The Harkive Project: popular music, data, and digital technologies

Craig Hamilton

Submitted in fulfillment of the requirements of the degree of Doctor of Philosophy to the faculty of Arts, Design & Media at Birmingham City University.

September 2017

Abstract

This thesis is about research around Harkive, an online project designed by this researcher, that gathers stories, reflections, and other data from people about their everyday engagement with popular music. Since 2013, over 1,000 people have contributed to the project, producing around 8,000 texts and highlighting the music reception activities of contemporary music listeners.

The thesis presents an analysis of the texts and other data generated, answering a key research question:

What can an analysis of the data generated by The Harkive Project reveal about the music reception practices of respondents?

To answer this question, the researcher developed an experimental, innovative approach that conceives of Harkive as a space in which people can reflect upon their engagement with music, whilst simultaneously acting as a place that is able to replicate many of the commercial practices related to data collection and processing that have recently emerged as influential factors in the ways that popular music is produced, distributed and consumed. By focusing on a set of findings about the way people reflect on their engagement with music within the Harkive space, this thesis engages practically and critically with these new conditions.

Simultaneously, the research explores how the systems of data collection and analysis that facilitate this are technologically complex, subject to rapid change, and often hidden behind commercial and legal firewalls, making the study of them particularly difficult. This then enables us to explore how the use of digital, data and Internet technologies by many people during the course of their everyday lives is providing scholars with new opportunities and methods for undertaking research in the humanities, and how this in turn is leading to questions about the role of the researcher in popular music studies, and how the discipline may take into account the new technologies and practices that have so changed the field.

Ultimately, the thesis makes the argument that a greater practical understanding and critical engagement with digital, data and Internet technologies is essential, both for music consumers and popular music scholars, and demonstrates how this work represents a significant contribution to this task.

Acknowledgements

I am grateful for the help and support offered to me by my supervisors, Prof. Nick Gebhardt and Prof. Tim Wall. Your guidance has been invaluable throughout this process, from exploring initial ideas for a PhD proposal back in 2013, right up until the final days and hours before submission.

I have been hugely fortunate to have my work funded by the AHRC Midlands3Cities Doctoral Training Partnership, so I extend my thanks to all involved with this. To all of the other M3C students I met along the way, I wish you all the best. I am very proud to have been part of such a talented group of people.

I have benefitted immensely from the support and encouragement offered to me by my colleagues in the Birmingham Centre for Media and Cultures Research (BCMCR) in the School of Media at Birmingham City University. Working alongside hugely clever and supportive people has made this process a lot easier than working alone. In particular I'd like to thank the members of the informal writing club I have been part of. The opportunity to meet regularly to discuss the process, problems, and joys of writing about music has been of tremendous help. To the scholars and researchers from other institutions I have met at conferences, from Cork to Berlin, or who have talked with me over email and Skype, thank you for the feedback and encouragement.

To Nick M and to Paul M: thank you for the technical assistance. To the various journalists, musicians, record label owners, technologists, artists, promoters, and those from many other walks of the musical life, thank you for writing about the project, providing promotional items, and spreading the word about my work. To all of the people who participated in The Harkive Project between 2013 and 2016, far too many to mention by name, I reserve a particularly special thank you. The response to the project exceeded all of my expectations, and this thesis

would not exist without your stories. I have enjoyed reading them, and I look forward many more years of collecting stories on sunny days in July.

To my friends and family, thank you for all the words of encouragement and support. In particular, thank you to the Independent Country boys (including, of course, Executive Producer Robson) for being my pressure valve and source of so much fun during this process.

I apologise to The Dude and Mimi for all the missed walks.

Finally, and most importantly of all, I should like to express my love and gratitude to my wife, Valerie, and my two sons, Mac and Bruce. You have supported me in so many ways during the completion of this work. I could not have done it without you. Thank you.

INTRODUCTION

Walking in the city with headphones on

Like millions of other people around the world I am passionate about music. It is something that is woven into the fabric of my everyday life. Music has helped and continues to help form my identity. It underwrites many of my friendships, and it is central to a lot of my activities, plans and memories. The exact moment when my relationship with music started is hard to pin point, but it feels very much as though it has been one of the true constants throughout my life. So much so that I am not entirely sure what I would do without it, and I hope that I never have to find out. Music fills my home, coming from the stereo in the lounge, the radio in the kitchen, or from the portable Bluetooth speaker that reaches the parts those static systems cannot. I often break headphones whilst gardening. Sometimes songs will just play in my head, cued up from the deep depths of my memory. Music can take me to weird and wondrous places, but even when it exists in the background it still feels somehow crucial. The total absence of it, or the wrong music at a given moment, can be jarring. In your own way, you are perhaps much the same. You and I are far from alone.

Whenever I leave my house I am almost always accompanied by music, thanks to portable devices and digital, data and Internet technologies. In the inevitable last-minute scramble to get out of the door, and along with ensuring that I have the mundane, essential things, like my keys, wallet, or whatever the other necessary items are for that day, I am always particularly careful to make sure that I remember to pick up my headphones, and that my iPhone is sufficiently charged to supply me with music on wherever my journey is taking me. A dead battery and no time to charge it can put me in a terrible mood. It disrupts my normal, preferred routine. I need music come with me when I walk my dogs. It has to accompany me around the supermarket. It provides the soundtrack for my occasional and sporadic forays into the world of running. And it comes with me when I am travelling to and from work at Birmingham City University.

One day in early November 2015 I was making that regular journey to BCU. Having completed the mental checklist of items I'd need, I successfully left the house on time and at the first attempt (not always a feat I accomplish!), complete with my headphones and a sufficiently charged phone. My choice for listening that day was the latest instalment of my Spotify Discover Weekly playlist, the automatically generated, personalised 30-song list delivered each Monday morning. As I walked down my street towards the train station, the playlist unfolded and in that particular week included the song 'Sexuality' by Billy Bragg. This was a song I knew well but had not listened to in many years.

Listening to this song led to me temporarily abandon the rest of the playlist and to follow instead the links within the Spotify mobile interface to the 1992 album from which the song came, 'Don't Try This At Home'. Continuing my walk to the station, I realised that apart from the records Billy Bragg made in the late 1990s and early 2000s with the American band Wilco – a fantastic collection of new songs based on Woody Guthrie lyrics – I had not really paid too much attention to Bragg's recent work. In fact, 'Don't Try This At Home' was probably the last of Billy Bragg's albums that I had properly paid attention to. At a rough estimate, it was something like 15 or 20 years since I had last listened to it.

As the album played, and I continued my walk, I was sent back to 1992, when I was 18 years of age and working in a record shop. I recalled buying an expensive, limited-edition version of the album (thanks to my dealer price staff discount) on a very strange format: all 16 songs spread across eight different 7" singles, which together came packaged in a red presentation box embossed in gold with Bragg's logo. I pondered how even at the time this was an unwieldy format for listening to an album, certainly in comparison to the connected mobile device I was presently using to listen to the very same record. I remembered also that – because of the impractical nature of my purchase – I had at around the same time borrowed a CD copy from the shop (as we were unofficially allowed to do) in order to tape it at home. Most of my listening to the album had been — certainly up until this point — from that copied tape version. I realised also that I still had

the tape. It is in a box in my loft at home, along with hundreds of others. For many years I've had the vague idea of bringing that box down from the loft and exploring its contents, but there never seems to be the time. The thought of retrieving that box of cassettes from the loft struck me once again. I then remembered that I still have that impractical 7" album box set. It has remained largely un-played over the years.

As these thoughts struck me, now arrived at the station and waiting for my train, I also began to remember other things related to Billy Bragg, and from my time working in that record shop. I remembered going to a Bragg gig in Birmingham with my then-colleagues on the eve of the 1992 British General Election. During the show, the Labour MP Roy Hattersley had joined Billy onstage, and together they gave speeches that had generated a wave of hope and optimism in the room. This had evaporated a mere 48 hours later when the election was lost, but the memory was nevertheless a fond one. My train pulled into the station.

As well as this wave of memories and nostalgia prompted by the music, stepping aboard the train I realised that I was also approaching the album from a different but simultaneous vantage point. I am no longer an 18 year-old youth; I am a 40-something man, a husband, and a father to two young children. Beyond the box set gathering dust in the numerous places I've lived over the years since I bought it, a great deal else has happened since I taped that CD back in 1992.

As the album continued to play, and along with hearing the uplifting 'Sexuality' once again, I heard a few long-forgotten, lost-love songs ("I saw them in the hardware store. He looked boring and she looked bored" (Bragg, 1992), which made me smile. I also met up again with other songs that I was hearing again for the first time in many years. In particular the song 'Tank Park Salute' – about the death of a father – provoked a surprisingly strong, emotional reaction in me as I rode on the packed 7.52am train into Birmingham New Street. Maybe it was the inherent sadness of the song, or perhaps that it made me think of my own children, which can sometimes lead me to thoughts of my mortality, or maybe it was all of these things coming so quick and fast after the fond, youthful nostalgia.

Or maybe it was simply that, because I rarely eat breakfast, my blood sugar was low and I was unusually susceptible to a particularly sad song. But whatever the reason, I suddenly felt very emotional as that song played. So much so, in fact, that rather embarrassingly I found myself close to tears, and had to try very hard not to let this show in front of a train carriage full of complete strangers. The shock at my reaction to the song – an involuntary welling up at a pop record, in public – was almost as unnerving as the reaction itself.

Discussing the incident with friends later that day it seems Tank Park Salute has elicited similar reactions from others over the years. One friend had cried on an airplane, prompting the cabin crew to check on his wellbeing. Another took the opportunity to joke that hearing Billy Bragg's somewhat abrasive singing voice could too bring them to tears. Later still, and sufficiently calm now to properly reflect on what had happened, what I found interesting about the incident was that my wave of nostalgia and narrowly averted public emotional collapse were, in part at least, prompted by the functional operation of an algorithm. Was it possible that an algorithm could make me cry?

Perhaps not, but had Spotify Discover Weekly not provided 'Sexuality' in my list that particular week, played by me on that particular day, it is highly unlikely – with millions of records to choose from at the swipe of a finger – that I would have found my way back to 'Don't Try This At Home', and subsequently to 'Tank Park Salute' and my emotional reaction. Of course, the exact same thing could have happened had 'Sexuality' been played on the radio, or mentioned by someone in my Twitter feed, or even if it had just involuntarily popped into my head, for reasons unknown, as songs have a habit of doing.

However, this leads to the crucial point about the incident: 'Sexuality' was not played or mentioned in a public forum, nor was it generated by the internal mechanisms of my memory. It was situated instead in a 'personalised' playlist, one that had been generated by a combination of humans and machines working together to perform analysis on data collected about me, about metadata associated with songs, and about the actions of many, many other people. It was –

in a peculiar way – a form of collaboration between people and things. In a paradoxical sense, however, it was not a shared experience, or a public one – it was a private and 'personalised' one that had its basis in the public broadcast of mine and many others' listening habits. This felt, to me, like a subtle but very important shift in how we collectively and individually engage with popular music. Ultimately, this extraordinary incident on a very ordinary train journey, helped to crystallise and prompt many of the questions that have contributed to the development of this thesis.

I began to think, for instance, about data collection and algorithmic analysis, and the role that technologies and practices associated with them play in the services, media channels and interfaces many of us interact with on a daily basis. I became interested in questions about the ways in which these activities, taken together, generate forms of cultural knowledge, and how this in turn may be involved to an extent in producing the experiences we have. I began to think also about how this spills over into our experiences with popular music more generally, back in the 'real world', exactly as it had done with me that morning on the train. I realised that I wanted to find a way of discovering the answers to these questions.

****

In February 2011, almost twenty years after I had first started working in that Birmingham record shop, and some seventeen years after completing a BA degree I had not taken particularly seriously, I returned to education and enrolled on the MA in Music Industries at Birmingham City University. Through the intervening years I had spent the early and mid-1990s working in several more record shops, this time all over London. A large part of that job involved buying collections of vinyl records from members of the public. They were abandoning the vinyl format in favour of the CD, and – several years before eBay and Discogs – this it was something of a golden period for the vinyl collector.

As a fan of records, I was never particularly enamoured with the CD. I bought a lot of those records that came across the counter during my time at the record shop. I still do buy a lot of records, but I wish I'd bought more of them back then. A lot of those records that were unwanted and practically worthless in the 1990s are now incredibly rare and expensive; vinyl has made an unexpected cultural and financial comeback. The endurance of older formats and associated cultural practices associated with them is another element of the story of this thesis, but we will get to that in due course.

From the record shop I eventually moved on to work in the Internet industry, in the late 1990s, and learned a little about how things worked in the online world, before combining the skills and experience I had accumulated to work in the nascent digital music distribution business, in 2004. Out of the initial handful of what became known as 'aggregators' of digital music that sprang up following the launch of Apple's iTunes store in 2003, two had set up shop in my hometown. Here I experienced first-hand the recorded music industries in transition from the physical to the digital world. I recall almost daily conversations with experienced label bosses and independent musicians. They had very little idea about how music in the online space was going to work. Many did not know how MP3 files were created, few understood the importance of the metadata they needed to contain, and fewer still had a grasp on how these endlessly replicable computer files would turn in to revenue. The label bosses and the musicians were not alone in this regard. The landscape of music online was still developing and it was developing so quickly that no one really knew the direction of travel. It was whilst in that job that I first saw YouTube, Last.FM, and several other services emerge that were to become closely associated with music online as it became commonplace over the next decade. It was an exciting time, and I learned a lot.

Along the way I became involved with the Birmingham Centre for Media and Cultural Research at BCU, who at the time were researching many of the changes in the landscape of popular music that I was working through on a day-to-day basis. It was here that the idea of studying the changes we were living through,

which fascinated me as much as they did the researchers at BCU, began to take hold within me for the first time. After completing some work with BCU as part of a Knowledge Transfer Partnership, and delivering some guest lectures around music in the online space, the chance to enrol on their new MA in Music Industries seemed a logical step. I was, at this point, a stay-at-home dad and was fortunate to be able to balance the commitment required of a part-time MA with looking after things at home. I don't think I would be writing this now had I not made that jump – it remains one of the best decisions I ever made.

It was during my MA studies that I developed The Harkive Project, which has subsequently developed into this thesis. Harkive was driven by me wanting to understand, document, and research the changes I had seen unfolding in the business of popular music, in technology, and in terms of being a music lover. I struck upon the idea that asking people who were perhaps like me, who had music woven into the fabric of their everyday lives, might be an interesting way of doing that. I devised a project that asked people to tell the story of how, where and why they listened to music and called it Harkive. To my immense surprise and satisfaction, thousands of people responded with stories and reflections about their experiences. It is the manner in which I have attempted to understand, analyse and think about those stories that forms the methodological and argumentative backbone of the work I present here. It has been a long and interesting journey from the record shop.

****

Over the last two decades digital, data and Internet and technologies have emerged as an important and influential factor in how popular music is produced, distributed and consumed. I am not the only person who has been on a journey that started in a record shop. These technologies, allied to practices of data collection and computational analysis, now play a significant role both in how audiences engage with music, and how those audiences are understood. A key point here is that popular music audiences are now highly individualised, and defined according to a growing number of new categorical variables. At the

same time they are also understood through the agglomeration of data points in a manner that recalls earlier conceptions of mass audiences. These intriguing new conditions invite us to revisit questions that have concerned popular music scholars for over 80 years, including issues of choice, agency, ownership, how audiences are constructed and understood, and how people derive meaning from popular music.

However, the systems of data collection and analysis that facilitate this are technologically complex, subject to rapid change, and often hidden behind commercial and legal firewalls. This makes the study of them particularly difficult. At the same time, the use of digital and data technologies by many people during the course of their everyday lives is providing scholars with new opportunities and methods for undertaking research in the humanities. This in turn is leading to questions about the role of the researcher in popular music studies, and how the discipline may take into account the new technologies and practices that have so changed the field. These are the inter-related issues my work addresses.

At the heart of my thesis is a piece of research on the music reception practices of contemporary music listeners, but what I then go on to do extends far beyond answering a single question, and into issues of research methodology and even the conceptualisation of music culture. My work, though, starts with a simple research question: What can an analysis of the data generated by The Harkive Project reveal about the music reception practices of its respondents?

To answer this question, I have developed an experimental, innovative approach that conceives of Harkive as a space in which people can reflect upon their engagement with music, whilst simultaneously acting as a place that is able to replicate many of the commercial practices related to data collection and processing. Through this space, I critically engage with the growing role of data and digital technologies associated with music consumption, whilst exploring the use of computational techniques in popular music studies research. The specific means by which this approach enables me to answer my research question can

be understood by considering the processes through which Harkive gathers text-based descriptions of music reception activities, the 'metadata' that accompanies those texts, the qualitative data gathered from a music listening survey, and the additional data produced through the use of computational analytical processing, including unsupervised machine learning algorithms. This means that the data about music reception activities available to me can be understood and analysed in a number of different ways, ranging from close readings of texts more usually associated with humanities research, through to the clustering, visualisation and analysis of abstractions generated through computational/algorithmic processing that renders the original texts as data. The method also allows for analyses that combine these approaches. Together they enable me to provide a number of answers to my central research question.

I show that Harkive respondents describe intriguing new cultural practices associated with music reception that can be understood as combinations of both new (digital) and existing (pre-digital) practices. For example, many of my respondents describe their use of digital interfaces in terms of vinyl record ownership, collection and use. Respondents also describe engaging in both 'new' and 'old' modes of engagement separately, and switch between all available modes with considerable dexterity. A number of the new conditions of engagement I have explored, specifically connectivity, digital interfaces, data-derived abstractions, and mobility, are also in evidence in respondent stories. This relates to the manner in which people switch between streaming and other online services and how they understand that process. There is very little evidence of Harkive respondents explicitly voicing similar concerns to those raised by a number of theorists regarding the potential impacts of the data collection and analysis – something I explore in my opening chapters. However, respondent narratives do contain acknowledgement of the role and function of these technologies, and it is in that context that interesting new questions arise.

The manner in which Harkive respondents also describe entering into a form of communication and/or relationship with data-derived abstractions (of their activities, of their selves, of available catalogues) indicates the importance of

undertaking further studies in this area. Of particular interest are automated recommender systems, the manner in which digital interfaces foreground (or not) content to audiences, and the new ways in which audiences are conceived of and organised, and how this relates to questions of choice, agency and identity. Relatedly, concerns over the potential consequences of data-collection and analysis that were not present in respondent narratives but were evident in the survey-gathering element of the process, suggests that further work in this area may also be fruitful.

In reflecting upon the issues and questions that have informed the development of my method, I consider how as a researcher I initially lacked the technical skills required to collect, prepare and analyse data in the manner I had identified as being of potential use. This project, then, became as much about how to conceive of new methods for studying the reception of music in the digital age. The approach I took drew on similar methods to those highlighted by the issues of debate above and are linked to both commercial practices in popular music, and to methods associated with the 'computational turn' (Berry, 2011; Hall, 2013) in humanities research. Data collection processes and computational techniques are shown, for instance, to be inherently reductive, which often prevents them from capturing and accurately reflecting complex cultural practices. In particular, text-based, qualitative data is a difficult form of data to process using computational methods and can lead to results that are problematic.

I also consider the extent to which the different modes of analysis afforded by my modular method have enabled me to arrive at different forms of insight that may not have been possible through methods usually associated with the humanities. Reflecting on these potential benefits and problems, I want to suggest that it is possible for popular music scholars to gain a better understanding of the new conditions of popular music's production, distribution and consumption through a combination of practical and critical engagement with processes of data collection and analysis. I suggest that the work I have undertaken in this project provides a springboard for that future work, and in particular for the creation of new tools, platforms, and research projects that

may enable both consumers and scholars to develop useful and productive epistemic responses to the role of digital, data and Internet technologies.

Ultimately, I argue that a greater practical understanding and critical engagement with digital, data and Internet technologies is possible, both for music consumers and popular music scholars, and I demonstrate how my work represents a significant step towards that.

***

In chapter 1 I begin by engaging with some of the key issues of debate highlighted above and explore some of the key features of contemporary cultural practice around music consumption. In particular I highlight the dominant position of digital, data and Internet technologies, and the structures and cultural practices associated with them. I then explore how we have arrived at this point through an historical examination of the development of the music industries, how audiences have been constructed and conceived, and how popular music scholars have addressed the various issues associated with these processes. This enables me to show how present day audiences have been individualised following developments over a 150-year period that initially saw them organised along broadly socio-economic lines, and then reorganized through smaller, sub-groups and according to more specific indicators such as age and taste. Central to this story are those technological developments that connect streaming to sheet music as part of general shift in the way that people engage in music reception. I also show that there is a parallel movement in popular music scholarship, which goes from theoretical perspectives of mass culture to sociological examinations of individual uses of music in everyday life. I argue that developments around digital, data and Internet technologies invite us to question what happens when audiences are constructed at the level of the individual.

