org/10.1007/s11135-011-9640-9).
- M. Lamont and A. Swidler. Methodological Pluralism and the Possibilities and Limits of Interviewing. *Qualitative Sociology*, 37(2):153–171, June 2014. ISSN 1573-7837. doi: [10.1007/s11133-014-9274-z](https://doi.org/10.1007/s11133-014-9274-z).
- R. T. LaPiere. Attitudes vs. Actions. *Social Forces*, 13(2): 230–237, 1934. ISSN 0037-7732. doi: [10.2307/2570339](https://doi.org/10.2307/2570339).
- D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. Computational Social Science. *Science*, 323(5915):721–723, Feb. 2009. ISSN 0036-8075, 1095-9203. doi: [10.1126/science.1167742](https://doi.org/10.1126/science.1167742).
- P. Le Guern, editor. *En quête de musique: Questions de méthode à l'ère de la numérimorphose*. Hermann, Paris, Feb. 2017. ISBN 978-2-7056-9300-8.
- E. Leahey. Convergence and confidentiality? Limits to the implementation of mixed methodology. *Social Science Research*, 36(1):149–158, Mar. 2007. ISSN 0049-089X. doi: [10.1016/j.ssresearch.2005.10.003](https://doi.org/10.1016/j.ssresearch.2005.10.003).
- N. L. Leech and A. J. Onwuegbuzie. A typology of mixed methods research designs. *Quality & Quantity*, 43(2):265–275, Mar. 2009. ISSN 1573-7845. doi: [10.1007/s11135-007-9105-3](https://doi.org/10.1007/s11135-007-9105-3).
- J. C. Lena and R. A. Peterson. Classification as Culture: Types and Trajectories of Music Genres. *American Sociological Review*, 73(5):697–718, Oct. 2008. ISSN 0003-1224. doi: [10.1177/000312240807300501](https://doi.org/10.1177/000312240807300501).
- K. Lewis. The limits of racial prejudice. *Proceedings of the National Academy of Sciences*, 110(47):18814–18819, Nov. 2013. ISSN 0027-8424, 1091-6490. doi: [10/15hjzw](https://doi.org/10/15hjzw).
- K. Lewis, M. Gonzalez, and J. Kaufman. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1):68–72, Jan. 2012. ISSN 0027-8424, 1091-6490. doi: [10.1073/pnas.1109739109](https://doi.org/10.1073/pnas.1109739109).
- E. S. Lieberman. Nested Analysis as a Mixed-Method Strategy for Comparative Research. *American Political Science Review*, 99(3):435–452, Aug. 2005. ISSN 1537-5943, 0003-0554. doi: [10.1017/S0003055405051762](https://doi.org/10.1017/S0003055405051762).
- O. Lizardo and S. Skiles. Cultural Objects as Prisms: Perceived Audience Composition of Musical Genres as a Resource for Symbolic Exclusion. *Socius*, 2:2378023116641695, Jan. 2016. ISSN 2378-0231. doi: [10.1177/2378023116641695](https://doi.org/10.1177/2378023116641695).
- P. Lombardo and L. Wolff. Cinquante ans de pratiques culturelles en France. *Culture études*, n°2(2):1, 2020. ISSN 1959-691X, 2118-4674. doi: [10/ghhr38](https://doi.org/10/ghhr38).
- T. Louail and M. Barthelemy. Headphones on the wire : Statistical patterns of music listening practices. *arXiv:1704.05815 [physics]*, Apr. 2017. doi: [10.48550/arXiv.1704.05815](https://doi.org/10.48550/arXiv.1704.05815).
- J. A. Maxwell. Expanding the History and Range of Mixed Methods Research. *Journal of Mixed Methods Research*, 10(1):12–27, Jan. 2016. ISSN 1558-6898. doi: [10.1177/1558689815571132](https://doi.org/10.1177/1558689815571132).
- J. W. Mohr, C. A. Bail, M. Frye, J. C. Lena, O. Lizardo, T. E. McDonnell, A. Mische, I. Tavorly, and F. F. Wherry. *Measuring Culture*. Columbia University Press, Aug. 2020. ISBN 978-0-231-54258-6.
- N. Moureau and M. Vidal. Quand préférences déclarées et révélées s'opposent : le téléspectateur, un cas paradoxal pour l'économiste. *Revue Française de Socio-Économie*, 4(2):199–218, 2009. ISSN 1966-6608. doi: [10.3917/rfse.004.0199](https://doi.org/10.3917/rfse.004.0199).
- C. Noûs, N. Robette, and O. Roueff. La structure multidimensionnelle des goûts. Une échelle de légitimité culturelle fondée sur les interactions entre diplôme, âge et sexe. *Biens Symboliques / Symbolic Goods*, (8), 2021. ISSN 2490-9424. doi: [10.4000/bssg.598](https://doi.org/10.4000/bssg.598).
- É. Ollion and J. Boelaert. Au-delà des big data. Les sciences sociales et la multiplication des données numériques. *Sociologie*, 6(3):295–310, 2015. ISSN 2108-8845.
- J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18): 7332–7336, May 2007. ISSN 0027-8424, 1091-6490. doi: [10.1073/pnas.0610245104](https://doi.org/10.1073/pnas.0610245104).
- L. A. Palinkas, S. M. Horwitz, C. A. Green, J. P. Wisdom, N. Duan, and K. Hoagwood. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and policy in mental health*, 42(5):533–544, Sept. 2015. ISSN 0894-587X. doi: [10.1007/s10488-013-0528-y](https://doi.org/10.1007/s10488-013-0528-y).