In chapter 2 I focus specifically on the conditions of the present day by exploring the role and function of digital and data technologies in the cultural practices associated with contemporary music reception. Through a close examination of

algorithmic recommendation systems, digital interfaces, data collection, and the use of mobile technologies, I demonstrate their potential collective benefits and consequences in terms of listener agency, identity and choice. I argue that the growing influence of data-derived knowledge in many areas of our everyday life – and certainly in terms of how we engage with music – suggests that work seeking to examine the contemporary cultural, technological and commercial conditions requires us to engage both practically and critically with those processes.

My third chapter then explores the difficulties of such an approach. I examine the epistemological positions of the emerging disciplines of cultural analytics and digital humanities in order to relate these to contemporary popular music scholarship. In particular I explore the issues and benefits of what Alan Liu calls the 'polluting proximity' (2016) of using computational and algorithmic tools and methods in an examination of the role and consequences of digital and data technologies. This chapter makes the argument that these new modes of working provide a difficult but potentially useful route towards new knowledge and understanding, but only if elements of reflexivity and modularity are built in to the fabric of our methods and approaches.

I then apply this to the central methodological challenge of my research project, described in chapter 4, which has been to devise a means by which I can understand the large collection of texts and other data generated by Harkive. In developing a method that harnesses the potential of the data-derived techniques associated with the computational turn in the humanities (Berry, 2011), I demonstrate how I am able to arrive at a kind of insight that would not have been possible through close textual readings alone. Using these tools and analytical methods enables me to examine the role that digital and data technologies play in contemporary modes of engagement with popular music whilst at the same time examining and reflecting upon the ways these same analytical tools can help produce that understanding. I do this in order to demonstrate what we as researchers can learn from reflexively using these techniques, and how an understanding of those processes can help further

knowledge in the field of popular music studies. This is achieved in the chapters that follow by using the modular method described in this chapter to address my central research question.

By deploying the method developed above in chapter 5, the first of my three findings chapters, I show that alongside the numerous benefits of deploying unsupervised computational analytical techniques in humanities research, and particularly in terms of processing large, unstructured text-based datasets, there are equally a number of issues with this approach. Through an analysis of activities related to the use of vinyl records and the iPod, I challenge many commonly held assumptions about both activities and demonstrate the possibilities and issues with computationally highlighting patterns and correlations as a means of knowledge generation. This allows me to unfold an exploration of questions regarding the type and efficacy of knowledge that is produced through computational analytical techniques when they are applied to both commercial audience construction and to datasets describing cultural activities in humanities research.

Chapter 6 is based upon a reflexive approach to my primary study and demonstrates that my methods of data collection and analysis are able to provide useful routes to specific insight within complex datasets. The particular analysis here focuses on regular users of streaming services, but the broader argument made is that the same techniques would facilitate other focused examinations, regardless of the dataset. In this instance my method enables me to show that despite the efforts of streaming services to engender a sense of brand loyalty amongst consumers, respondents appear to reject aligning themselves with particular streaming brands and instead view streaming as primarily a new, additional mode of music reception that augments existing listening practices. Related to this I also show that despite their growing use, streaming services do not in and of themselves lead to expressions of greater/better levels of consumption in terms of the discovery and recommendation of new music. Finally I demonstrate that an analysis of respondents' stories shows very little evidence of concern over activities related

to issues of data collection, privacy or algorithmic curation. The presence, however, of quantitative data that demonstrates otherwise, gathered by the survey element of my data collection, suggests that such concerns may be worth exploring further.

This leads to chapter 7, which explores the issue of how the reductive and productive processes of data collection and computational analysis are able to deal with the complexity of everyday life, whilst simultaneously showing that use of technologies that facilitate precisely such analysis are inspiring new cultural practices related to engagement with music. I argue that it is these new cultural practices that can provide popular music scholars with a route towards new work in the field. In particular I show that respondents to Harkive are demonstrating intriguing degrees of engagement with their abstracted, digital selves produced by and through their use of digital interfaces. From this I argue that new means and models of understanding the activities and outcomes of music reception should be explored in order for people to be able to make sense of both this process, and the conditions by which their practices are increasingly monitored and guided. I explore how might we come to terms with this, both in terms of popular music studies and also in the management of our everyday lives, and argue that one way this could be achieved is through further development of the work I have begun.

In my concluding chapter I undertake a reflection on the successes and failures of this project, and provide recommendations for future activity. In many ways my thesis could be understood as an argument for the continued development of the work I have undertaken so far. The systems, processes and analysis I have conceived of are limited in size and scope, particularly when compared to the sociotechnical systems deployed in commercial and others operations. They are limited in scope also to systems used in other areas of contemporary scholarship, particularly more technically advanced work in areas of cultural analytics and digital humanities. Nevertheless I argue that my work opens up important possibilities for the field of popular music studies. We may consider, for instance, the types of new insights that could be arrived at if the techniques I have

developed were to be applied to historical texts and data sets (for example, lyrics, sales figures, reviews and criticism), or to the ongoing data collection activities I have planned for Harkive. Ultimately I show that my work is growing into a potentially valuable archive of data of contemporary cultural practices linked to music, and a repository of modular digital research tools that have useful applications within and outside of popular music studies.

***

When I initially embarked upon my research, I had envisaged I would be operating within the field of popular music studies. My belief was that I would be looking at the changing ways in which we experience music in light of digital technologies. My engagement with a broad range of literature, along with ongoing reflection brought about by events such as the one on the train to Birmingham, soon led me the realisation that what I was also trying to do was something else: I wanted to understand not just contemporary cultural practices associated with music, but also practically and critically engage with digital, Internet and data technologies, the cultural practices associated with them, and the role these now play in different social contexts.

This was a subtle but important shift in object and focus that helped reveal a significant gap in my skills. I had embarked on a path that required a degree of practical knowledge related to data, analysis and computational research techniques that I did not possess. Interestingly, and as this thesis also reveals, I discovered that as a humanities scholar I am far from alone in this regard. The increasing importance of data collection and analysis on many aspects of everyday life, along with the growing availability of data and tools related to it as potential research resources, has presented a compelling set of questions, opportunities and problems. My work is situated amidst those, and this thesis is an attempt to first of all engage with them, and to then provide a way forward.

A considerable amount of my research energy has been spent on practical questions. My proposed approach made it necessary to learn how to collect,

organise, analyse, and interpret data from the standpoint of the computational, rather than from the position of more conventional interpretive methods we usually associate with the humanities. This was not based on a belief that the computational should replace these methods, but rather that one could be significantly augmented by recourse to the other. Along with an approach based on a close engagement with literature and texts, my research has also involved, for instance, learning how to write computational code within R, a popular software language and suite of resources located in the burgeoning field of data science. This has also led to engage with academic work that takes similar approaches, but which is occurring outside of popular music studies and in the emerging disciplines of cultural analytics and digital humanities.

This approach has revealed an interesting intersection between critique and practice that can be observed in the laborious detail of my practical learning. Despite now being able to perform some reasonably complex computational analyses, however, I'm still very much a novice coder. This disconnection between ability and ambition has been – strange as it may seem – enormously helpful. A computer script can be understood as a logical argument in much the same way that a theory can, and when you build a computer script very, very, very slowly, as I have been forced to do by my technical limitations, the myriad of assumptions inherent in your 'argument' are constantly foregrounded. When we focus upon computational techniques, there is a fascinating tension between the empiricism of numbers, data and scientific process, and the reflexive realisation that the methods used in pursuit of that are ultimately creative acts. This is not a situation that is unique to the computational; it is a condition of research and knowledge creation more generally conceived. It is an epistemological conundrum.

Interestingly, this tension and conundrum is very much what also occurs when the cultural practices of people engaging with popular music meet with the rationalized, empirical systems of data collection and analysis. How could something as innately unique as a specific experience with a pop song ever be adequately explained away or understood by a data point? Conversely, to what

extent could a data point ever be seen as significantly implicated or influential in that highly individualised action/reaction? These are the type of additional questions I have wrestled with during the course of my research.

Ultimately, I will have raised more questions than answers, but I will show that along the way I have provided both an argument for and perhaps even a suggested path towards, a more practical engagement with data technologies than currently exists within the field. I will show that I have provided a means by which scholars may be prompted to think about the efficacy of data systems when they are applied to the popular music. I will challenge numerous assumptions around ideas related to what a data point can and does represent, and what the potential consequences for popular music cultures. Specifically, what happens when social action is reduced to data points, which are in turn subjected to statistical and other mathematical analyses that can often remove or else under-privilege what are seen as statistical anomalies and outliers? This ultimately begs the question of the extent to which we are now collectively involved (as consumers, practitioners, technologists, musicians, researchers) in the creation of a different kind of popular music culture.

This thesis is also an invitation to popular music scholars interested in similar questions to consider collaborating with me on them. It is also an invitation to scholars from outside of popular music studies to consider working collaboratively towards those ends. My ultimate hope and aim in undertaking this work has been to make the case for, and begin drawing together, a movement of sorts in popular music scholarship; one that could begin to engage both practically and critically with ideas and techniques related to the intersection of digital, data and Internet technologies, the everyday, and the cultural practices associated with music. How we get to that point, and the difficulties we face in doing so, drives the narrative of my thesis.

But in closing this introduction I should also add like to add the following: despite a focus on technology, this thesis is also about music. It's about the sheer, atavistic, punch-the-air joy it can bring us, and also the tears it can occasionally

reduce us to. Without a fascination for music, something I share with the thousands of people who have contributed to this research, there is no Harkive project, and there are no questions to answer. 'Without A Song', as the famous philosopher Frank Sinatra once said, 'we'll never know what makes the rain to fall, or what makes the grass grow so tall' (Rose & Eliscu, 1929)

The limitless wonder of songs, and the questions they can point us towards: that is the real story of this thesis.

CHAPTER 1

Where we are and how we got here

In this chapter I explore the role of digital, data and Internet technologies in popular music, and the structures and cultural practices associated with them. Through a historical examination of the development of the music industries, I focus on how audiences have been constructed and conceived over time, and how popular music scholars have addressed the various issues associated with those processes. I then show how the technologies, structures and practices associated with data collection and analysis have emerged as important new elements in the production, distribution and consumption of popular music.

1.1 – Taylor v Apple

On Sunday 28th June 2015 a post appeared on the American singer Taylor Swift's Tumblr page. Apparently written by the singer herself, the post was titled: 'To Apple, Love Taylor'. "I write this to explain", she opened, "why I'll be holding back my album, 1989, from the new streaming service, Apple Music."

Apple had announced their new service on 8th June 2015[1]. Due for launch on 30th June it would open with a three month trial period during which Apple's existing customer base of five hundred million people[2] would be able to use the service free of charge before billing began in September 2015. When details of the contracts between Apple music and rights holders (record labels, artists, and so on) began to circulate online, it emerged that rights holders would not be paid royalties for plays of their music during the trial period. It was this that had prompted Swift to withhold her music. "This is not about me", she continued:

[1] The original announcement is available on Apple's site: https://www.apple.com/uk/newsroom/2015/06/08Introducing-Apple-Music-All-The-Ways-You-Love-Music-All-in-One-Place-/

[2] 500m is the figure used by Apple to refer to what it calls its 'installed base' of existing users. See: http://uk.businessinsider.com/iphone-8-installed-base-2017-9

> This is about the new artist or band that has just released their first single and will not be paid for its success. This is about the young songwriter who just got his or her first cut and thought that the royalties from that would get them out of debt. This is about the producer who works tirelessly to innovate and create, just like the innovators and creators at Apple are pioneering in their field…but will not get paid for a quarter of a year's worth of plays on his or her songs[3].

She closed the 488-word post with the words: "We don't ask you for free iPhones. Please don't ask us to provide you with our music for no compensation".

Swift here was providing added weight to a growing discord over the issue of royalties during the Apple music trial phase, which was the subject of much discussion in media outlets and on social media. In a statement issued a week earlier, on 18th June, the UK-based Beggars Group, which represents independent record labels such as 4AD, Rough Trade and others, had said "We struggle to see why rights owners and artists should bear this aspect of Apple's customer acquisition costs.[4]"

What followed in the next 24 hours, however, generated further headlines for both Swift and Apple. 'We hear you @tayloreswift13 and indie artists. Love Apple', tweeted Apple CEO, Eddy Cue[5], as part of an announcement that Apple would now pay royalties during the trial period after all. 'Taylor Swift prompts Apple Music U-Turn over artist payment', ran the headline of an article in The Independent newspaper[6], in a variant of a story that ran in media outlets worldwide[7]. Compelling as this narrative was, however, some commentators were sceptical[8]. Was it likely that Apple would alter their position, costing potentially millions of dollars in the process, based on a blog post by a single artist? This line of reasoning led some to suspect that a deal was already done,

[3] Swift's original post has since been deleted from her Tumblr page but is reproduced in full here: http://www.stereogum.com/1810310/read-taylor-swifts-open-letter-to-apple-music/news/
[4] See https://www.theguardian.com/music/2015/jun/18/beggars-group-statement-apple-music-three-month-free-trial
[5] Cue's original tweet can be viewed here: https://twitter.com/cue/status/612824947342229504?lang=en
[6] See http://www.independent.co.uk/news/business/news/apple-music-will-pay-artists-during-three-month-trial-period-after-taylor-swift-writes-open-letter-10335488.html
[7] See this article in the New York Times: https://www.nytimes.com/2015/06/22/business/media/taylor-swift-criticizes-apples-terms-for-streaming-music-service.html?mcubz=3
[8] See: http://www.techtimes.com/articles/63421/20150625/was-taylor-swift-apple-feud-a-conspiracy-to-publicize-apple-music-streaming-service.htm

and that Swift's post had presented an ideal opportunity to announce it with considerable impact under the aegis of a compassionate U-turn. Indeed, a press release statement issued by WIN (Worldwide Independent Network), dated 24th June, appears to suggest that the issues over royalties had been resolved at least 24 hours before Swift took to Tumblr:

> I am pleased to say that Apple has made a decision to pay for all usage of Apple Music under the free trials on a per-play basis, as well as to modify a number of other terms that members had been communicating directly with Apple about. With these changes, we are happy to support and endorse the deal[9]

Whether or not Taylor Swift did indeed influence Apple's decision is something we may never know, but one thing we can observe from the incident is that it was not the first time in recent years that artists, rights holders, and the operators of online music services, had struggled over issues of ownership and control of the digital space. In 2012 it was another streaming service, Spotify, who were the subject of complaints over their royalty rates, which were publicly criticised by artists and labels, with Radiohead's Thom Yorke being particularly vocal[10]. Similarly, YouTube and parent company Google have in recent times also been embroiled in public conflict over the revenues derived from their service[11], and in April 2015 it was the turn of Soundcloud, who were served with an injunction by the UK collection society PRS over their royalty rates following a number of years during which extensive back and forth discussions had failed to produce a mutually agreed solution[12].

These debates, and the various positions that different groups take, are interesting to observe and help highlight some of the key issues of debate I am engaging with in my work. Each of the parties involved above have their own

[9] The full WIN statement is available here: http://winformusic.org/news/win-statement-on-apple-music-agreement-24th-june-2015/
[10] In October 2013, Yorke called Spotify 'the last fart of a dying corpse' - https://www.theguardian.com/technology/2013/oct/07/spotify-thom-yorke-dying-corpse
[11] Performing Rights Societies in numerous countries, and particularly GEMA in Germany, had been negotiating with YouTube since 2009 before reaching a settlement in 2016. See: http://musically.com/2016/11/01/youtube-ends-gema-standoff-with-music-licensing-deal/
[12] Soundcloud and PRS eventually reached a settlement in December 2015: https://www.prsformusic.com/press/2015/prs-for-music-and-soundcloud-reach-a-multi-territory-licensing-agreement-ending-legal-proceedings

vested interests to protect, and with digital delivery mechanisms being a relatively new mode of consumption (and thus revenue), it is understandable that these debates take place. Some figures for the commercial value of recorded music in the UK released on 3rd January 2017 by the British Phonographic Institute (BPI, 2017) provide further context here. One of the headline figures showed that 45 billion audio streams were 'served' in the UK in 2016, an increase of 68% on the previous year. The figures showed also that the annual number of UK-based streams had increased by 500% since 2013. Clearly, and certainly in terms of commercial revenues, the growth in music streaming represents a major change in the means by which people in the UK consume music, and one that is being replicated in other geographical sectors (IFPI, 2017)[13]

However, the report also revealed that in the same period the retail value of digital downloads had fallen by almost 30% on the previous year, while revenues from the sale of vinyl records were travelling in the opposite direction, growing by 53%. The commercial resurgence of vinyl records after several decades during which the format was thought by many to be moribund, coupled with the fact that download revenues are now falling so relatively soon after they had first appeared as a mainstream commercial proposition after the launch of Apple's iTunes service in 2003, adds complexity to our understanding of the overall picture. Although the new technologies of streaming now facilitate cultural and commercial activities engaged in by many people, the physical record still has a place, despite being based on a technology that is now over 100 years old. Downloads, meanwhile, and much like the record and the CD[14] before them, have been overtaken by a new technology following a period during which they were the dominant format.

The figures above, along with the Taylor v Apple story, reveal three key activities that have existed in various inter-related forms since the emergence of a popular

[13] The IFPI reported in 2017 that streaming revenue worldwide had increased by 60% on the previous year. Full report available here: http://www.ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2017

[14] The BPI figures revealed that 47m CDs were sold in the UK in 2016, with physical formats (CD & vinyl) representing 43% of total revenues.

music industry in the mid 19th century (Russell, 1997). Once we peer past the layers of fame, of powerful entities wrestling for commercial control, and the apparent centrality to the Taylor v Apple debate of recently emerged digital and Internet technologies, we can see three things: the commercial exploitation of products related to music; the manner in which audiences for those products are understood; and the many different ways in which audiences engage with those products. The old and new playback technologies mentioned in context of the BPI figures above are important to the ways these activities are closely inter-related, and we can understand these activities more broadly in terms of the processes of production, distribution and consumption. These in turn each contain technological, commercial, social and cultural aspects that together further complicate our understanding of them.

Although their reasons and methods have variously differed and occasionally converged, arriving at an understanding of how audiences engage with music has long been the goal of both popular music scholars and commercial entities. This has resulted in various configurations of audience construction over time, each helping to produce different consequences and forms of knowledge. The relatively recent introduction of digital, data and Internet technologies invites us to revisit questions that have emerged as popular music has developed over time. At the same time, however, they also raise important new issues.

1.2 – Old and new issues of debate

We may consider here, for instance, the Apple customers invited to trial the new streaming service and observe that 500 million is as conveniently round a number as it is an impressive one in terms of size. It highlights the manner in which audiences are constructions, and how these constructions have often been commercially driven through various processes, including market research techniques, promotion and marketing activities, and structures such as the sales charts. Yet, alongside this, there have also developed paradoxical rhetorics of individual and collective taste and meaning making, and the sense that both popular music texts and the way audiences interact with them can be read and

considered in a number of different ways. Critical theoretical, cultural studies, and sociological examinations of music in everyday life have all variously observed that the conditions of popular music's production, distribution and consumption contain interesting tensions, and this study is no different in that regard. I will show, however, that conditions within the related areas of production, distribution and consumption, and particularly as they relate to the manner in which both audiences engage with music, and how they are constructed and understood, have recently altered in important ways that amplify questions raised by popular music scholars since the 1940s. These relate to ideas of how audiences can be organised, and how the activities of those audiences are understood in matters relating to choice, agency, identity and meaning-making. The changes that invite us to revisit old debates and consider new issues are closely linked to the emergence of digital and Internet technologies, and specifically for my purposes, to associated practices and technologies surrounding the collection, analysis and use of data about audiences.

Because of their growing centrality it is sometimes easy to forget that the digital and Internet technologies that have transformed the business, cultural and technological environments of popular music over the last three decades are relatively recent developments. None of the digital music services mentioned in the opening section, for instance, existed before 2003, and as such it is perhaps no surprise that our understanding of them should still be in a process of development. One way we may begin to think about developing our understanding is to consider a consequence of the fact, illustrated by the BPI figures above, that many people now engage with music through online and digital interfaces. Because of this a growing amount of data is gathered from and about peoples' activities (the songs they play, the reviews they read, the artists they discuss on social media, or search for on Google) and is now available to those who are seeking to understand audiences. As Housley et al show, the use of online and digital interfaces generates data as an "accountable trace and functional pre-requisite for network and system integration" (2014:2). These technologies and the issues they raise are explored in more detail in chapter 2,

when I examine how they alter the ways in which audiences are constructed and understood. Digital, data and Internet technologies and the practices associated with them are also facilitating new forms of academic enquiry that can allow us to attempt to account for those changes. These new forms of enquiry are creating new debates about how research can and should be designed, performed and disseminated – something I will discuss in more detail in chapter 3.

It is the exploration of these inter-related issues, as they relate to both audiences for popular music and how we may study them, where I find the foundation for my own enquiry. In other words, digital, data and Internet technologies both enable and produce the necessity for my study. This is because the very means that make my study possible are also those which enable those operating within the companies working in the digital and online spaces of popular music to understand audiences in new ways, and – specifically – in much greater detail than has previously been possible. This raises questions about, for instance, the consequences of the reduction of individual and collective cultural practices associated with music to quantifiable, real-time metrics that in turn are becoming important factors in the way music is produced, distributed and consumed. These same processes are also an influential factor in many of our other everyday experiences (see: O'Neil, 2017), and now play a role in the way we understand the world (see: Lynch, 2016), our place in it (see: Cheney-Lippold, 2017), and – equally importantly for my purposes – how we conduct research and approach ideas related to the generation of knowledge (see: Kitchin, 2014).

This study will focus, then, precisely on those technologies as they relate to the construction of audiences for popular music, the practices that those audiences engage in, and how we may begin to understand what this could mean for the ways in which we can perform popular music studies research. My three findings chapters explore each of these areas in turn. Before we reach that point, however, the first step is to situate this particular study within the field of popular music studies. Because of my focus on digital, Internet and data technologies, I require a starting point, and later in this chapter I show how from the 1980s onwards a series of inter-related technological, commercial, social and cultural

developments – beginning with the emergence of the CD format – have brought us to the present point of study. As Frith (2002) observes, however, when studying popular music there is often more to be learned from continuities than change, and indeed we can understand much of the contemporary landscape – and the older debates engaged in by popular music studies scholars that new developments invite us to revisit – through studies that have focussed on periods prior to the 1980s.

1.3 – Looking back

Russell (1997) and Sanjek (1988), for instance, demonstrate that the early development of what was to become known as popular music in the UK and USA during the mid-late 19th century helped set up many of the structures and cultural practices we may recognise today, and in particular the establishment of popular music as simultaneously a cultural form and a commercial product around which business models underpinned by the exploitation of copyright coalesced[15]. Russell shows also how the then new technologies of the telegraph and the railway enabled music publishers and promoters to create commercial networks that brought songs to regional, national and sometimes international recognition. Several decades before the inventions of radio, records, television, or the Internet, there is evidence of songs being familiar to millions of people (Palmer, 1977) . The arrival of recorded music (Chanan, 1995; Laing, 1991; Milner, 2010) and radio technologies in the early decades of the 20th century augmented this process. Recording technology effectively froze a few moments of time and space – the performance of a song in a given location – and encased it in physical, productised form that could then be experienced at innumerable geographical and temporal removes, making it the primary means by which music was experienced for those not within the immediate geographical proximity of its creation or performance. Radio disseminated live and recorded music to large audiences in simultaneous, real-time (Hendy, 2013), enabling

[15] Russell describes how song writing, for instance, became a professional process controlled by a small number of music publishers, the decisions of whom helped establish and reinforce certain stylistic conventions in the songs made available to publics. Weber (1977) shows that music publishers in 1780 carried around 100 songs combined, a figure that had increased to 200,000 by 1830. Many of the stylistic conventions established by the commercial practices of publishers – the verse/chorus structure, a catchy, memorable melody, or subject matter relating to love (or lack of it) – remain with us to this day and can be observed, for instance, in the work of Taylor Swift

songs and performances from other locations and cultures to be brought to wider audiences in an instant. The technologies of television, transistor radios, FM Radio and cable television that emerged from the 1950s onwards provided further and incrementally more varied means for the dissemination of music (Wall, 2013:54-60), while the introduction of the cassette in the 1970s[16] enabled consumers to "enjoy music at any time, at any place, in any wanted programme or programme compilation, in any desired sound quality and almost at any wanted price" (Andriessen, 1999:12). We can thus see that the digital and online delivery mechanisms offered by download and streaming services, whilst new, are nevertheless part of a longer arc of structures, technologies and practices. Further, we can see the large catalogues of music available to consumers today, along with the speed and variety of means by which they can engage with them, are also the continuation of series of developments that predate digital and Internet technologies by many decades.

Similarly, we may consider one of the other key issues of debate I am exploring in a similar way, and observe that the ways in which audiences can be now understood at the level of the individual by the data collection practices facilitated by digital and Internet technologies, is also part of an equally long period of development. Ohmann (1996) and Lears (1995) allow us to see how the developments described by Russell and Sanjek above occurred alongside of the emergence of mass culture during the 19th century, and particularly in terms of the ways that audiences for cultural goods were initially conceived of in terms of social status, gender and purchasing power[17]. Later studies (see: Cohen, 2004) show how from the 1940s onwards[18], and aided by advances in market research techniques derived by academics from disciplines such as psychology and

[16] By the end of the 1970s the cassette would peak in its popularity following the introduction of the Sony Walkman, although it had been available as a consumer electronics product throughout the 1970s. There are clear parallels between the introduction of the Walkman and that of the iPod some 30 years later. Both were hardware devices that took established audiences for existing 'software' and added the element of portability. However, what made the cassette (and the MP3 some 30 years later) additionally attractive propositions to consumers was the heightened level of individual agency and potential for choice they introduced.

[17] Richard Ohmann (1996) offers a definition of this new, emerging 'mass culture' as publics being provided with 'voluntary experiences produced by a relatively small number of specialists, for millions…to share, in similar or identical form, either simultaneously or nearly so; with dependable frequency; mass culture shapes habitual audiences, around common needs or interests, and it is made for profit'.

[18] .Cohen (2004) describes life for consumers in the 1950s and beyond as a 'pursuit of prosperity' in what Thomas Hine (1986) defines as a spirit of the 'populuxe', where goods were 'fashionable and luxurious, yet affordable to many' for the first time, contributing to 'one of history's great shopping sprees'. Between 1945 and 1959, the number of companies registered in the United States grew by 45% to meet the demand of the newly affluent population.

psychoanalysis, audience were constructed according to smaller, sub-groups based categories such as age or musical genre, and had their perceived behaviours, wants and preferences catered for as the logic of mass communication was scaled down to consider homogeneity within a given segment, but within a growing heterogeneity of segments. Armed with more detailed statistics and computational techniques, the attention of marketers from the 1970s onwards turned towards ever smaller subcultural groups. The strategic targeting of market segments had become so successful by 1970 that Business Week had declared that concepts such as 'mass audience' and 'mass media' were now misnomers and, as Cohen argues, "what resulted was a new commercial culture that reified – at times even exaggerated – social difference in the pursuit of profits" (2004: 309). Here we can consider also Frith's (1992) observation, that the business of popular music by the time it had reached the 1970s was a highly efficient service industry, capable of exploiting new styles and markets at speed, that itself is part of a longer process of corporate concentration and rationalisation that dates back to the early part of the 20th century (see: Murphy, 2014; Southall, 2009). Thus, when we consider issues regarding the digital monitoring of individuals – and particularly the relative novelty of such a practice – was can see too that this forms part of a longer structural arc.

It is within the context of the various inter-related developments sketched above that some of the key debates of popular music studies have emerged. Adorno (1942) focussed on the "standardised", "pre-digested" form of popular song that implicitly linked popular music to the uncritical use of leisure time. Popular music, he argued, acted as "social cement" through generating the illusion of meaning, where the popular music consumer was either "rhythmically obedient", and thus primed for mechanised labour, or else "emotional", and drew false meaning from the pseudo-individualised material, "consuming music in order to be allowed to weep". This critique of popular music is one that, as later scholars have argued, denied any degree of agency to listeners (see: Bogdanovic and Longhurst, 2012:6) but Adorno's critique of the mass production and consumption processes of popular music is one that could be usefully applied at

any point along the timeline sketched above, or indeed beyond. Sterne (2006a), for instance, revisits Adorno's idea of pre-digestion in his analysis of MP3 technologies. But it is the recent emergence of automated recommendation services facilitated by digital, data, and Internet technologies – discussed in more detail in chapter 2 – that invites us to revisit and reconsider issues of agency and choice (or lack of it) within present day consumption practices.

Cohen argues that the consumer culture of the 1950s helped drive a fundamental economic, political and cultural shift that meant ideas of the citizen and the consumer were no longer diametrically opposed, as they had been in critiques of mass culture. Instead, they overlapped and thrived on the tension that existed between them. Hall's model of encoding/decoding (1980) is instructive when conceptualising this, and particularly so in light of later work that Hall influenced. Audiences, according to Hall, actively decoded messages contained in cultural texts differently depending on their social, cultural and economic backgrounds and experiences, and made sense and use of them through a process of reproduction. This represented a challenge to the top-down, passive conception of audiences offered by Adorno and was a theory developed alongside Hebdige's (1979) sub-cultural work exploring working class youth culture. According to Hebdige, by resisting dominant ideologies through popular cultural expression and consumption, groups were shown to be in possession of considerably more agency than the critical theoretical model allowed. Hebdige also argued that, ultimately, the practices of subcultural groups were co-opted or subsumed by powerful economic interests. The emergence of digital, data and Internet technologies enables us to reconsider this conceptualisation of an on-going negotiation between consumers and producers, and is explored in more detail in chapter 2 through the examination of automated recommendation services and mobile technologies, that capture detail not only at group and individual levels, but also within the spatial/temporal environments of the everyday. In later work, the role of personalised listening using mobile devices in the negotiation of individualised, everyday life has been examined by De Nora (2000) and Bull (2006, 2000), who conceptualise the Walkman and the iPod respectively as tools for the individual mediation of space, time and inter-

personal relationships in late 20th century environments. As I will discuss in chapter 2, however, these examinations took place before elements of connectivity and data collection were added to the functionality of mobile listening devices. Whereas Bull and De Nora both demonstrate that users were able to create unique, personalised soundtracks that were used to manage and sometimes disengage from everyday, public situations, the contemporary equivalents of the Walkman and the iPod now mean that – voluntarily or otherwise – their owners must remain connected. As such the use of individual mobile listening devices needs to be reconsidered also in light of the emergence of digital, data and Internet technologies.

1.4 –Towards the present

The arrival of the CD in the early 1980s, a joint innovation by Sony and Philips, was to play a major role in the manner in which popular music's present and past were packaged to those segmented audiences and individual listeners. Major labels turned their "dusty old vaults into cash dispensers" (Murphy, 2014: 314) by making their legacy catalogues available on a new format. The CD had the additional benefits for rights holders of being the product of more rationalised and thus cheaper production process[19] and of successfully being positioned at a higher retail price than legacy formats. The CD eventually became the dominant format in terms of global sales by 1992. Vinyl and cassettes were slowly squeezed from the racks of retail outlets[20] as the larger retail chains in the US, such as Walmart and Best Buy, began to dominate the retail space for the lucrative CD market. In 1999 Walmart claimed 20% of the US retail share for recorded music, despite only carrying 750 titles, and along with Best Buy held 65% of the market by 2001 (Murphy, 2014).

[19] The CD had an inauspicious start, however. As Milner (2010:214) explains, label executives were wary of the capital investment required of an unproven format and largely unconvinced by the audio quality of the new medium, shouting down a demo of the CD at an industry conference with chants of 'The Truth Is In The Groove!".≈.

[20] During this same period the larger parts of the recorded music industry further centralised through a process of buyouts, mergers and acquisitions and switched their focus to broader conceptions of rights exploitation. Stylistically innovative independent labels that emerged during the 1980s era were aligned to majors through distribution and other affiliate deals (Hendy, 2000; Hesmondhalgh, 1998). One of the biggest 'Indie' bands of the 1990s, Oasis, had their records released on the maverick Creation Records imprint, but were in fact ultimately contracted to Sony. Their main 'Britpop' rivals, Blur, similarly released records on the indie Food label, which was tied itself to Parlaphone/EMI. John Harris' 'The Last Party: Britpop, Blair and The Demise of English Rock' is a useful resource for this period.

Straw (1997a) describes a fundamental architectural rethinking of the retail space by major music labels and retailers at this time as they attempted to cater for both casual and committed music purchasers. The size of stores enabled a deep and slow-moving inventory that would serve the more "knowledgeable" buyer, whilst listening booths, point of sale recommendations, canonical reissue campaigns, and in-house magazines sought in combination to create "chivalrous" guidance to the casual purchaser who right holders and retailers had identified as a crucial "swing vote" in the success or otherwise of a hit record. As Straw observes, "these forms and technologies all presume a redefinition of record buying as an activity stimulated by discovery, previewing and experimentation, rather than repeated exposure to peer group reinforcement" (1997a:63). The issue however, as Straw points out, was that highly rationalised attempts to cater for 'individual' taste at big retail scale ultimately led to stylistically and socially homogenous buckets into which consumers were placed, closing off the potential for crossovers and convergence that the social reproduction of popular music afforded.

What is intriguing also about Straw's description and analysis of the distribution and consumption landscape of the mid-1990s is that, just as the previous section demonstrated how many of the structures and cultural practices we recognise from today were present at earlier points in time, the same is true of the 1990s regarding the particular technological and commercial practices I am exploring in this study. Two decades before the arrival of streaming, we can observe through Straw a joined-up, rationalised supply chain reliant on data collection and computational processing that attempted to cater for all tastes within segmented audiences. We can also see elements of the contemporary consumption experience, and in particular the apparently intimidating nature of having seemingly boundless choice. It is interesting to note here that the 'problem' of too much choice is something that predates the present period, but I will demonstrate in the next chapter how has become a driving force behind much of the 'chivalrous' curational rhetorics of data-driven music interfaces and automated recommendation services.

The arrival of digital formats and distribution in the late 1990s is often cast as a major disruptive force. Taking such a view necessarily provokes the temptation of conceptualising them as a break from any form of historical continuity. It is more productive, however, to place the developments of the 1990s and early 2000s within a broader framework, with digital distribution and consumption technologies better understood as a continuum. The widespread sharing and downloading practices that coalesced around services such as Napster (discussed in more detail below) initially presenting challenges to commercial practices and business models, particularly in relation to those surrounding physical media, but there was then the relatively quick re-emergence of a new variant of a familiar paradigm.

It is this that highlights another key issue of debate that I will address. The consumers who had briefly seemed beyond the control of rights holders now have much of their activity subject to increased levels of measurement and analysis that is also in line with the historical trajectory of consumer scrutiny and segmentation. Exploring in more detail how we reached that point is instructive in terms of framing this present study, however, particularly since it demonstrates how many important positions related to the processes of distribution and consumption came to be occupied by companies that had very little prior involvement with popular music. It is these companies, many of which only came into existence post-2000, or else can be understood as previously focussed on activities not inked to popular music, that are now providing the means and the impetus for the BPI figures related to digital music discussed above, and also the digital monitoring of individuals engaging with popular music.

Witt (2015), Milner (2010), and Sterne (2006c) have all shown that the emergence of MP3 technology during the 1990s was a not only a gradual process, but also one that was taking place with the knowledge of the music and entertainment industries. Witt (2015:91) describes how Karl-Heinz Brandenberg and his team of developers from the Fraunhofer Institute who developed the MP3 had continually failed to gain the support of the RIAA and

other commercial bodies in their attempts to have the MP3 adopted and supported as a legitimate format[21]. That frustration eventually led the developers to take the decision of releasing their technology under considerably less strict licensing conditions into the emerging online space of the World Wide Web. Baym (2010) has described how the 1990s saw the gradual, and then rapid uptake of Internet use in Western societies, and charts how the manner in which Internet technologies were 'domesticated' is in line with how the use of older technologies, such as the telegraph and the telephone, were socially negotiated over time. From its initial adoption in academic and military circles, the web became a mainstay in millions of UK and US homes by the mid-1990s, and the technology of the MP3 dovetailed with the growing subcultures around online activity that were driven not only by curiosity (and fears) of new technologies that Baym theorises, by also by the practical facts of ever and rapidly decreasing costs for hardware, storage and connectivity. At the same time, the larger parts of the music industries were – as they had been for some considerable time – more concerned with physical piracy, particularly as it related to the then lucrative CD market. Walter Yetnikoff (2005) and Maurice Oberstein, the chairmen of CBS Records in America and the UK respectively during the 1980s and 1990s, continually battled with their parent company, Sony, over the manufacture of hardware that enabled home copying[22]. Dell Glover, the employee of the Teure Huete CD pressing plant around which Witt's book is largely based, and who was allegedly responsible for the online leaking of over 1,800 major label releases in the late 1990s, was himself largely motivated not only by the excitement and buzz of the online space, but by the money he could make from selling bootlegged CDs around his neighbourhood. As Jeremy Silver (2013) demonstrates, there was scant willingness or ability within the wider music industries of the mid-to-late 1990s to consider the possibilities for online music distribution and consumption. A 1995 BBC Radio 1 documentary (BBC, 1995), about music and the internet, talked excitedly about chatrooms and the world-

[21] There were numerous reasons for this, as highlighted by Witt, and they relate to issues over potential financial agreements in terms of the licensing of the Fraunhofer technology, coupled with the fact that the RIAA, although nominally in charge of industry standards, was essentially powerless over the decisions of the right holders they represented, and that within those organisations there was widespread disdain for the audio quality of MP3s from studio engineers.

[22] In a newspaper feature on Witt's 2015 book, Oberstrein was quoted, talking of CDs and the possibility for physical copying, as saying, 'we're making a huge mistake. We're putting studio quality masters into the hands of people' see https://www.theguardian.com/music/2015/may/28/how-the-compact-disc-lost-its-shine

wide-web but nevertheless spent much of its 60-minute airtime discussing how CD-ROMs would revolutionise popular music in the years to come.

As we now know, however, it was not CD-ROMs but instead the widespread adoption of the Peer-to-Peer (P2P) technologies that enabled MP3 (and other digital files) to be shared online that came to represent a tipping point in digital music delivery. Like MP3 technology, Peer to Peer (P2P) delivery mechanisms had also experienced close shaves with commercial legitimacy [23] before eventually becoming freely available and spreading rapidly (Menn, 2003). Bhattacharjee et al (2007) identify what they call a 'watershed period' from mid-1998 to mid-2000, when the arrival of P2P and MP3 technologies mixed with a rise in popularity in other cultural goods (such as DVDs and computer games), along with a general downturn in the US economy, and combined to contribute to a brief period which came to be widely referred to as a 'digital revolution' in popular music (see: Kusek et al., 2005). Revenues for recorded music, having reached an all-time peak in 1999 as a result of CD sales, began to drop rapidly shortly thereafter and would continue to fall year-on-year for over a decade.

Academic and media texts from this period reveal that much was in a state of turmoil and flux. Many music and popular press articles predicted variants along the lines of the 'death of the music industry' amid talk of a "revolution" (Kusek et al., 2005; McLeod, 2005), and even the potential for a new 'middle class' of musicians emerging from a particular reading of Anderson's Long Tail theory (Anderson, 2007) [24] that would see traditional roles disappear through disintermediation (Jones, 2000). Other readings of the landscape, such as that provided by Fox (2004), were ultimately proved more accurate. Fox's assessment suggested a model very close to what would become known as streaming, some four years before the arrival of Spotify, whilst Burkhart & McCourt (2006)

[23] The most famous of these, Napster, was developed by a team led by 17-year-old Shawn Fanning. Attempts by early investors in Napster to pursue industry legitimacy were hampered by the a collective inability to address issues around copyright, and a chaotic internal dynamic that saw Fanning's uncle represent a significant stumbling block to progress. Napster's story, examined by Joseph Menn (2003), is as much a cautionary tale about digital-age business models as it is about the inability of the music industries to harness the potential for online music.

[24] Anderson's idea was essentially a manifesto for those who were considering online aggregation businesses. The success of companies such as eBay, Amazon, Uber, and other high volume/low margin operations, is more closely aligned with his observations, rather than the polemical readings that emerged alongside rhetorics of the imminent 'death of the music industry'.

correctly saw a fight for control coming in the online space as a new configuration of the political economy in the field began to emerge that placed technology companies in central positions in the chain of distribution.