- D. A. Parry, B. I. Davidson, C. J. R. Sewall, J. T. Fisher, H. Mieczkowski, and D. S. Quintana. A Systematic Review and Meta-Analysis of Discrepancies Between Logged and Self-Reported Digital Media Use. *Nature Human Behaviour*, 5(11):1535–1547, 2021. doi: [10.1038/s41562-021-01117-5](https://doi.org/10.1038/s41562-021-01117-5).
- L. D. Pearce. Mixed Methods Inquiry in Sociology. *American Behavioral Scientist*, 56(6):829–848, June 2012. ISSN 0002-7642, 1552-3381. doi: [10.1177/00027642111433798](https://doi.org/10.1177/00027642111433798).
- V. Pérez Bentancur and L. Tiscornia. Iteration in Mixed-Methods Research Designs Combining Experiments and Fieldwork. *Sociological Methods & Research*, page 00491241221082595, Mar. 2022. ISSN 0049-1241. doi: [10.1177/00491241221082595](https://doi.org/10.1177/00491241221082595).
- A. Prieur, L. Rosenlund, and J. Skjott-Larsen. Cultural capital today. *Poetics*, 36(1):45–71, 2008. ISSN 0304422X. doi: [10.1016/j.poetic.2008.02.008](https://doi.org/10.1016/j.poetic.2008.02.008).
- M. Prior. The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, 73(1):130–143, 2009. ISSN 1537-5331, 0033-362X. doi: [10.1093/poq/nfp002](https://doi.org/10.1093/poq/nfp002).
- Y. Renisio. *L'infortune Des Sciences Sociales : Sociologie d'une Illégitimation Scientifique Récurrente*. These de doctorat, Paris, EHESS, June 2017.
- Y. Renisio and R. Sinthon. L'analyse des correspondances multiples au service de l'enquête de terrain. Pour en finir avec le dualisme "quantitatif"/"qualitatif". *Genèses*, 97(4):109–125, Oct. 2014. ISSN 1155-3219. doi: [10.3917/gen.097.0109](https://doi.org/10.3917/gen.097.0109).

- W. S. Robinson. Ecological Correlations and the Behavior of Individuals. *American Sociological Review*, 15(3):351–357, 1950. ISSN 0003-1224. doi: [10.2307/2087176](https://doi.org/10.2307/2087176).
- M. J. Salganik. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press, Princeton, 2018. ISBN 978-0-691-15864-8.
- M. Savage and M. Gayo. Unravelling the omnivore: A field analysis of contemporary musical taste in the United Kingdom. *Poetics*, 39(5):337–357, 2011. doi: [10.1016/j.poetic.2011.07.001](https://doi.org/10.1016/j.poetic.2011.07.001).
- A. Sen. Behaviour and the Concept of Preference. *Economica*, 40(159):241–259, 1973. ISSN 0013-0427. doi: [10.2307/2552796](https://doi.org/10.2307/2552796).
- E. Silva, A. Warde, and D. Wright. Using Mixed Methods for Analysing Culture: The Cultural Capital and Social Exclusion Project. *Cultural Sociology*, 3(2):299–316, July 2009. ISSN 1749-9755, 1749-9763. doi: [10.1177/1749975509105536](https://doi.org/10.1177/1749975509105536).
- S. Stier, J. Breuer, P. Siegers, and K. Thorson. Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, Apr. 2019. doi: [10/gf3xxw](https://doi.org/10/gf3xxw).
- O. Sullivan and T. Katz-Gerro. The omnivore thesis revisited. Voracious cultural consumers. *European Sociological Review*, 23(2):123–137, Dec. 2006. doi: [10.1093/esr/jcl024](https://doi.org/10.1093/esr/jcl024).
- I. Tucci, J. J. Fröhlich, and I. Stock. Exploring the Nexus between Migration and Social Positions using a Mixed Methods Approach. *Social Inclusion*, 9(1):114–129, Feb. 2021. ISSN 2183-2803. doi: [10.17645/si.v9i1.3538](https://doi.org/10.17645/si.v9i1.3538).
- Q. Villermet, J. Poiroux, M. Moussallam, T. Louail, and C. Roth. Follow the guides: Disentangling human and algorithmic curation in online music consumption. In *RecSys'21: Fifteenth ACM Conference on Recommender Systems*, RecSys '21: Fifteenth ACM Conference on Recommender Systems, pages 380–389. Association for Computing Machinery (ACM), Sept. 2021. doi: [10.1145/3460231.3474269](https://doi.org/10.1145/3460231.3474269).
- J. Webster. Music on-demand: A commentary on the changing relationship between music taste, consumption and class in the streaming age. *Big Data & Society*, 6(2):1–5, 2019. ISSN 2053-9517. doi: [10.1177/2053951719888770](https://doi.org/10.1177/2053951719888770).
- J.-M. Weller. Le mensonge d'Ernest Cigare : Problèmes épistémologiques et méthodologiques à propos de l'identité. *Sociologie du travail*, 36(1):25–42, 1994. doi: [10.3406/so-tra.1994.2155](https://doi.org/10.3406/so-tra.1994.2155).
- B. Zhang, G. Kreitz, M. Isaksson, J. Ubillos, G. Urdaneta, J. A. Pouwelse, and D. Epema. Understanding user behavior in Spotify. In *2013 Proceedings IEEE INFOCOM*, pages 220–224, Apr. 2013. doi: [10.1109/INFOCOM.2013.6566767](https://doi.org/10.1109/INFOCOM.2013.6566767).