Indeed, far from the chaos of revolution and disintermediation that characterised by the early years of the 21st century, we have instead seen a reasonably fast and steady process of reintermediation (Young and Collins, 2010), where aggregators, retailers and social networks provide important new functions, and the previous centrality of rights holders is effectively reduced to the role of content providers. There has been the establishment of new loci of exchange (Burkart, 2014), typified by streaming models that provide access to copyrights (Wikström, 2013) under certain conditions, rather than ownership, where, as Edwards at al (2015) show, 'good' consumers are those who pay their monthly fees, whilst 'bad' consumers are those who rip songs from YouTube in what is all together a rather reductive binary. Following decades of audience segmentation and fracture, and now armed with instant access to a huge number of old and new methods through which to engage with music, it is questionable whether anyone one truly fits this definition.

From 2003 onwards, when Apple opened its iTunes store, we have seen a change in the political economy of the music industries, with companies that are primarily, or had previously been, technology brands controlling a large share of the music market and wielding considerable influence as a consequence[25]. Meanwhile, Facebook, Twitter and Google control access to millions of online users in variants of 'walled gardens', and each are able to gather detailed, individual-level, real-time data about consumer/listener habits and their preferences that are in line with trend in historical trajectory drawn by this chapter. From the homogenous masses of the early years of mass culture, divided according to purchasing power and class, to being conceptualised as

[25] Some headline figures from recent years: In 2012, 70% of download revenue come via Apple's iTunes store (Forde, 2014), and 10% of all digital music sales originate in a Shazam transaction (Jenke, 2017). Spotify reported having 100m users in 2017 and in 2015 generated $2bn in revenue for rightsholders (Sweney, 2017). Even the smaller, niche players are significant: Bandcamp has generated $126m in artist income since 2008 (Hogan, 2015), and Beatport, a relatively minor retailer, was valued at $50M when it was sold in 2013 (Khawaja, 2016). To put these numbers into some sort of perspective, the entire value of recorded music in 2014 was $15Bn (IFPI, 2015), which is roughly 8% of Apple's reported $178Bn cash reserve at that time (Bowers, 2015).

male/female, rich/poor, young/old, black/white, audiences are now understood at individual levels according to variables that are the products of the unexpected correlations derived from individually tracked and agglomerated data points.

The emergence of access models from 2005[26] onwards, typified by companies such as Spotify, Rdio and Deezer, took the digitalisation of engagement with music to a further level of intangibility, as well as providing the potentiality for further monitoring and segmentation of audiences through real-time data collection and analysis. Rather than trading digital files (legally or otherwise) and storing them on personal devices such as smart phones, home computers and hard-drives, listeners instead gain instant access to extensive catalogues of material in return for either monthly fees or the placement of advertisements between songs. This is, in effect, a legalised and considerably more rationalised and efficient mechanism of online music delivery to that which the likes of Napster had been offering less than a decade previously, and one, furthermore, that was built on the very same technologies that Shawn Fanning, the Fraunhofer institute and others had harnessed and developed, yet failed to fully bring to commercial/legal fruition.

In terms of understanding some of the issues that surround these new technologies and the role they play in the experiences of listeners and audiences, the work of Sterne (2012, 2006a, 2006c) is a useful starting point. Although he emphasises the "pre-digested" nature of digitised music formats in a manner that recalls Adorno, in the sense that the technological processes involved in the creation of digital files "plays" the listener (as well as vice versa) by utilising psychoacoustic masking to remove inaudible frequencies and thus reduce file size, and while he highlights the sonic efficacy of such files in noisy environments for the distracted, busy listener, which also recalls critical theoretical perspectives, he nevertheless conceptualises them as inherently more "social" forms. The "promiscuous" nature of digital music technologies make them an

[26] Although initially conceived as an online dating platform and not as a music delivery mechanism, the emergence of YouTube in 2005 can be understood as perhaps the first truly mass-market music streaming service. There were numerous other services that emerged around this time also, and Last.FM's model was a hybrid of traditional radio and online 'free choice' that remains popular to this day. Others, such as SpiralFrogs's ad-supported access model, failed to gain traction in 2005/2006 but can be seen as forerunners to the now incumbent streaming platforms.

ideal fit for the segmented, fractured, internet-enabled audiences of today, but his crucial observation is the fact that "even in the most digital situations…most of the actual musical event still happens as sound in the non-digital parts of the social world" (Sterne, 2006a:106). This grounds such technologies within the lineage drawn by this chapter – they are in that sense not too different from sheet music, vinyl records, or the cassette. What is now different is how audiences are constructed and understood, and how audiences engage with music through digital, connected devices. What we do not know are the potential outcomes of those inter-related activities. Which brings us to the present day.

It is worth considering at this point, however – and as the BPI figures discussed earlier in this chapter demonstrate – that many of the practices that pre-date digital technologies are still in existence, including the production and consumption of sheet music, audio cassettes, vinyl records, and CDs. Consumers are now engaging with music in ways that combine the old and the new, in what Nowak (2014) describes as "fragmented and heterogeneous" consumption, or, as Magaudda (2011:16) observes, in a "circuit of practices", where the material objects of consumption are not just the records, CDs, radios and stereos, but also the multi-function devices carried by listeners, and the online services they use, all of which create new and diverse pathways (Finnegan, 2007) to meaning. As Nowak goes on to say, "the digital age is characterised by more diverse and uncertain modes of consumption where one format does not necessarily replace another, but complements it to further establish the differentiated omnipresence of music in everyday life"[27].

The contemporary landscape is, in other words, complex, and although it is seemingly dominated by recently emerged digital technologies (and companies), it is also simultaneously a reflection of the history of tensions that have developed between the related processes of production, distribution and consumption, and the technological, commercial, social and cultural practices each contains. The introduction into this mix, however, of new technologies that

[27] Nowak's article appeared online on the First Monday website, hence there is no page number to cite this quote. The full article is available here: http://www.firstmonday.dk/ojs/index.php/fm/rt/printerFriendly/5550/4129

are able to capture and analyse the "accountable trace" (Housely et al, 2014:2) of audience behaviour, is a complex, new phenomenon and raises another key issue of debate that requires closer attention.

1.5 – Data and Popular Music

Around six weeks before the BPI released the UK recorded music revenue figures discussed above, the UK trade publication AdWeek reported how the streaming music service Spotify had launched its largest ever OOH (Out of Home) advertising campaign to date. This campaign ran simultaneously in a number of countries where the service is available, including the UK. According to the report the creative focus of the campaign centered around putting Spotify's "vast trove of listener data to playful use…with executions that playfully highlight some of the more bizarre user habits it noticed throughout 2016". Adweek observed that the campaign was "a clever, engaging way to use data to humanise technology. And it works particularly well for music, since people do have such a passionate emotional connection to it[28]".

Besides being a useful asset in marketing messages to potential new users, the "vast trove of listener data" available to Spotify also serves another important commercial purpose in terms of advertising. As Seb Joseph notes in a news article in The Drum, the company has "been at pains to stress to media planners the value of using someone's streaming data to understand their mood"[29]. In other words, the data collected about users is valuable to the company not only in terms of developing experiential adjuncts to its service – something discussed in more detail in chapter 2 – but also as a means of generating additional revenue through advertising. Spotify's UK head of consumer marketing, Karen Staughton, quoted in the same article, said, "our community of users (and the data their behaviour creates) is one of our most amazing marketing assets". Staughton here was talking specifically about the deployment of user data in the

[28] http://www.adweek.com/adfreak/spotify-crunches-user-data-fun-ways-new-global-outdoor-ad-campaign-174826
[29] Joseph's article is available online here: http://www.thedrum.com/news/2016/10/29/spotify-why-it-s-launching-its-first-uk-brand-campaign-now

Spotify ad campaign described by AdWeek, but the same could equally be said about their use of data in the generation of advertising revenue that flows in the opposite direction, from media buyers towards Spotify. Individual and agglomerated user data gathered by companies such as Spotify provides them with two main benefits: it can help them improve and promote their services to users, but it can also be used to derive income from advertisers. This highlights how the use of digital, data and Internet technologies by publics facilitates the construction and organisation of audiences in a number of new ways.

Puschmann & Burgess' (2014) analysis of business and technological literature around metaphors of the type of "big data" practices exemplified by the activity of Spotify shows that discourses are often framed in terms of the harnessing and controlling what is seen as a "natural" resource to be "consumed". The idea that "data is cheap" and "plentiful" is a prevailing one, but in a manner that often elides its origins and eventual uses. This is a practice not unique to Spotify, of course, and is one instead replicated across numerous commercial and governmental sectors (see O'Neil, 2017) The term big data itself can be better understood in terms of Laney's (2001) defining characteristics of the "3Vs" of volume, variety and velocity. Large amounts of data (volume) – conceived in terms of being more than which can be processed by one person, or one machine – which is hugely diverse (variety) in that it can be drawn or mined from numerous different sources such as server logs, social media activity, financial transactions, and so on, can be collected, stored, categorised and analysed in close to 'real time' (velocity). This definition gives some context to the common framing of such activity in terms of metaphors related to abundant natural resources. It is also worth considering that the capabilities and affordances of such systems are in historical terms relatively new, certainly at the scale at which we are presently witnessing them, and so one can perhaps understand the often excitable tone redolent of exploration and possibility within business circles – data represents a 'frontier' to be crossed. Clearly, however, the idea of data being a 'freely available' resource is not one that should be accepted without question and relates to wider debates around privacy, ownership and surveillance that are currently being raised by academics in a number of fields

(see: (Boyd and Crawford, 2012a; Tufekci, 2015; Van Dijck, 2014; Zimmer, 2015, 2010). These will be discussed at length in the next chapter.

A further concepualisation of the manner in which systems of this kind function is provided by Rieder (2016:42-43), who offers four key characteristics: big data systems are characterized by large sets of logged behavior and cultural tastes (1), that are subject to new analytical techniques, including machine learning (2), within operating systems where capture, analysis and outputs are integrated (3) and which are well suited to a networked economy and the managerial spirit (4). Certainly, much of the activity currently being undertaken by Spotify can be understood in these terms, but such a definition could also usefully be applied to media companies more generally, and to practices that exist outside of the cultural industries. These same technologies are also now being used within the research process, generating debates that I will discuss in more detail in chapter 3.

Indeed, in terms of both commercial and academic work, data is used in the generation of new knowledge about audiences and so it is worth considering briefly at this point how definitions of the word data have changed over time and have been used by the scientific and computational imagination. The latin origin of 'that which is given prior to argument' is at odds with the aura of objective truth data now carries and (Gitelman 2013) has demonstrated the oxymoronic starting point of assuming that data should be taken at face value:

> [Data] are not facts, they are 'that which is given prior to argument' given in order to provide a rhetorical basis. Data can be good or bad, better or worse, incomplete and insufficient (2013:7)

As Boyd & Crawford (2012b) observe in Gitelman's work, data needs to be "imagined as data" in the first instance, which implies a subjective rather than objective process. Puschman and Burgess (2014) trace this movement towards perceived objectivity through its origins in an 18th century shift in focus towards scientific research and discovery, through to the emergence of computing in the 1940s, whereupon "data was increasingly used to refer to digital objects that can

be manipulated using a computer rather than generally accepted facts or the outcomes of experimentation or observation" (2014:4).

The prevailing idea that a unit of data can contain all that it is claimed to represent is thus highly problematic. In the case of music reception through digital interfaces we can understand this as the 'datafied' (and thus 'objective', 'truthful') representations of the complex cultural act of listening to songs, or placing two songs together in sequence within a playlist, or purchasing concert tickets, or discussing any of those activities with friends on social media platforms, all of which enable the reduction of social action to a finite number of data points. Gerlitz and Helmond (2013) demonstrate that activity in online platforms can have many different meanings, and that the platform-interoperability (Bodle, 2011) of services (where the same service can be accessed by mobile, desktop and tablet versions) allow for new and different reconstructions of data depending on the origins of its generation. This is because these spaces all contain what Gerlitz (2016:1) calls specific "grammars of action". Some of the data collected mirrors these, whilst other data does not. As such any reading of data must take the conditions, constraints and affordances of its method of collection into account. Added to this, Tkacz (2015)[30] has shown that once individual behaviour can be reduced to a data value, anomalous action can be corrected by interface design. In other words, the possibility for 'deviant' behaviour can be written out of a system. Thinking about popular music in those terms, we can understand Spotify and similar services, in part at least, as a response to the piracy that disrupted commercial activity in popular music around the turn of the century. However, the relationship between data collection/analysis and dynamic interface design also raises issues about how this new, highly rationalised variant of Thrift's knowing capitalism (2005) may potentially affect the music we hear. This is an idea I develop in the next chapter, through updating Straw's (1997a) work. In the meantime, we could consider Bernard Rieder's argument that these systems are not merely sets of

[30] Tkacz's work comes from The Poplitics of Big Data conference held at Kings College, London in May 2015 (https://www.kcl.ac.uk/artshums/depts/ddh/eventrecords/2015/bigdata-conference.aspx). Tcatz discusses the project that has was presenting on in the following web article: https://quarterly.demos.co.uk/article/issue-4/keeping-an-eye-on-the-dashboard/

technologies, but rather "performative forms of function", which, following Herrnstein-Smith (1979), can be seen as important and new dynamic elements in systems of value creation. Indeed, as Chapter 2 will also demonstrate, recent work from Webster et al (2016) and Prey (2015) has attempted to conceive of digital interfaces for popular music in precisely these terms.

At the risk of being reductive, it is nevertheless a useful starting point in terms of unpacking the large and somewhat nebulous subject of big data and related practices to consider the formulation of useful, distinct questions. What is a unit of data? How is it constructed? Does the nature of it, and its relationship to other units, change depending on the manner in which it is processed? How are algorithms and other processes of analysis constructed, and to what ends? What form does the knowledge produced through a process of analysis take, and does this change the nature of that knowledge? Who has access to it? How does algorithmically generated knowledge inform interface design, and—coming full circle—would this effect the type of data such interfaces are able to collect? These are the questions that will inform the design of my methodology, which will be detailed in Chapter 4.

1.6 – Summary

The debate entered into by Taylor Swift and Apple that was discussed at the beginning of this chapter can be better understood less in terms of new or disruptive technologies, and more accurately in terms of a wider historical framework of how technological, commercial, social and cultural practices have both changed and remained the same over time in relation to the inter-related processes of production, distribution and consumption of popular music. The period often characterised as a disruptive 'digital revolution' running from the mid-1990s until the early 2000s, is often seen as a tipping point that explains how we have arrived at the present day landscape. It was, however, the period that immediately followed the disruption that should be of far greater interest. The contemporary conditions can be understood as individualised consumption patterns meeting with the apparently boundless choice of material through on-

demand, digital and connected access methods, that in turn facilitate more and more detailed measurement, analysis and segmentation of audiences than has previously been possible. Taken together, these factors introduce a new set of issues that relate specifically to the consequences of the stability achieved post-2000. In the next chapter I will look specifically at what those consequences are for people engaging with popular music, before turning my attention, in chapter 3, to the issue of how we scholars may address these new conditions.

CHAPTER 2

Popular music and digital, Internet and data technologies

This chapter explores the issue of how both 'new' and 'old' activities related to engaging with music cross over and are subject to varying degrees of digital monitoring. In doing so I engage with debates about the role of digital, Internet and data technologies in the cultural practices associated with music reception. Through a close examination of algorithmic recommendation systems, digital interfaces, data collection, and the use of mobile technologies, I demonstrate their potential collective benefits and consequences in terms of listener agency, identity and choice. I argue that the growing influence of data-derived knowledge in many areas of our everyday life – and certainly in terms of how we engage with music – suggests that work seeking to examine the contemporary cultural, technological and commercial conditions requires us to engage both practically and critically with those processes.

2.1 - The ultimate personalised playlist

We can begin to examine these issues by exploring an announcement made a little over three weeks after Taylor Swift posted her Tumblr critique of the financial details of Apple's new music service, which opened the previous chapter. In this example it was instead a different music streaming service, Spotify, who were launching something new. Described in the headline of their announcement on 20th July 2015 as "your ultimate personalised playlist"[1], Spotify Discover Weekly promised to deliver a 'tailored' 30-song playlist each Monday morning to every one of Spotify's reported 75 million users[2]. Both the songs contained within each playlist, along with the manner in which the songs

[1] https://press.spotify.com/uk/2015/07/20/introducing-discover-weekly-your-ultimate-personalised-playlist/.
[2] As of June 2016 Spotify's user base had grown to 100m, according to company figures: https://press.spotify.com/co/about/

would be ordered, were to be derived from data and analysis of each users' own listening habits, and those of others, as the announcement explained:

> For the first time ever, we're combining your personal taste in music with what similar fans are enjoying right now. This means every song in Discover Weekly is based both on your own listening as well as what others are playlisting and listening to around the songs you love – making your playlist completely unique and full of deep cuts and new discoveries. It's like having your best friend make you a personalised mixtape every single week. (Spotify, 2015)

The growth of streaming as represents an important facet of modern day listening experiences for many people, and for global music industries revenues. After over a decade in decline following the emergence of P2P technologies in the late 1990s, revenues for recorded music rose again for the first time in 2013, and largely as a result of digital music delivery services, particularly Apple's iTunes download service. According to IFPI (2014) figures, revenues from digital delivery mechanisms exceeded those from sales from physical formats (CD, etc.) for the first time in 2014. In the following year figures from the RIAA (2015) revealed that revenues from streaming in the US had then exceeded those from downloads for the first time, contributing 34.3% of total revenue following growth of 29% on the previous year. Streaming now occupies a major share of corporate revenue in terms of recorded music, and the introduction of Spotify Discover Weekly highlights many of the key changes that have taken place in music consumption over the last three decades (Nowak, 2014), including issues of copyright and corporate control, the centrality of computer technology companies and the role of data-driven business models, and the ways in which people now listen to and engage with music.

To present and explore some of the issues related to the study of music consumption at this present time Spotify, as a relatively mature entity in the streaming market, represents an ideal starting point, and for two reasons. Firstly, in 2007 Spotify was amongst the first to commercially launch a service based on

a model providing metered access to catalogues of recorded music through digital delivery mechanisms. Although several competitors have emerged since (most notably Apple, who launched their AppleMusic service in 2015) each of these streaming services operates largely upon the same principle as Spotify. At present Spotify is also the largest streaming service, available in 60 different domestic markets worldwide, and reports having over 100 million active users, 50m of whom pay for a monthly subscription that enables offline, ad-free listening.

It is Discover Weekly in particular that demonstrates the second reason. Although the manner in which Spotify collects and processes data about its users and their activity in order to create Discover Weekly is unique to its own operational methods, these methods are indicative of what Hartmann et al (2014) have described as data-driven business models (DDBMs), which they define as models that rely on data as a key resource. The authors' subsequent examination of 100 new businesses revealed that 73% used external data sources (and in most cases 'freely' available social media data), and 76% had analytics as a core activity. The emergence of what Chen et al (2011) describe as two new types of business in "data as a service" and "analytics as a service" that each seek to benefit from data as a resource were very much in evidence in Hartman et al's findings. The collection of what is often framed by businesses as 'freely' available data (Puschmann and Burgess, 2014), its organisation, processing and deployment, are key components in the leveraging of a competitive advantage for Spotify, as product development manager Matthew Ogle has stated:

> I think that the music catalog [sic] is more or less commoditized at this point…if you can do the deals and set the service up, you can stream the same music that we have or that Apple has to millions of people. But something we're really keen on is establishing a relationship with our users and evolving that over time. It's the

> same way you would get to know the staff at your local record store.[3]

Spotify's competitors in the music streaming space each engage in similar data-derived activities and although the exact details vary from company to company, each is a means for them to generate an understanding of listener activity and taste through forms of digital monitoring. Alongside digital monitoring of listening activity taking place within the interfaces of streaming companies, other activities related to what Negus (1997:9) calls the reception of music, such as concert ticket sales, Google searches, the purchases of physical products, and social media discussions, now often take place within digital interfaces that make data collection and analysis possible (Amatriain, 2013; Vanderbilt, 2016). This wide-reaching data grab enables different and disparate data from possibly unrelated sources to be reciprocally contextualised in new ways, facilitating the production of a form of knowledge that is deployed via services' interfaces and that in turn can directly influence user experiences.

2.2 – Automated Recommendation

Along with on-demand access to large catalogues of music, a key feature of Spotify is that its users can create personal playlists from the large catalogues of songs the service carries. Matthew Ogle has described the playlist as the "atomic unit" and the "common currency" of the service (Dredge, 2016), and since 2007 the Spotify user base has generated over 2Bn unique playlists. Each of these playlists generates digital traces of who created it, the manner in which the songs within it are organised, how popular it has been with other users, and other metrics. This data has been recognised by Ogle (Dredge, ibid) and his team as a valuable proprietary asset and forms a key component of the technical process behind the weekly, automated creation of Discover Weekly playlists.

[3] Ogle was quoted in John-Paul Titlow's article for FastCompany: https://www.fastcompany.com/3049231/inside-spotifys-plan-to-take-on-apple-music

In addition to the playlist data that provides Spotify with an at-scale insight into how people navigate, organise and engage with the catalogue of music it carries, the company also gathers data on how people listen within their service more generally. Their interface captures at an individual level which songs are played, skipped, shared, added to playlists, via which devices, and at what time of day. Further to this, and following Spotify's acquisition of the self-defined music intelligence operation Echonest in 2013, the service also holds data on each song within its catalogue. Some of this information, often referred to as metadata, contains information related to both the value chain (i.e. publishers, rights holders, etc.) and information pertaining to the genre and sub-genres, song length, year of release, and so on. This information is augmented by Echonest who provide additional information about the tempo, musical key and other structural elements through computational analysis of recordings, and also through web-crawling technologies that use natural language processing and machine-learning to analyse online conversations occurring outside of the Spotify interface in relation to particular songs, artists or genres, on blogs, social media networks, and through articles published by media outlets. This second element of Echonest's activities highlight the manner in which cultural practices related to music that take place outside of Spotify's interface can be directly linked to activity that takes place within it. These data generation and processing activities enable both the Spotify catalogue and their user base to be organised and segmented according to a large number of variables. The deep level of metadata around each particular song, coupled with the collection of data related to user activity, facilitates the creation of detailed, individualised "taste profiles" of users and of abstracted versions of artists, genres and labels (Pasnick, 2015).

User taste profiles are then used in conjunction with collaborative filtering to help identify songs from the available catalogue that users have not yet listened to, but to which Spotify's internal processes indicate they may have a potential affinity for based on statistical similarities with songs from previous listening. It is these songs that are considered for inclusion in Discover Weekly playlists. The placement and ordering of the songs identified is determined by analysis of the 2Bn existing user-generated playlists, which helps Spotify determine what Ogle

refers to as the "flow" of the new playlists. The order in which songs in 2bn playlists have been placed by users – where a song from one genre may have been situated next to another from a completely different one in a manner that is perhaps only understandable to the creator of the original playlist – are converted to data points which help provide a degree of abstracted contextualisation for machine-derived selections. Through an iterative process, Ogle and his developers settled on an amount of 30 songs, or roughly 2 hours' worth of music for the Discover Weekly playlists, which they felt was "more human, approachable", and thus more like the advertised "personalised mixtape" than a machine-derived selection. Spotify's engineers run this process weekly and deliver a new playlist to each user every Monday morning. The rollout of the playlist was seen as a considerable success for the company, garnering favourable coverage in news reports and in online forums describing the efficacy of the playlists in terms of their recommendations as being "magic" (e.g. Pasnick, ibid). Furthermore, and according to Spotify figures[4], users streamed a combined 1bn songs from these algorithmically generated playlists in the first 10 weeks of release, a figure that had almost doubled before the end of 2015. The success of the rollout appears to have provided Spotify with a degree of validation for the use of data-derived, machine-learning approaches in the construction of a listening experience that aid customer retention. Ogle, speaking in March 2016, confirmed that Spotify intends to develop further features of a similar nature in the near future that create "appointments to listen" via the service[5].

2.3 – Choice, agency, identity

By capturing, analysing and reflecting back individual and collective tastes and activities in the form of recommendations, Spotify are engaged in a practice that is not entirely new in terms of commerce (see: Cohen, 2004), the media

[4] Spotify provided these figures in a press release, available in full here: https://press.spotify.com/ca-fr/2015/10/08/discover-weekly-reaches-one-billion-tracks-streamed-in-10-weeks/
[5] Ogle here was quoted in a MusicAlly interview, full text here: http://musically.com/2016/03/21/matt-ogle-discover-weekly-spotify/

industries (see: Lears, 1995) or indeed music specifically (see: McCourt and Rothenbuhler, 2004). However, the scale and degree of fine detail involved with the processes of creating machine-derived recommendation systems through digital interfaces, along with the relative novelty yet growing centrality of streaming as a mode of music reception, suggests that long-standing debates around individual choice and agency (Adorno and Simpson, 1942) and the role of the cultural intermediaries in terms of cultural goods (Bourdieu, 1984) need to be revisited in light of the recent technological, commercial and cultural developments exemplified by Spotify Discover Weekly.

Webster et al (2016) argue that a key function of automated recommendation in digital music interfaces is the leveraging of a competitive advantage in a crowded and undifferentiated marketplace. As Ogle suggested above, when all services offer the same (or largely the same) catalogues of music, at similar price points and in similar ways, one of the only competitive spaces that remains is the quality of listening experience delivered. This view is also in evidence in commercial reports (e.g. IFPI 2015, MMF 2016) that underpin research in the fields of computer and data science that seeks to improve the performance and efficiency of experiential systems. As Vanderbilt (2016) shows, in the construction and iterative rationalisation of automated recommender systems, implicit feedback – which can be understood as data gathered about which songs are played, skipped, shared, or added to playlists – is often viewed as a more useful 'raw' material than the explicit feedback volunteered by users in the form of star ratings, purchases or reviews. For Vanderbilt, this is where "abstract, 'unaccountable' notions of taste run into the empirical order of the Internet"[6]. It is also where digital monitoring and related systems play a role in relationships between lived experience and the consumption of cultural goods, thus raising issues of choice and agency.

[6] In promoting his book, "You may also like: Taste in an age of endless choice", citied in full elsewhere in this thesis, Vanderbilt also wrote numerous articles that outlined some of the key ideas. This particular quote comes from an online article here wrote in 2016 for the Star.com website. Full text here: https://www.thestar.com/news/insight/2016/06/11/how-netflix-knows-what-you-want-to-watch.html

Vanderbilt's examination of the film and TV streaming service, Netflix, demonstrates that a shift in focus towards implicit feedback data enabled them to avoid the inferential pitfalls of what they referred to as "aspirational rating", in other words the Veblensque reporting of the consumption of 'superior' cultural goods. Netflix found an evidential gap between what people said they watched (explicit feedback), and the films they actually watched (implicit feedback). Real-time tracking of activity within interfaces (where a movie can be abandoned, or a song skipped, or a product left in the virtual shopping basket), appears to offer much more compelling evidence of actual (and thus future) behaviour than a purchase receipt or a rating. Crucially, however, in automated recommender systems the data points of individuals (i.e. the skipping of a song by one user) are in isolation largely useless to broader aims and functionalities. As Vanderbilt observes, "the aggregate level is where, through sheer numbers, the noise can be filtered, the outliers marginalized, and statistical consensus achieved". (2016:5). This invites us to think about what the potential consequences on cultural forms and practices may be given the investment many service providers have placed in implicit feedback and algorithmic recommendation systems. It is here also where dynamic interfaces for cultural goods become linked to issues of debate that are concerned with the potential impacts of data-derived systems.

Tania Bucher's concept of "the algorithmic imaginary" (2016) is as useful way of thinking through this. It allows us to understand both how data-processing impacts upon experience, but also how experience impacts upon the design, function and use of algorithms. The algorithmic imaginary can be observed in action through a consideration of automated music recommendation services in particular: data is gathered on listener activity from which abstracted inferences of taste are derived; leading to recommendations that can positively or negatively influence choice; which in turn creates data about listener tastes; and the process repeats. Interestingly the decision of a user to ignore a recommendation is equally important here because it too creates data that is used to tweak recommendation algorithms. It is important also because although ultimately listeners can choose whether or not to follow

recommendations, they cannot chose whether or not their activities are recorded and subsequently used in the creation of recommendations. Our relationship with such systems is thus not entirely top-down, but rather one of co-production that is based on an unequal relationship. The conditions under which the user performs cultural work (Banks, 2007) are not entirely known and are inescapable as long as the user continues to use Spotify, ultimately helping produce consequences in the form of recommendations and dynamic changes to the user interfaces that foreground or not particular content. It is in around these points where debates of the potential outcomes of relationships between cultural practices and digital monitoring find their foundation.

Tufekci (2015), for instance, highlights the negative reactions to Facebook's emotional contagion study (Kramer et al., 2014), which measured the emotional effects on users of different types of content, and argues that questions around the potential harms/benefits of the algorithmic production of experience have moved "beyond hypotheticals" now that "algorithms act as de facto gatekeepers of consequence" (2015: 206). In terms of popular music this is not limited to the delivery of recorded music via digital interfaces. Bucher's algorithmic imaginary is also at play in the foregrounding of media content and advertising through social media and news media platforms, and increasingly in the promotional and A&R activities of record companies (Thompson, 2014). These are all areas where algorithms, fuelled by consumer activity data and cultural content metadata, are deployed as subjective decision makers. Tufekci defines algorithms as "computational processes that are used to make decisions of such complexity that inputs and outputs are neither transparent nor obvious to the casual human observer" (2015:205), and through which we can understand much of the work currently being undertaken by companies offering cultural goods through digital interfaces, and of Spotify's Discover Weekly in particular.

Drawing on Latour, Tufecki conceptualises algorithms as "actants", but then demonstrates the imprecision of such a conceptualisation because real-world analogies are hard to alight upon due to the complexity, fluidity and opaque nature of algorithms and systems they form part of, the latter point supported

by Boyd and Crawford (2012a) and Annany (2015). This operational opacity introduces "new obstacles in the quest for accountability and transparency in consequential gatekeeping" in a manner that "reverses or significantly modifies…traditional gatekeeping with regard to visibility, information asymmetry, and the ability of the public to perceive the results of editorial work" (Tufecki, 2015:209). Alongside this difficulty, Tufekci shows that the affordances of "latent trait inference" – where computation and statistical analysis can accurately reveal undisclosed information about users – introduces an even more complicated set of questions about privacy and civil rights. Latent trait inference allows marketers and other interested parties to go "beyond lumping people into rough categories and [to use] individualised data...to create a deep profile of a person with increasing levels of accuracy, improving every month' in a manner that is 'completely unaccounted for in law [or] policy" (2015:211)[7]. The case of Spotify, its users, and the mechanisms and rationales behind automated recommendation in the form of Discover Weekly, offer a way of relating these wider concerns and debates to popular music studies.

The irony amidst all this is that the systems and practices causing the concern are facilitated partly by users' engagement with digital interfaces. Spotify would not be able to offer Discover Weekly in its present form if people did not use their system to create playlists. Likewise, Facebook would have no contagion study without user feeds. Van Dijck (2014:1) argues that user data has become "a regular currency for citizens to pay for their communication services and security – a trade-off that has nestled into the comfort zone of most people." This uneasy covenant has been described by Barnes (2006) as a "privacy paradox", where people are uneasy about the information collected about them, which they know will be packaged and monetised, but nevertheless accept this as a condition of using certain services. As such, listeners, users, and publics have their roles, agencies and choices and are by degrees similarly implicated in the concerns raised by Bucher, van Dijck, and Tufecki. Although these scholars do not focus on popular music services, there are clear parallels with the activities and concerns they describe. The relationship between systems and publics is

[7] For more information on recent work around the ethical questions of digital monitoring, see Zimmer (2015, 2010)

further complicated by the fact that processes of this kind are often framed in rhetorics of personalisation, or in terms of being otherwise beneficial to users. We can consider here the language used by both Ogle at Spotify, and Amartriain at Netflix, that assume a spirit of willing, collaborative production between services and users. Amartriain (2013) provides an example of such in a paper outlining the technical aspects of the algorithmic recommendation system used at Netflix:

> We want members to be aware of how we are adapting to their tastes. This not only promotes trust in the system, but encourages members to give feedback that will result in better recommendations. (2013:40)

Amartriain goes on to claim that implicit feedback can provide crucial nuance "without intrusion". This qualification necessarily raises invites debate issues about the nature of such intrusion, which must surely be relative to the person being intruded upon and not a matter for the 'intruder' to define. This returns us to Barnes, to ethical ideas of informed consent (see: Zimmer, 2015, 2010) , and also to considerations of Banks' idea of cultural work (2007). Amartriain's position is indicative of many of the assumptions that underpin the rhetoric around the deployment and iterative optimisation of data systems and digital monitoring. As we saw in the previous chapter, Puschmann & Burgess (2014) show that prevailing discourses in business and technological literature frame such activities in terms of the harnessing and control of what is seen as a 'natural' resource to be 'consumed'. The characterisation of data as "the new oil" by the World Economic Foundation (WEF) in 2011[8] is a clear recognition of both the perceived abundance and economic potential of 'freely' available data that forms the basis of Hartmann et al's DDBMs. As Ogle and Amartriain's comments demonstrate, these metaphorical characterisations have been widely adopted and repeated in business circles in the manner suggested by Lakoff and Johnson (2008), all of which has the effect of domesticating (Baym, 2010) what Mayer-Mayer-Schönberger and Cukier call the datafication of social, cultural and economic everyday life (Mayer-Schönberger and Cukier, 2013).

[8] https://www.weforum.org/agenda/2015/08/is-data-the-new-currency/

The idea that 'data is cheap' and 'plentiful' is coupled with the advertised beneficial outcomes to users in a manner that often elides questions of origin or consequence. This can be observed in the comments of Matthew Ogle regarding the 2Bn user-generated playlists being the "common currency" of the Spotify system, which the company is now consuming in the furtherance of a business advantage, while the cultural work or informed consent of the users who initially created the playlists is not taken into account. We can recall that this language is also echoed in the initial promotional material accompanying Spotify Discover Weekly, where "every song..is based both on your own listening as well as what others are playlisting and listening to" (Spotify, ibid), which appears to ignore entirely questions of whether 'others' wish to have their listening repackaged and ultimately monetised in such a way. Effectively this monetisation is achieved when the implicit data generated as a byproduct of the cultural practices of users is computationally processed in order to generate positivistic insights that help companies retain users in an undifferentiated marketplace. These insights, as Tcatz (2015)[9] shows, are then deployed via the same interfaces that are used to generate data: Spotify Discover Weekly recommends songs to listeners, which in turn generates data that leads to more recommendations and listening activity. It thus exemplifies the way in which a number of new and recently emerging technologies, practices and outcomes are becoming commonplace, everyday and mundane for both users and companies alike, while little is known so far about what the short or long-term consequences may be of their increasing centrality.

2.4 – Understanding recommender systems

When Webster et al (2016) argue that in cultural terms systems of this kind exert influence over listeners, they do so in order to explore questions about the

[9] This is taken from Tcatz's paper given a conference I attended, The Politics of Big Data, in May 2015 - http://www.politicsofbigdata.net.

extent to which recommendation systems promote diversity or homogeneity in taste. A key line of their enquiry is how such systems negotiate traditional social dimensions of taste, such as age or gender, and whether these dimensions are replicated or transcended. They examine these questions through the lens of Pierre Bourdieu and in particular his work around cultural intermediaries (Bourdieu, 1984), but argue that the perspective has shifted from 'occupational categories of the intermediary to the networked practices of intermediaries'(2016:2), and so employ Latourian actor-network theory (Latour, 2005) and its relational ontology of human and non-human actors to help account for this. Through such a conceptualisation the elements of software, hardware and code involved with automated recommendation are considered on an equal performative footing with the feel for the game of human actors, such as the coders, engineers, and managers like Ogle and Amartriain. Through this Webster et al argue that the cultural capital of a given system is thus a product of both human and non-human actions and suggest that a broader understanding of what constitutes the technological elements of the processes involved is required. It is interesting, then, that Webster et al only briefly include the cultural practices (and thus work) of users as part of this system and its processes. They do this specifically in terms of the explicit and implicit ratings generated through consumption practices that were discussed above, before ultimately shifting their focus to how the data is socio-technically processed within the recommender system. In other words, despite demonstrating that users play a part in what they call the "assemblage" that constitutes the cultural intermediary role, the user contribution is in the end neglected in favour of a focus on what happens within the internal mechanisms of the recommender system. As I will demonstrate in subsequent chapters, although my work shares Webster et al's interest in the socio-technical aspects and functions of computational systems, it begins instead at the precise point where they step away from the cultural activities of users. Their work is necessary, however, given the aforementioned operational opacity (Boyd and Crawford, 2012b) of such systems, and they demonstrate usefully, via their update of Bourdieu, how systems such as Spotify Discover Weekly now play an important cultural role.

As a useful next step, Prey (2015) offers a different analysis of music streaming services that can further help to conceptualise the ways they are used, the capture of implicit feedback, and the potential cultural consequences of the deployment of knowledge derived from this via interfaces. Building on Lefebvre's trialectics of space model (Lefebvre, 1991), Prey argues that the production of music streaming services as experiential interfaces are an ongoing dialogue between ideas of form, concept and practice in a manner that links with both Bucher's idea of the algorithmic imaginary and to Webster et al's internally-focused, socio-technical conception. For Prey, how we perceive music streaming spaces, in terms of their interfaces and the experiences we have, is influenced by how they are conceived of by designers, engineers, and the business needs of the services concerned. However, he notes that this process also works in the opposite direction. The actions (or non-actions) of users inform the on-going, iterative (re)conception of interfaces because both human and non-human elements of what Webster et al call the "assemblage" react to the information gathered as a consequence of both user and developer actions. For Prey, the site of this ongoing negotiation is lived space, which renders streaming interfaces as never fixed in the production of what Lefebrve called the abstract space of capitalism, and this is because they are constantly informed by the actions taking place in social space. As Prey puts it, "the qualitative..is [thus] never completely absorbed by the quantitative" (2015:17). The process discussed earlier, whereby data leads to recommendation, which can guide listening activity, and thus generate more data, is an updated, highly rationalised form of how the music industries has commonly operated over time (we can think here in terms of sales charts, or radio airplay, for instance). This manner in which this process plays out, however, becomes much more tightly conceived of with the arrival of digital interfaces. But it is nevertheless important to remember that this process remains – both conceptually and in reality – a negotiation. People can, after all, choose whether or not to use Spotify, and to reject or accept recommendations when doing so.

An interesting question Prey raises through his conceptualisation of constant negotiation, both from the point of view of streaming services and the listeners who use them, is how ideas of identity begin to shift. This, in turn, has implications for issues of agency and choice. For Prey, algorithmic recommendation systems (for music) and dynamic interfaces (more broadly), facilitated by implicit and explicit data collection about users and their activity, mean that users are effectively conceived of as more than one user, all of the time. To this I would add that we can observe the same thing happening to ideas of musical genres, which are also constantly updated by systems exemplified by Echonest that collect and analyse data in order to categorise and organise music. In both scenarios, existing labels of identity, such as gender, race and age, (or rock, jazz and funk) disappear in what Cheney-Lippold has called the "cybernetic relationship to identification" (2011:168). Through this, users and genres are constantly (re)constructed in 'real-time', all of the time. Cheney-Lippold's more recent idea of "measurable types" (Cheney-Lippold 2017) further expands upon this. Under such systems the conception of a type (for example, a user's gender, or a musical genre) are constantly shifting according to newly available data, which has the effect of decoupling those identities from lived experience. Prey suggests that under the gaze and whilst operating within such systems we are therefore not individuals, but rather the result of a process of data-derived individuation. The question then changes: it is not whether the recommendation engine gets it 'right', but instead to what extent the user, or the genre, is a construct of the process. Where Williams once argued that "there are in fact no masses, but only ways of seeing people as masses" (1985:58), Prey suggests that – within the interface and the socio-technical assemblage behind it – there are in fact no individuals, only ways of seeing people as individuals. We can consider this point also alongside the metaphors of personalisation discussed earlier. Given the close relationship between digital monitoring, dynamic interfaces, and engagement with music (and many other everyday experiences) – all of which in turn feed further data collection (and the process repeats) – we can begin to see important connections between the roles of these systems and the experiences we have of our place, actions and even our understanding of the world around us. Following Cheney-Lippold we may ask whether measureable types make it

more difficult to believe in your identity as – for example – a young man who likes jazz (or an old woman who likes rock), if one of the places in which we perform and experience such identities – i.e. the streaming interface – creates categorical boundaries that are changing in computerised real-time, rather than negotiated in social time.

What is particularly informative about Prey and Cheney-Lippold's observations in terms of my own research is that together they point towards the idea of an analysis that should seek to pick at exactly the problems they describe regarding shifting conceptions of experience and identity. Such a route may enable a useful exploration of how publics and individuals are constantly (re)constructed and how experiences (of popular music, of identity, of choice and agency) may alter as a consequence of what van Dijck calls the "digital transformation of sociality". We can see the constant, real-time individuation process Prey describes when looking at Spotify's Discover Weekly service. It is an example of how digital constructs of human subjects experience the real-world consequences of their classification and other analytical outcomes, and within interfaces that are designed to facilitate the constant repetition of that process. Here the line between the virtual/abstract and the real world becomes blurred, and it is here also where the technologies involved with online music environments have real-world consequences.

2.5 Theorising digital music space/place

These consequences can be considered further by thinking through ideas related to one specific but common end product of digital monitoring and data analysis, which is the practice of predictive modeling. This is the process whereby statistical analysis of data helps to create models that are used to predict future behavior or trends, with the intended outcome of guiding or otherwise influencing behavior or trends through the foregrounding (or not) of certain types of content through dynamic digital interfaces. Spotify's Discover Weekly is

an example of this. Both the following or rejection of recommendations by users generates new data points that inform new iterations of predictive models that incrementally improve their efficacy. We can return to Prey's question, then, of whether the recommendation engine gets it 'right' or whether the user is a construct of the process. More broadly, however, the epistemological question raised by such a process is as follows: if, from gathered and analysed data, it is possible to predict or guide what happens next, and if the efficacy of a recommender system can be constantly improved through further data collection and statistical modeling, then is a theoretical understanding of what occurs, or may occur, still required? To put this another way, if life is constantly monitored and analysed, does it ever reach the point at which an understanding of its past, present or future, direction is needed? If the computerised real-time, all the time of the interface is the temporal reality, is there any way of stepping outside of this structure? And where does this leave us in terms of individual and collective agency and choice?

We can also consider the differing results we each see from Google searches based on our individual search histories (and the measurable types the analysis of those histories place us within), or the manner in which – as Tufecki shows – social media feeds can be tailored to what are understood to be our interests. Alongside this we should also consider the manner in which many of us are willingly participating in the processes that facilitate these systems through our regular, everyday use of them. Google, Facebook, Spotify and several other services based on Hartmann et al's DDBMs, have become, to greater or lesser extents, common, everyday and mundane. We can begin to see why practices of digital monitoring, algorithmic data analysis, dynamic interfaces and their growing centrality to both popular music listening and other areas of everyday life are potentially problematic. Specifically, we may ask what are the conditions that mark out the current configuration of technologies, practices and everyday use for our attention. And what can an examination of these new conditions tell us about ideas of choice and agency as they relate to music?

A starting point for this is Meija's (2012) work in which he argues that an

increasing portion of our communication is "structured by an (online) environment in which we cannot be present; we can only engage with an interface, which then communicates with others on our behalf" (2012:113). Clearly we can see this through what has been discussed so far, when cultural practices associated with music reception, such as listening, discussion, purchasing records and tickets, reading reviews, discovery and so on, are mediated through exactly the kind of interfaces Mejia describes. We have seen also, through Prey, how these interfaces act as a two-way conduit between, on the one hand, the systems of digital monitoring and analysis that reduce, abstract and derive positivistic insight, and on the other our real-world, lived experiences. Broadly characterising this latter activity as "movement", where "choosing to walk down one path as opposed to another adds both a content and a form to that choice which would not have existed otherwise" (2012:113-114), and – following that – where movement is a means of arranging social reality and as such "modes of governance have coalesced around its regulation" (ibid), Mejia builds on De Certeau (1984) to suggest that rather than being a natural phenomenon (the choice to take a particular path; the decision to listen to a particular song), movement instead "resonates with questions of power" (ibid).

It is through his next move, however, that Mejia offers a way of thinking about what may be the central contemporary problematic of this on going negotiation. He argues that it is the advent of mobile information and communications technologies (mICTs) that are the key development because of the way these technologies have been designed not only to be used in 'real-time', all the time, and as conduits between the real and virtual worlds, but – crucially – to do so within existing, real-world spatial arrangements. Mejia demonstrates this by arguing that previous mobile music technologies such as the Walkman and the iPod – unconnected devices, in other words – are primarily used, as Michael Bull (2006, 2000) has shown, to transform the aesthetic experience of a particular place without significantly altering the structural experience of that place. For our purposes here we may also add other items to that list, including the transistor radio, the portable record player, and even the piano in the parlour. Mejia claims that mICTs differ in that, along with replicating the functions of

their predecessors, they also engage directly with the present spatial arrangements we experience. Mejia argues that because of our increasing reliance on mICTs as a means through which we conduct aspects of everyday life, users and non-users in physical space are reduced to avatars in virtual space: the landmark in the city is a dot on the screen in Google Maps; the album is a thumbnail in an iTunes window; the human navigating towards both leaves a digital trail; and through digital monitoring, the mICT interface reacts to that trail in real time, all the time. Considering Cheney-Lippold's measureable types once more we may think about, for instance, the manner in which mICT applications that provide us with directions are able to react, and redirect us, according to traffic or other information. This type engagement with dynamic mICT interfaces necessarily relies on spatial arrangements being quantified and converted into the code that computational systems can understand. Because of this Mejia then moves to suggest, via Manovich (2001), that the possibility thus presents itself that that which can be reduced to digital code can also be programmed. Through Mejia we can begin to see where the negotiations between Lefebvre's abstract and social space take a "computational turn" (Berry, 2011). This is where the new development resides, and it is from this point where questions of choice and agency need to be revisited. It is here also that De Certeau's theoretical conceptions of place and space, strategies and tactics, are useful in terms of understanding contemporary conditions, and in terms of framing my own work.

Particularly this relates to both the "movement" I will study – i.e. my focus on the experiences of listeners through The Harkive Project and the manner in which Harkive gathers and interprets data about that movement. De Certeau's concepts therefore require some explanation in terms of my own approach. It is here also where the issues of debate begin to become enfolded into issues of method, and thus enable me to address both through my work.

Place, for De Certeau, refers to an established area, and for him this was the city. We can view not only the established places of popular music consumption (the record shop, the Spotify interface) in the same way, but also the points at which

data is gathered through social networking sites, search engines, and other digital interfaces where music reception can occur. Each of these places has a set of conditions and regulations that both normalise and legitimise their arrangement and existence, and which facilitate the creation and regulation of those who enter them. Just as walkers in the city must keep to the pavements, or not drop litter, so the Spotify user must pay their subscription, and can only press the play, skip, or share buttons. The Harkive Project is also a place in these terms: it asks people how, where and why they listen to music, and it specifies a date on which to do so – this too can be understood as a *de facto* regulatory framework for the individual narratives and data Harkive collects. Further to this, it should be considered also that Harkive seeks to capture (a word not used accidentally here) acts of engagement with music that more often than not occur within the regulations, traditions, and legality of the commercial music industries. These too can be understood in terms of place. Such interaction is further shaped and regulated by the interfaces and modes of listening that engagement occurs in and through, which are also places. Ultimately, data in all of these cases is gathered through systems and interfaces that shape and regulate narratives (Gerlitz and Helmond, 2013); Spotify provides only the option to play, skip, share or add a track to a playlist in the same way that a tweet sent in contribution to Harkive can only contain a maximum of 140 characters. The existence and function of place, then, can be understood in terms of the construction of a particular spatial reality, ultimately rendered in computational code, that resonate with questions of power.

De Certeau's conception of space, meanwhile, "refers to the practices that exist within a given place that work to create, shape, modify, and/or destroy that spatial reality" (2012:114). A place can thus only be maintained if it is practiced, and we can recall here Prey's examination of the on-going conflict between abstract and social space, or the fact that Spotify Discover Weekly is only possible (as a place) because of users engaging in cultural (spatial) practices within the Spotify interface (another place). The disruption in the recorded music industries witnessed at the turn of the 21st century discussed in Chapter 1 serves as an illustrative point to demonstrate the fragility of place. Once people

began to circumvent the regulatory, financial and ethical frameworks underpinning the exchange of popular music goods, the spatial reality of place began to rapidly unravel. For a place to remain intact, then, the details of its creation and structure must remain suppressed, according to De Certeau. "Should this secret be acknowledged", continues Mejia, "the sanctity of the place would be undermined by the possibility of equally legitimate alternative spatial practices." (2012:114) This is precisely what has occurred in popular music at several points throughout its history. Home-taping, pirate radio and illegal downloading all illustrate what occurs once practices of space temporarily wrong foot regulatory scaffolds of place. The relationship between place and space is thus symbiotic, an on-going negotiation. Although place would appear to have the advantage of power structures that maintain it, space can nevertheless reveal their often illusory nature, or their structural fragility. The site of this conflict is everyday life and, in terms of the specifics of my work, the ways people engage in acts of music reception.

De Certeau's concepts of strategies and tactics, meanwhile, enable us to look beneath the surface of the on-going place/space negotiation. Strategies can be understood as the mechanisms by which powerful entities maintain place through regulating 'movement'. Tactics are instead movements that occur within spaces that often (although not always) help to reproduce (or dismantle) places. They are a means by which individuals and groups carve out a space that is tolerable, liveable, and more human within regulated places controlled by strategic action. De Certeau talked of the movement of walkers within the city, of people cutting across wasteland rather than following the paths deemed legitimate by the regulatory framework of place. Considering digital interfaces, we can view the music listeners who rejected or else reconfigured legitimate spatial practices, most recently through the widespread practice of P2P downloading. In each instance what often follows, whether in the city or within the interface, are tactical attempts to reconfigure place towards strategic aims. After instances of tactical movement, and according to De Certeau's model, strategic action will attempt to close off the possibility for emergent, non-sanctioned practice. Consider here the construction of a fence preventing access

to wasteland, or legislation such as the Digital Millennium Copyright Act[10] intended to curb music piracy. Alternatively, non-sanctioned tactical activity can sometimes be incorporated into a strategic whole in a manner similar to that described by sub-cultural theory (Hebdige, 1979), where new practices (tactics), such as styles of music and fashion, emerge from certain groups as an attempt to react against, or differentiate themselves from, dominant ideologies. Those that do not dissipate are eventually co-opted by powerful interests of strategic control. The development of music streaming has been precisely this movement in action: a tactical spatial practice (illegal downloading) based on a technology that helped bring it about (P2P delivery mechanisms) eventually informed the strategic creation of a mode of music consumption that has ultimately redirected spatial practices in a direction that maintains the fabric of place. Streaming has its roots in tactical practice, but is now the dominant mode of legitimate, strategic music consumption within legitimate places such as Spotify.

Through digital monitoring and mICTs, and through the fragility of ideas around identity, agency and choice brought about by practices discussed above such as predictive modelling that in turn help facilitate the constant recreation of measureable types, the distances between ideas of place and space, and thus between the strategies and tactics that form, maintain or dismantle them, can now potentially be reduced through much quicker means than social negotiation. The symbiosis between the real and imagined locations and practices of everyday activities cross-fertilise in real-time/all the time as well as in social, lived time, in a manner that has the potential to remove elements of the social negotiation stage entirely. Tkatz (2015) has demonstrated such a process already exists to an extent by showing that digital interface design can be used to remove the possibility for unwanted actions. In other words, the very possibility of tactical action can be removed. It is through this that Manovich's (2016) conception of the programmable user becomes not just a possibility but instead part of the strategic arrangement of place.

[10] A 1998 piece of US legislation, available here: https://www.copyright.gov/legislation/dmca.pdf

2.6 – What does this mean for popular music?

Clearly, however, we are not yet at this nightmarish stage, either in terms of the ways in which we engage with music or in aspects of everyday life more broadly conceived. Even though people are increasingly engaging with music (and related activities) within and through systems that can both reduce action to quantifiable metrics and potentially facilitate the "programmable user", we would do well to pause and remember once again Sterne's observation (2006) that – when it comes to music – much of the act of listening occurs in the non-digital parts of the physical world. We may pause also to consider that Harkive, although capturing data through similar digital interfaces, is able to capture discourses from respondents that detail both digital and non-digital engagement. The question remains, however, of what happens when listening, and many of the activities associated with music reception can be monitored, measured, strategically countered, and thus potentially programmed. If we are to follow Mejia and Manovich, and consider also the work of other writers discussed earlier in this chapter, including Prey and Webster, then there are potentially serious implications for the cultures of music. What might those implications look like?

We can consider, for instance, that there are only a limited number of musical building blocks (musical notes, chords, generic conventions, etc.) that together – in the main – form the basis of the material (the songs, recordings, performances) that people engage with. Further, this engagement often takes place within an economic and regulatory framework – copyright – that effectively polices the configuration of those limited building blocks. This ongoing negotiation can also be understood in terms of De Certeau's model of space/place and strategy/tactics, and given that, we may wish to consider what it is that has driven, or otherwise allowed popular music to have been so endlessly (spatially, tactically) creative within those constraints. Although followers of Frankfurt school critical theory would argue that this creativity and the perception of it is at best illusory, later scholars (see: Bogdanovic and

Longhurst, 2012:6) recognise that ideas of the role of listener agency, choice, meaning and taste are missing from that critique. I would speculate here that what has consistently driven what many perceive as the seemingly boundless creativity in recorded music has rarely been the various strategies of place. In other words, innovative movements that emerge from the status quo are rare. It has instead been the spatial movements that belong (initially, at least) to those who operate tactically within a place that have driven (and perceived) that creativity. Although, as Wall (2013) points out, the idea of the dichotomy between 'mainstream' and 'margins' as an exploratory system is an oversimplification, it does allow us to understand that "there are differences between different parts of popular music culture"(2013:9). I would offer speculatively that it is largely unregulated cross-pollination of ideas, cultures, and technologies – and sometimes the unexpected uses of them – that emerge from the cultures of music emanating from 'users' that have almost always beaten paths to new sounds and styles, and thus to new experiences. Mason (2009), for instance, demonstrates how the unexpected, unsanctioned and what we can understand as tactical uses of digital technologies, which commercial interests have then strategically reacted to (or not), are important factors in what has helped to create much of the present day landscape of music. We may ask: is this potentially at stake once what Mejia and De Certeau call 'movement' becomes programmable, or once the negotiation between place and space has its vocabulary limited by Tcatz's dynamic interfaces, or the confidence that underpins starting positions is undermined by Cheney-Lippold's measureable types? As De Certeau shows, it is those cutting across the wasteland who are likely to produce the new pathways (Finnegan, 2007), not those who walk along the main street and wait obediently at every red light. The possibility of the 'programmable user' should therefore be considered in these terms. It is perhaps a stretch to call it a threat, but at the very least we could consider it a new (and potentially disruptive) element within popular music cultures[11]. It is for this reason, and for the reasons discussed throughout this chapter, that I suggest we should seek to understand the consequences of the new strategic

[11] A recent article in the Guardian newspaper in the UK, which reported on the use of artificial intelligence in the creation and recording of music, illustrates further questions regarding creativity and our perception of it. https://www.theguardian.com/technology/2017/aug/06/artificial-intelligence-and-will-we-be-slaves-to-the-algorithm

systems of place that are associated with music reception. As shown above, Vanderbilt observed that a common feature of computational and statistical systems operating a Hartmann et al's (2014) DDBM approach is the removal of marginal outliers and the arrival at a statistical consensus. It is precisely these systems, in the form of Spotify, and in the form of the mICTs we use to access Spotify, that we can now see occupying central roles in the ways in which people engage with music and thus derive meaning, perceive creativity, exert agency, and so on, through that engagement. We might think here about pausing to consider the type of reception cultures such systems could eventually help to produce[12].

To examine this in more detail we may look once again towards the work of Straw (1997a), which was prescient in this regard in its examination of the emergence of music retail 'megastores' in the mid to late 1990s. Straw shows that these stores were intended to feed and grow the then-lucrative market for CDs and back-catalogue by helping to guide the bewildered consumer through the problem of what Straw described as selection stress. This can be understood as the 'problem' of having too much choice leading to an inability or lack of confidence in selection behaviours. With much more of music's back catalogue available to users today via digital services than within the megastores of the 1990s, the potential for selection stress has by no means diminished since the time of Straw's work. His examination of the role of computerised, data-driven rationalisation within the music retail environment can help us to think through the issues raised above. Despite his observation that economies of scale afforded the possibility of the delivery of an "authentic" experience to some individuals in specific generic segments, his prognosis for popular music culture more widely was less than positive:

> As record super-stores beckon with their pluralist abundance, magazines, radio formats and the broader logics of social differentiation have circumscribed tastes and buying patterns within predictable

[12] On August 5th the US-based artist, Kyle McDonald, posted a number of data visualisations to his Twitter feed based on his analysis of the Million Song Dataset (https://labrosa.ee.columbia.edu/millionsong/). The visualisations appeared to indicate a growing degree of homogeneity over time in terms of the BPM and 'loudness' of recordings– see: https://twitter.com/kcimc/status/893853684676702208

> clusters. What has been lost…are those...moments of crossover and convergence. Stumbling around the record superstore, 'lost, driven crazy', the paths we follow are likely...to map the stubborn lines of social division (Straw, 1997a:65)

In one sense, here, we can consider again the critical theoretical position of the Frankfurt school and draw a neat line back to Adorno's (1942) critique of the culture industries and of standardisation, or to Benjamin's (2008) work of art that loses its 'aura' through rationalised, industrial production. Straw is similarly arguing that art should have the potential to bring us closer to a form of truth, knowledge, or beauty, but that commercially disseminated art somehow impedes that process. It is worth pausing briefly to consider this idea in light of the issues raised in this chapter regarding engagement with music in an age of digital monitoring, data analysis, predictive modelling, and automated recommendation. Spotify Discover Weekly is a case in point: it raises concerns over the programmable user and could be considered in terms of Straw's work also in that it is entirely based on the idea of predictable clusters in similar ways to a 1990s CD store. But, equally, many (including respondents to this research project) have praised Spotify Discover Weekly for the quality of the recommendations it provides (see Pasnick, 2015). That being said, Straw's work, takes on additional significance when one considers Mejia and Manovich, and also De Certeau's model. In the twenty years since Straw wrote, we have moved from the 'chivalrous' guiding of consumers through the headache of selection stress, to a point where selection is automated, based increasingly on the real-time analysis of individual and collective activities that produce destabilising measureable types, and occurs within interfaces where the possibility for tactical movement is diminished. If we are to take Straw's warning seriously, and also consider the potential benefits some see in the efficacy of – for example – automated recommendation, we should examine the conditions of and forms through which such 'chivalry' now takes place. This in turn leads me to the suggestion that we should seek to examine what happens when the cultural and the computational meet in an always-on, ubiquitous (Kassabian, 2002) networked environment, and how that relates to ideas of choice and agency in the lived experience of the everyday.

For Lynch the present networked age is characterised by "information [that] doesn't come in discrete packages but in a structured whole"(2016:111). Using Harry Beck's 1931 design of the London tube system map, which ignores the physical geography of London and instead prioritises the nodes of the travel network in abstraction, Lynch argues – and here we can again consider the mechanisms and outcomes of systems of digital monitoring – that "thinking of something as a network is useful when what matters is a complex pattern of distribution points rather than the points themselves" (2016:113). Based on that he asks, in an echo of Prey's observation above, whether we are either "increasingly composing a knowledge of network, or it [is] composing us?" (ibid) Again, we could pause here to think about the relationship that exists between data collection, recommendation and listening within interfaces such as Spotify, and how this spills out into wider cultural practices. The question then becomes about which is the more powerful entity, an individual's knowledge, or the knowledge of the group encoded within the network. In terms of the process of selecting songs to listen to, do we trust our real-world, lived experiences or do we instead trust automated recommendation? The answer is, of course, that to varying degrees we have a degree of trust in both. One, however abstracted, represents the other. For Lynch the answer is also both, but only in so far as the extent to which each feeds the other: if individuals in a particular network are poorly informed, then so is the network. This is, of course, reasonable.

Considering, then, Webster el al's position that algorithmic processes influence publics through their role as cultural intermediaries, we can begin to consider some of the potential consequences. For Lynch these questions ultimately relate to the idea that the "growing networked nature of knowledge makes independent thinking more, not less, important" (2016:113), in much the same way that we may reasonably reject the recommendations of an algorithmic process based on our critical faculties and 'real world' tastes. These critical faculties, for Lynch, are based in the difference between the accumulation of information and our ability to derive understanding based on that information. A song is recommended to us, or a particular piece of content is foregrounded (or not), but to what extent do we understand the rationale behind that process,

particularly when it may have been based on the measurable type version of ourselves that exists in a place far removed from our lived experience?

Posing a similar question, Lynch suggests that it is the very availability and use of such technologies that makes critique of them so difficult. It becomes harder to step away and view such activity objectively, and to take responsibility and ownership of our identity, choice and agency, when we adapt to our digital form of life. Here Lynch offers a minimal definition of knowledge as a correct belief that is grounded or otherwise justified in order to guide our actions. Understanding, meanwhile, comes from a "creative insight into how..evidence hangs together, into the explanation of the facts, not just the facts themselves" (2016:14). In other words, having taken us from 'what' to 'why', our understanding subsequently leads us to 'which' questions we should ask next. For Lynch, that we appear to be exhibiting as a matter of course trust in what he refers to as "Google-knowing" is inherently problematic. Consider the automatic, instinctive way in which millions of us use relatively new digital, ubiquitously connected methods as our primary means of acquiring information (and music recommendations), that we in turn use these to guide our actions[13]. The root of this, for Lynch, is that to be reliably receptive to information about the world around us is crucial to our survival – it is how we and other animals acquire the basics of food, safety and shelter. This is something we instinctively do without knowledge or understanding of the action concerned, and with an innate and unreflective trust of our reaction to the information we receive – I am thirsty; this water is clean; I will drink. And so on. The difference, Lynch points out, is that processing information reflexively and exchanging information reasonably (both in terms of cordiality and in terms of the application of reason) are crucial elements in our route to understanding. He suggests, and I agree, that these qualities are at risk due to an over-reliance on an automatic acceptance of 'Google-knowing', to which we can also add automated recommendation. Alongside this we can consider the ideas discussed throughout this chapter, where many of us engage daily with interfaces that are able to change

13 A recent example from NYMag illustrates this point. Google search queries are auto-completed based on existing content, which makes false claims difficult to investigate since counter evidence is sometimes not available: http://nymag.com/selectall/2017/03/googles-dangerous-identity-crisis.html

dynamically based on new, available information that decouples our lived experience from our abstracted digital selves.

Further to this, is the idea of societal fragmentation brought about when like-minded groups, or 'tribes' as Lynch refers to them, coalesce around different types of information likely to support and enhance their existing beliefs. This hinders routes to conflicting positions, and thus a reasoned understanding of them, which Lynch argues is problematic in Kantian terms for democracy in that the reasonable exchange of ideas, opinions, facts, and so on, is the very means by which authority is evaluated. It is not so much, then, that we should all agree (on musical taste, on political opinion), but that we should engage in discussion of our respective positions. Lynch argues that the division of "tribes" into knowledge silos by Internet technologies is thus a further cause for concern. It can be argued this is also the case for music reception. Consider two users of digital music services who have their 'discovery' and thus subsequent experiences informed by commercial analyses of their tastes and activity. Then ask – as Webster et al do above – whether the result is likely to be ultimately divisive or cohesive. Even if division or cohesion as outcomes are unlikely (or even unwanted) polar extremes in matters of taste, the implications for the reasonable cross-pollination of ideas become clear when one tribe can exist in hermetic isolation from another, or when one 'user' can – as Prey suggests – be endlessly reconstructed as a measurable type for a wide variety of aims, none of which necessarily speak to ideas of the self or the collective. For Lynch not only are conceptions of social artefacts changing in light of digital technologies, including those related to property and privacy, but so too are ideas of the self, itself a social artefact. Flanagan and Dennett (Lynch, 2016:74) have suggested that the self is an ongoing narrative construction, and it is clear that music is a hugely important element of that process (see De Nora, Finnegan (2000; 2007)). However, because much of our narrative and narrative construction now takes place online and within interfaces that (re)produce and facilitate networked knowledge, there are ramifications for objectivity since our digital form of life obscures the boundary between the real and the virtual, or between what is made and what is found. Lynch describes that which is made is being

understood by its function, and that which is found to have individualised meanings that differ from person to person. Ultimately, Lynch asks, "if the real is virtual, [then] how important can the truth be?" (2016:74).

2.7 – How can we research this?

An immediate problem with forming an enquiry based on an engagement with the issues and debates explored in this chapter, and of using Spotify Discover Weekly as an exemplar, is that the scale of the activity engaged in by Spotify and other processes of computational knowledge creation, are difficult to grasp. In the case of Spotify, the daily interactions between twenty millions songs, 100m users, and 2Bn playlists produces a large amount of information that can only be processed through complex socio-technical systems (see: Webster et al, Prey). These systems have commonly been understood through the term 'big data', practices related to which are evident in the activity of Spotify in the creation of their automated Discover Weekly playlists, and the manner in which users of digital and mICT interfaces are "simultaneously the informed, the informant and the information" (Michael and Lupton, 2015:1) . As this chapter has shown, together these raise interesting new issues and debates around identity, choice and agency. A useful next move, then, may be a normative understanding of the processes and systems through which such activity occurs.

The collection, processing and analysis of data as a key resource, and the implications this has for issues of identity, agency and choice, leads Lynch, Tufecki, van Dijck and others ultimately to the position that some form of transparency and accountability in terms of the internal mechanisms of data systems would be desirable, perhaps even beneficial. But is such a thing even possible? Vanderbilt (2016:53), for instance, suggests that the creators of the algorithmic processes that aim to deal with the abundance of information that companies such as Spotify gather, "have admitted that ever more complex mathematical regimes can become, in effect, HAL-like "black boxes" whose precise behavior can no longer be actually determined or predicted". There is contradictory evidence, however, from the computer science community that

challenges such bleak conceptions of who (not what) controls computational systems. Dunietz (2016)[14], for example, observes that although processes such as machine learning are becoming increasingly more sophisticated, they are still largely guided by human hands: "Choosing a learning algorithm just means choosing which patterns a machine will be bad at". Presented with a dataset an algorithm may help an analyst reveal infinite possible patterns within it – far more than a human alone could – but an algorithm presently lacks what Lynch would call the understanding required to select the correct pattern for the problem at hand. The learning element of machine learning, then, is the unknown element, and in Dunietz's view, "we'll be teaching machines to teach themselves for many years to come". I raise these conflicting observations here because it is indicative of wider debates around data and related technologies that, as van Dijck (2014) has suggested, clouds our thinking when attempting to examine them. Ultimately this is a human problem, and one that requires a socially constructed and negotiated solution, and in acknowledging this we can then move towards questions of how we as researchers may critically approach this. Potential routes through this are discussed in greater detail in the following chapter but have been described by Berry (2011:1) as the experimental processes by which scholars attempt to:

> "take account of the plasticity of digital forms and the way in which they point toward a new way of working with representation and mediation, what might be called the digital 'folding' of reality".

The work Berry is talking about here concerns itself, to greater or lesser extents, with the relationship between human experiences and data systems. Berry offers a note of caution that accompanies this that is worth bearing in mind before proceeding. He states that the process of representing complex elements of real-world experiences through the medium of data necessarily leaves some elements out. It is a process that involves, at almost every step, reductions and abstractions. The choice to play a particular song within a digital interface can,

[14] http://nautil.us/blog/the-fundamental-limits-of-machine-learning?utm_content=buffer9ffc0&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

for example, be recorded as a series of data points that include the listener's location, the device they are using, and so on. From this positivistic insights can be derived – people in group *a* like *x* music in *y* weather, people in group *b* do not. We see precisely this occurring in Spotify with automated recommendation. Such processes, however, either exclude or are else incapable of capturing other elements related to those choices – the smell of a certain food that brought the song in to mind, for example, that may occur in Sterne's non-digital parts of the physical world. As Berry observes, 'a computer requires that everything is transformed from the continuous flow of our everyday reality into a grid of numbers that can be stored as a representation of reality which can then be manipulated using algorithms' (Berry, 2011: 2).

At each stage of this process something is removed, or else interpreted, grouped, abstracted, statistically smoothed out, or otherwise processed, to produce the actionable knowledge or insight. We can see, then, that the methods for understanding reality and generating knowledge via digital mediation are inherently subtractive in nature. This has been identified by Berry as the key problematic in work of this kind, whether academic, commercial, cultural, or computational. The issue thus becomes two-fold: we must first consider what kind of knowledge is produced and how it is produced, before we can consider what sort of realities these may lead us toward.

This in turn further highlights the difficult methodological and epistemological issues related to a research process that seeks to engage directly with these technologies: if the generation of knowledge via digital techniques is problematic, how can a research project that seeks to critique and examine Berry's key problematic achieve its aims if that project uses the very techniques it must acknowledge as being flawed? We are thus presented with the problem of how we can study ideas related to datafication, using the techniques of datafication, without replicating and reinforcing the assumptions inherent in datafication. In so doing we acknowledge, for instance, that any knowledge generated through such a process must also be reductive, regardless of how robust and well designed those processes are. What follows from such a realisation is the temptation, perhaps even the logical decision, to wash our

hands of such approaches and seek alternate methods. As my opening chapters demonstrate, however, the growing influence of data-derived knowledge in many aspects of our everyday life – and certainly in terms of how we engage with music – suggests that work that seeks to examine the contemporary cultural, technological or commercial conditions of that engagement must look closely at and engage with processes of data-derived knowledge. This is an approach that is not without issues, and the next chapter will explore these in more detail.

CHAPTER 3

‘Polluting proximity’ – researching data technologies

In Chapter 2 I demonstrated why new research that can critically engage with the technologies and associated practices linked to recent changes in music consumption is necessary. In this chapter I will explore the many issues related to how that research may be undertaken. Specifically, the issues of how we may go about researching the technologies that now underpin the major means by which music is distributed and consumed, and what that research may be able to tell us about the role of those technologies in terms of our engagement with cultural goods and everyday life. To begin that process, we may look once again towards another recent development from the field of popular music.

3.1 – Content targeting

On 20th July 2016, an announcement from Spotify further highlighted new technological and commercial developments in data collection and analysis that directly relate to music reception. Under the headline, “Spotify launches programmatic audio globally”, their website explained how advertisers were being given access to the 70 million users on Spotify’s free tier. Through ‘content targeting’ – the delivery of individually targeted audio adverts between songs – media buyers would be able to “reach users with particular habits, mindsets, and tastes that align with your target persona[1]”.

[1] Quote taken from Spotify’s https://www.spotify.com/us/brands/targeting/

This kind of relationship between audiences, media outlets, and advertisers is as old as the media industries, as Lears (1995) has shown. What the Spotify announcement highlights, however, – and we can recall here specifically Mejia's work around mICTs discussed in Chapter 2 – is a change in the degrees of reach and in the spatial-temporal conditions of that relationship. An advert can now be placed at the level of an individual, and according to where and when that individual is at a given moment[2], and this adds new dimensions to the ways in which the audience segments can be constructed and conceived. With advertising revenue linked to digital monitoring technologies, audiences for certain types of music, genres or playlists, for example, can be understood in new ways, and some segments can be better understood than has previously been possible as demonstrably more or less valuable than others. According to AdAge, a trade publication for the advertising industry, "an ad buyer can now tweak and serve audio creative to a user based on what they know about the user – listening habits and log-in data from Spotify and first-party data from a client, for example – in a matter of milliseconds"[3].

The benefit to advertisers is the possibility that individuals choosing particular songs allows for the delivery of adverts based not only on those songs, but also on users' locations, the devices they are using, and other information gathered about them. Spotify's head of programmatic content, Jana Jakovljevic, told UK trade publication Campaign, "we probably have the most unique insight into our users…We know when users are commuting, know what devices they are listening from. We know that they might be interested in cooking because they have a cooking playlist, and I think very few publishers can make that claim"[4]. The relative novelty of content targeting is further illustrated by the fact that it was only in March 2016 that the International Advertising Bureau (IAB) released a new protocol for their

[2] The idea and technology behind this move was further illustrated in March 2017 when Spotify, working with clothing manufacturer North Face and the band White Denim, made a song available only to users who were in locations where it happened to be raining at the time: See https://www.fastcompany.com/3068935/spotify-and-the-north-face-collaborate-to-make-a-song-only-available-in-the-rain
[3] http://adage.com/article/agency-news/spotify-ramping-programmatic-efforts-ad-tech-partnerships/305019/
[4] http://www.campaignlive.co.uk/article/q-a-programmatic-audio-will-change-game-spotify/1403452#fEtfHXRyCeXD1Elc.99

delivery[5]. In terms of the wider financial context, the global annual market for ad content of this kind is predicted to be in excess of $20Bn[6] a similar figure to total global revenues for recorded music achieved in recent years (IFPI, 2016).

We can consider the above alongside the changes outlined in Chapters 1 and 2, where popular music has been shown to have relatively recently moved towards models of consumption based on subscription and/or attention models, and in terms of how the consumption of media products and many other everyday social and cultural activities, including shopping and administrative tasks, now take place in and through online environments where users willingly or otherwise leave behind data, and where data "capture, analysis and output are integrated" (Rieder, 2016). By using music streaming services as a means of framing my opening chapters I am not, however, concerned with that particular sector alone. Streaming is used here not to focus attention entirely on modes of music reception that occur within digital interfaces but rather to demonstrate how methods of engaging in acts of music reception are changing in ways that highlight what is a growing relationship between everyday experience and data-driven systems.

We can acknowledge, for instance, that just as everyday life continues in the 'real' world alongside our activities in the 'virtual', not all experiences related to popular music occur within online environments. We may recall here also Sterne's observation from Chapter 1, that much of the act of engaging with digital music technologies takes place in the non-digital parts of the physical world. There are still many ways in which music can be experienced that do not directly involve digital delivery mechanisms. Streaming has not, for example, brought about the complete disappearance of CDs, vinyl, or even sheet music, and live performance remains both culturally and commercially relevant. However, through our use of services such as online retailers, ticket sellers, social media platforms, search engines, and so on, we are creating

[5] http://www.iab.com/wp-content/uploads/2016/03/OpenRTB-API-Specification-Version-2-4-FINAL.pdf
[6] This figure is according to IPG's Magna Global ad spending forecast, and was reported in the AdAge article linked in previous footnote.

new links between those existing practices and newly emerging systems of data collection in what Maguadda describes as a "circuit of practices" (2011). This leads us to interesting issues of debate in terms of research, briefly introduced at the end of chapter 2, and dealt with in greater detail here.

3.2 – The problem of researching data technologies

Taking Sapsford's definition of methodology as the "philosophical stance or worldview that underlies and informs a style of research" (2006:1), the approach I will develop in Chapter 4 utilises an experimental analysis based on similar data collection and computational analysis methods to those used in commercial settings to inform and explore new forms of knowledge about contemporary experiences of popular music. Rather than positivistic testing of hypotheses, my analyses will instead be driven by a practice-led exploratory methodology that will attempt to identify, explore and address the issues inherent with such an approach. This experimental mode of performing research has the potential to reveal new insights, but the aims and reasoning behind my approach needs to be explained further.

A starting point is Hall's definition of the computational turn in humanities research. This he describes as "the process whereby techniques and methodologies drawn from computer science and related fields..are used to create new ways of approaching and understanding texts in the humanities" (Hall, 2013:2). By using some of the techniques identified by Hall – specifically the "mining, aggregation, management and manipulation of data' and 'interactive information visualisation" – we may begin to explore the ways in which such processes and technologies are shaping experiences of popular music. By using these techniques critically and reflexively, we may also develop a better understanding of the role of those same techniques in the process of undertaking research. I am thus situating my work at a position where the fields of digital humanities, cultural analytics and social computing intersect in an attempt to form a new variant of popular music studies.

As such the developments at Spotify described above provide a useful means by which we may begin to develop questions about both the conditions of contemporary engagement with popular music, and the manner in which we may research them. Despite the greater amount of user data available to companies and the incremental advances in computing power that enable them to process this data in the furtherance of a competitive advantage, the marketing of cultural goods remains a process that contains elements of guesswork and gut-reaction, a feel for the game, that even the most technically well-equipped cultural intermediary relies upon (Bourdieu, 1984; Webster et al., 2016). Largely this is due to the unpredictable ways in which consumers respond to cultural goods, and how elements of that complexity do not – even in the present day environment – easily lend themselves to digital monitoring or measurement. We can think here, for example, of how an individual's memory of a person or a place may inform their choice of a song. A digital interface may record that a song was played, along with the user's location, and other information, but the memory (or some other internal/external trigger) remains unmeasured. Nevertheless, content targeting and digital monitoring, along with the automated recommendation services discussed in chapter 2, invite us to look at the related processes of production, distribution and consumption in new ways. A radio advertisement, a high street billboard, a magazine advert, or the release of a new single by a recording artist, may appear speculative when compared to an individually targeted commercial message, delivered at the optimal time and place, and based on constantly updated metrics. Messages can be foregrounded only to those groups for whom advertisers, marketers and other commercial interests have specifically identified a likelihood of success, with unsuccessful messages used to iteratively improve both delivery models and content. As was also discussed in chapter 2, this recalls and repurposes debates around agency and choice because this is a process facilitated by technologies of data collection and analysis capable of operating at the level of the individual that have only recently become available.

Systems such as content targeting demonstrate that commercial operations and the practice of engaging with music are now linked in new ways because the choices people make about the goods they consume are now more closely monitored than ever before. This can lead to recommendations and the foregrounding of particular content which in turn helps generate new data, in real-time, all the time. Chapter 2 demonstrated how this process is self-referential, in that consumer responses to cultural goods ultimately inform the types of cultural goods they are subsequently exposed to, but considered Cheney-Lippold's (2017) notion of measureable types as a problematic in that process. Indeed, despite what may be seen as the benefits of recent technological developments for music listeners – the ability to listen to huge catalogues of recorded music on demand, for instance, or the ways in which we can easily purchase concert tickets or vinyl discs with a single click – the questions remain as to what the long-term outcome of these changes will be.[7] In particular they raise the issue of what happens to the ways in which we experience popular music, and to everyday life more generally, when online platforms and related digital monitoring technologies – including algorithmic processing – become, as Van Dijck & Olsson suggest, "not transmitters but rather producers of sociality"? (van Dijck and Olsson, 2013:57)

The growing role of technologies in this process suggests that our focus should necessarily shift towards the systems concerned, as Prey (2015) and Webster et al's (2016) work discussed in the previous chapter does. This in turn leads us to debates about how such work may be undertaken, and van Dijck argues that we must "look into the credibility of the whole ecosystem of connective media" (2014:298), a position that ultimately leads to questions regarding the type of critical questions we should be asking. For van Dijck this means considering the ways in which we can critique ideas related to what Mayer-Schoenberger and Cukier (2013) call datafication in a manner that avoids replicating and reinforcing the assumptions, positions and beliefs

[7] A recent investigation by journalist Liz Pelly into the ways in which playlists are foregrounded by the Spotify interface is a case in point. It revealed that, far from being the meritocratic system understood by many users, artists/labels and playlist creators, songs and playlists from major label partners are increasingly foregrounded https://watt.cashmusic.org/writing/thesecretlivesofplaylists. We may also consider here Susan McWilliams' update of Chistopher Lasch's critique of progressivism - https://home.isi.org/sites/default/files/MA58.4_McWilliams.pdf

that underpin it. The starting point for this is what van Dijk refers to as the "digital transformation of sociality" (2014a:199) , which means not only examining the commercial realisation and conception of value inherent in data, but also the perceived value data and related technologies have for new types of academic enquiry.

In other words, we cannot begin to critically engage with systems such as Spotify's content targeting, or automated recommendations, until we know how to critically engage with them, and this means we must acknowledge the starting points (the assumptions, beliefs, and so on) of our own modes of enquiry. This requires a degree of detached, objective scholarship that is complicated by the availability and use of similar tools, techniques and datasets to those used in commercial settings – in other words, the objects of such a study. In the case of my own work this is exemplified by the functionality of the Harkive Project and the methods that underpin my enquiry: I am using digital platforms and data collection techniques as a means by which to examine the implications of digital platforms and data techniques. For van Dijck, the pitfalls in an approach such as this can be observed in situations where researchers treat "social media data as natural traces and...platforms as neutral facilitators" (2014a:200). These are precisely the types of assumptions and beliefs that can fail to take into account the reductive and productive processes inherent in the transmission of real-world activity into the creation of a set of data points raised by Berry (2011) at the end of Chapter 2, or – in other words – what is understood as datafication.

Consider here the complex set of individual and contingent choices involved in the decision to play a particular song, and then the extent to which data-derived abstractions – however complex – can be said to faithfully represent those choices. Could those data points alone be used as a means of examining ideas of musical choice? As van Dijck observes, "promoting the idea of metadata as traces of human behavior and of platforms as neutral facilitators seems squarely at odds with the well-known practices of data filtering and algorithmic manipulation for commercial or other reasons" (2014a:200). The

manner in which digital interfaces demonstrably influence or otherwise potentially direct activity, particularly in terms of how they foreground (or not) particular content to particular audience segments (see Tufekci, 2015), is but one of the ways in which data is not so much collected, but rather produced (see: Gitelman. 2013). In much the same way, we can see the role that interfaces have upon experiences when capture, analysis and output are integrated (Rieder, 2016). As Boyd and Crawford (2012) have argued, by using what is essentially corporate data as the unreflexive basis for enquiry (where a tweet is treated as a true reflection of a belief), such an enquiry is inherently flawed. The very process that seeks to speak to power – i.e. independent, critical research – instead becomes part of what they describe as the "deep government and industrial drive toward gathering and extracting maximal value from data, be it information that will lead to more targeted advertising, product design, traffic planning, or criminal policing." (van Dijck, 2014:14). There are, then, some important methodological and epistemological issues that need to be addressed before we can begin to properly account for broader questions about the role of digital and data technologies in music reception. These questions revolve around the issue of how we can understand and critique the problems of datafication, using the tools and techniques of datafication, without falling into the traps of datafication.

The difficulty of this "polluting proximity" (Liu, 2016) between object and mode of study is further complicated when we consider Van Dijck's (2014a) argument that the trading of data and metadata in return for communications and other services has "nestled into the comfort zone of most people", in what Barnes has conceptualized as a "privacy paradox" (2006): people are often uneasy about the data collected about them, but have come to accept it as a condition of using certain types of digital and online services. This on the one hand raises ethical issues about the data we collect (see: Zimmer, 2015, 2010)[8], but at the same time to consider that the widespread use of services

[8] The ethical concerns of my own work are discussed in full in Chapter 4

that engage in datafication also contributes to what Baym has called the domestication (2010) of such technologies. This is a process Baym describes as where technology goes from "wild to tame and gets laden with meaning along the way", and which often proceeds in a manner that replicates the conceptual and metaphorical frameworks originating from commercial vendors (see: Hagen, 2016). As Van Dijck observes, the widespread adoption of digital services and online platforms by publics is seen as being of great potential benefit to academics who see a revolutionary research tool that can help them investigate human activity through the availability of the data such activity produces, but we must also consider the potential issues that arise when the very processes that form the questions driving our research are also the ones that provide us with, and influence the content of, both the data and the analytical tools that underpin our methods.

If this conundrum were not already tricky enough, a further difficulty is raised by Ananny (2015) in his attempt to move us towards an ethical accountability of the algorithmic technologies intrinsic to data-driven business models. He argues that "existing approaches to media accountability that assume stable technologies and clear questions are outstripped by the dynamic and contested nature of algorithmic assemblages" (2015:16). In other words, they are a fluid, moving target. However, in considering the algorithms that lie at the heart of the systems operated by companies such as Spotify, Ananny provides a useful first step in developing the concept of NIAs (Networked Information Algorithms). This concept is an attempt to first of all differentiate algorithms from their mathematical definitions in the field of computer science in order to produce a conception that is not reliant on code, or indeed on human action using code, but instead on "an intersection of technologies and people that makes some associations, similarities, and actions more likely than others" (2015:2). Thus, when Ananny observes that NIAs have three key affordances in providing a means by which to organise, predict and influence publics, he arrives at a similar place to Webster et al when he conceives of them, following Latour (2005), as:

> [assemblages of] institutionally situated computational code, human practices, and normative logic that create[s], sustain[s], and signifies relationships among people and data through minimally observable, semiautonomous action (2015:7)

From here we can begin to understand that the internal components of a given analytical system (i.e. the code and mathematical logic of an algorithm) are largely meaningless in their own right to the critical eye: it is only in combination with other factors within an assemblage (the available data, the coders, the aims of the business or organisations concerned, and so on) that their implications or results can become observable. In much the same way, research that attempts to explore issues with computational systems must also consider the role and skills of the researcher, the disciplinary philosophies the researcher(s) emerges from, and other factors related to the research process, including funding bodies and the intellectual projects of institutions and research centres. This is a productive route since it enables us to consider both data-derived business technologies and the research process in terms of relationships comprising of identifiable parts that can be explored, but it also is an approach that is not without certain practical problems.

Considering first of all the data that routes through such assemblages, Andrejevic (2014) points out, citing Tim Berners-Lee's complaint that there is a "growing separation of people from their data" (2014:2), that an individual's data is largely useless when taken in isolation. This is because its 'value', or potential value, is only attained when it is combined with the data of others. This appears to be a defining argument and justification in the epistemological project of big data and datafication (and, indeed, some modes of research that harness similar technologies, which will be discussed later in the chapter) – that the potential value/knowledge inherent in a single unit of data is hidden, and can only be unlocked once the volume of available data is scaled. Knowledge, or the potential of it, is situated in the observations gleaned from the abstracted aggregate, and the problems with this were discussed in Chapter 2 in terms of the "removal of outliers" and the drive for a

"statistical consensus" discussed via the work of Vanderbilt. Data systems, argues Andrejevic, are omnivorous in their consumption of data because they have "embarked on the project of discerning unexpected, unanticipated correlations" (2014:4), which is precisely what can be observed in the use of data around music listening by Spotify. This, for Andrejevic, increases the divide between what Boyd & Crawford (2012) have defined as the data rich and poor, in a manner which outstrips concerns limited to questions of who has access to data – the data haves and the data have-nots – and extends to questions about how data is collected, sorted, stored and analysed, all of which have an influence on the (access to) knowledge produced by such processes. In other words, the availability (or not) of data to researchers, and the availability (or not, due to factors such as a lack of the required skills) of analytical tools such as algorithms that may help researchers understand their data, does not so much represent a potential solution, but rather a whole new set of issues.

If data in isolation is useless without more data, and more data requires exponential increases in collection, storage and processing power in the form of complex algorithms, researcher knowledge, and funding, it may follow that understanding can only be arrived at when taking into consideration – and attempting to replicate – the processes as a whole. In other words, the very process of undertaking research in this manner becomes a related object of study. This is by no means an easy process, and particularly in fields such as popular music studies, where such work has thus far been rare. Nissenbaum (2009) is instructive in framing the challenge of the current era as being the need to develop contextual norms for the use of data whose uses can be decontextualized in unpredictable and radical new ways. In a similar vein, discussions of a "coming crisis in empirical sociology" (Savage and Burrows, 2007), can be understood in terms of the historically privileged position of academic researchers to speak to power coming under threat from markets and governments having access to comparatively more data and better

systems that can facilitate the generation of a form of knowledge or truth[9]. These problems also present an opportunity, however. Savage, for instance, suggests a radical rethink of theoretical and methodological repertoires, which in part can be facilitated by the very thing that threatens existing ones: i.e. the data generated at scale by millions on a daily basis and the new means by which it can be analysed. As Housley et al argue:

> digital societies are self-referential, in the sense that they generate data as an accountable trace and functional pre-requisite for network and system integration (2014:3)

There is a huge amount of data available to academics that could augment existing approaches and enable the digital re-mastering of classic questions, according to Housley et al (Housley et al., 2014). In pointing this out, they argue, a collaborative approach is required: simply linking 'new' data to 'old' methodologies will not be sufficient, and will not take advantage of the opportunity (or address the problems) digital technologies and data systems present.

The problems outlined above are closely linked to notions of knowledge that, according to Lynch (2016), are not only changing in conceptual nature (in particular 'networked knowledge') but also in terms of shape. The printed word (and the academic paper) was once what Lynch calls a "stopping point", when that which was known at a given time was fixed. This allowed for the questions to be moved along until a new stopping point was reached, and so on. This is the philosophy that underpins academic enquiry, even now in the age of "networked knowledge". My own work, for instance, will ultimately be judged on the basis of its new contribution to knowledge, which implies a body of knowledge that exists previously and which is fixed. As Lynch points out, the issue with this – particularly in work that examines digital technologies – is that the Internet doesn't involve stopping points, it is instead constantly updating. As such, knowledge becomes fluid. It has

[9] The British sociologist, David Beer, writing in the aftermath of the 2016 UK European Union and US Presidential elections, and amid widespread discussion of the role of social media platforms in the respective outcomes, explored the question 'why is everyone talking about algorithms' http://discoversociety.org/2017/01/03/why-is-everyone-talking-about-algorithms/

changed in the time it has taken you to read this paragraph, and not in the time it has taken me to research, write and publish my work. We can recall here the discussion of Cheney Lippold's conception of measureable types from chapter2, and much like identity, or musical genre, knowledge itself is thus always new, shifting, fluid, and this in turn challenges foundational conceptions of what knowledge is. For Lynch the Cartesian pyramid takes on a different shape if its building blocks of information constantly flow, are boundless, and are (often) beyond the control of gatekeepers. The problem with this, according to Lynch, is that the reasonableness with which we evaluate the truth of knowledge is adversely affected if we do not step outside of the network that creates it: this is as equally true for the choices we make as music listeners as it is for the work we undertake as researchers. The need for critical thought becomes paradoxically more and not less important given the availability of information in greater volume, and Lynch suggests that it is not a question of how we know with and through these technologies, but of how we should know. This, equally, applies to the research process. Digital and Internet technologies thus raise questions not only about the responsibility we take for our own beliefs, but also the ways in which we work creatively to grasp and reason how information fits together. This is the epistemological problem facing both music listeners and the researcher seeking to study their experience.

Kitchin (2014) highlights exactly some of those epistemological issues that developments in data collection and processing bring into focus for researchers. Giving the example of census data as emblematic of large datasets as they have been previously understood, he points out the time-consuming nature of their compilation, their ultimately limited scope in terms of analysis, and – perhaps most importantly –that they were usually designed with specific questions in mind. What has changed, Kitchin argues, is that data sets are now dynamic and flexible, in that more data and variables are constantly added (in real-time, all the time), and that datasets are often created without specific questions in mind. In fact, datasets are often byproducts of activities such as music listening within digital interfaces,

or other forms of transactional and/or implicit activity data. As the experience of the music listener is increasingly channeled within and through the interface, the researcher's task becomes more complex.

Advances in computational power, largely centered around concepts such as machine learning, artificial intelligence and predictive modeling, are providing routes through the problem of what to do with such abundance, both for companies such as Spotify and for critical researchers. As Kitchin observes,

> Here, literally hundreds of different algorithms can be applied to a dataset to determine the best or a composite model or explanation…a radically different approach to that traditionally used wherein the analyst selects an appropriate method based on their knowledge of techniques and the data. (2014:2)

It is here, then, where the epistemological shift moves into a conceptual rather than purely practical gear, and its importance should be considered here in more detail before proceeding. It is the shift hinted at briefly in chapter 2, where we move towards a point where theories are not tested by recourse to data, but rather insight is derived 'born from' data. Although Kitchin questions the notion of a paradigm shift in and of itself, which he claims can produce "overly sanitized and linear" (2014:3) narratives, through this potential move beyond theory he nevertheless observes that we may be in a transitional phase towards what has been referred to as a 4th paradigm in science (Hey et al., 2009)[10]. In examining how such a shift relates to the humanities and social sciences, Kitchin notes first that the empiricist position that data can "speak for themselves", exemplified by Anderson (2008), has recently gained considerable traction within business circles[11]. Secondly he notes that alongside the empiricist position, a more nuanced version has emerged within the scholarly work of scientific researchers attempting to

[10] This idea suggests that scientific research developed from experimental modes of scientific enquiry that can be traced back to Galileo, through the theoretical science of Newton, and to the relatively more recent (20th Century) activities around the modeling of complex phenomena through computational science. The shift suggested by Hey is associated with the abundance of data and the increases in computational processing, and is referred to as data-intensive, or eScience.

[11] Hey, it should be noted, was and is a research executive president at Microsoft

harness the possibilities of data-intensive science to help make sense of complex phenomena. It is these related positions that can provide useful pointers for social science and humanities researchers who seek to explore and/or use similar technologies. My research falls exactly into this third category and so Kitchin's examination of the two epistemological positions of empiricism and data-driven science, and how these relate to the humanities, is of considerable use.

Kitchin's examination of the empiricist epistemology reveals that it collapses under the weight of its own logic on four key points: It is based on a presupposition that data captures a true, objective reflection of the whole, which is problematic since data is not exhaustive – it is instead an inherently reductive product of systems and interfaces. Secondly the empiricist position is not free from "the regulating force of philosophy" (Berry, 2011) – i.e. theory – since collection, analysis and so on, have their roots in scientific and other forms of thinking and training. Thirdly, because human interpretation is often required to sift through the results that are foregrounded by automatic process, it cannot be said that insight, knowledge or other products of such processing are necessarily free from theory, or of Berry's regulating force, because algorithmic processes produce innumerable correlations and patterns and it does not necessarily follow that all of these patterns will be useful or relevant. Dunicek's observations from chapter 2, regarding the realities of operating at the front line of machine learning, will resonate here. Finally, the idea that data can speak for themselves presupposes that domain knowledge is not required for its interpretation, which divorces insights from wider debates and knowledge within a given field. It is on this final point in particular where humanities scholars can be of use, a point that I will illustrate in my findings chapters.

Based on the above, then, we may once again return to Sterne (2006) and reframe his question on the "amount of digital" there is in "digital music" when much of the act of listening takes place in the non-digital parts of the physical world. This question can be repurposed and can instead ask the

extent to which empiricist, data-derived knowledge generation occurs in the non-data parts of the philosophical world. Sterne suggests approaching digital technologies from their "many exteriors", and – considering van Dijck and Boyd and Crawford's observations above – it appears that reflexivity and critical distance in the deployment of data research practices could provide us with a similarly useful vantage point. The reason this is needed, as Kitchin ultimately argues, is that the empiricist position "operates as a discursive rhetorical device designed to simplify a more complex epistemological approach and to convince vendors of the utility and value of Big Data analytics" (2014:5). It is a position driven by commercial hype, a concept that would be familiar to students of the music industries. We should therefore exercise extreme caution before blithely accepting, or unreflexively replicating, commercial rhetorics when undertaking work of this kind. As such, Kitchin's insights into the comparatively more objective and detached processes of data-driven science are useful in plotting our next steps.

The data-driven scientific position is emerging in scholarly work that employs many of the same technologies as empiricists. In addition Kitchin demonstrates that this mode of work uses a hybrid of "abductive, inductive and deductive approaches..that seek[s] to generate hypotheses and insights" (2014:5) in a manner that also takes data, rather than theory, as a starting point. Much of this has to do with the affordances of computational procedure and availability of data, but differs from other modes of scientific analysis in that rather than start from the point of theory and hypothesis, proceeding then to data collection and analysis, large data sets and computation are instead used to reveal possibilities for further enquiry, model building and theory creation. This is similar to the approach I will take in Chapter 5. Through such a process insight is not the end-point, as it is in empiricist epistemologies (and particularly those linked to commerce), but rather a step in processes leading to new hypotheses and further testing.

For the social sciences and humanities, however, the picture is less clear when employing these methods. Although Kitchin points out that positivistic

approaches to social sciences can benefit certain studies in terms of the greater availability of data and efficacies in analytical procedure (we can think, for example, of the field of economics here), the field he refers to as digital humanities is somewhat more complex. This is an area I will return to in more detail later in the chapter, but in the meantime it can be observed that the field relates closely to the alternative Kitchin proposes and is indeed closer to where I am attempting to situate my own work. It is a field that acknowledges the epistemological issues with a data-driven approaches and "[draws] on critical social theory to frame how..research is conducted, how sense is made of the findings, and the knowledge employed" (2014:9). The aim is to arrive at a mode of working able to recognize that the research itself does not come from nowhere: there is instead an 'inherent politics' involved at each step.

> ..the researcher is acknowledged to possess a certain positionality (with respect to their knowledge, experience, beliefs, aspirations, etc.), that the research is situated (within disciplinary debates, the funding landscape, wider societal politics, etc.), the data are reflective of the technique used to generate them and hold certain characteristics (relating to sampling and ontological frames, data cleanliness, completeness, consistency, veracity and fidelity), and the methods of analysis utilized produce particular effects with respect to the results produced and interpretations made (2014:9)

To Kitchin's observations I would also add some additional, reflexive thoughts specific to my own work at this point. The first relates to the limitations of my own technical capabilities in terms of coding, computational analysis and other areas of knowledge not usually associated with popular music studies. These can be considered in terms of Boyd and Crawford's (2012) divide between data "haves" and "have nots", where possession is defined not only by access to data but also the ability to store, process and analyse that data. My research is exploratory, experimental and developing, but it is limited in technical scope in a manner that highlights the second part of Boyd and Crawford's point. Without a background in statistics, computational science, or related disciplines, I can only take such an

approach so far. This presents a different set of implications. To the problem of a creative or experimental approach, then, we must also add technically limited. An interesting, additional tension arises too if we consider that these technical limitations would be obvious to anyone from a scholarly or commercial background in a relevant field, but they may not necessarily be so to someone from a popular music studies or cultural studies background, where work of this kind is less common (or to audiences for more general public dissemination activities). Indeed, practice-led work focusing on data and algorithms in cultural studies is as relatively new as the technologies and questions it concerns itself with, and so by deliberately revealing and reflecting upon the difficulties and limitations I face and I am able to make a second observation that is specific to popular music studies. The approaches described by Kitchin above, and particularly positivist and empiricist approaches obfuscated through technical complexity and/or commercial opacity, are precisely those that are being deployed by the business interests who are occupying powerful positions within the chains of production, distribution and consumption, as Chapters 1 and 2 have demonstrated. Similar technologies and practices also impact across social media, where a considerable amount of cultural discussion and commercial promotion of popular music takes place. In other words, and recalling both the analysis of Prey's work in the previous chapter, and van Dijck's notion from earlier in this one of the digital transformation of sociality, they can be said to be having a cultural as well as a commercial impact. A study of contemporary popular music cultures may benefit, then, from taking data-intensive epistemologies into account and as such my intention is to develop a field-specific understanding of them. My work – even with the caveats of being technically limited, creative/exploratory – by aiming to be simultaneously reflexive in terms of the issues inherent with it, contributes important first steps in that direction for popular music studies. Indeed, following Kitchin, it may well be that it is precisely the limitations and problematic elements, revealed and reflected upon by my process, that help frame the questions regarding what we as popular music scholars do next. In this way, it is the research process, rather than the data, that is 'allowed to speak for itself'.

There are hints in emerging popular music studies work as to this combination of a focus on data systems with critical social theory that was suggested by Kitchin. Prey's framing of streaming services in Lefebvrian terms, for example, explores changing ideas of the individual and classification, while Webster et al look towards Bourdieu and Latour in order to understand the same. This work, as I have argued previously in chapter 2, focuses primarily on the activity of socio-technical systems as they relate to digital music interfaces. It does not explicitly attempt to account for the manner in which these fit into the wider trajectory of "fractured and heterogeneous" (Nowak, 2014) cultural practices associated with popular music that in the main take place "in the non-digital parts of the human world" (Sterne, ibid). The extent to which, for example, Spotify listening influences or is influenced by vinyl, radio and CD listening, is not considered in work that focuses on the internal systems of data-derived business models, since this is data that is often not captured during the process of delivering recommendations.

This brings me to another important point central to my enquiry, that of the role of digital technologies on wider conceptions of knowledge and experience, and to the question of what happens when the real and virtual worlds meet (we can recall here Bucher's concept of the "algorithmic imaginary" from chapter 2), and how that may be addressed using a mode of working informed by a digital humanities approach that draws on "critical social theory to frame how the research is conducted" (Kitchen, 2014:9). For this the work of Michel De Certeau, discussed in detail in Chapter 2 has been beneficial. Although, as Dant (Dant, 2003) points out, De Certeau was not strictly speaking a critical theorist, he nevertheless shared a critical interest in consumer society with the Frankfurt school but, unlike the critical theorists with whom his work can be placed in continuity (Gardiner, 2000), he was focused less on instrumental reason and more concerned with the manner in which "individuals manage[d] to maintain some control of their everyday lives despite the impact of mass culture" (Dant, 2003, p. 75). De

Certeau's conception of place/space, strategies/tactics and the ongoing, symbiotic negotiation involved provides a useful theoretical framework that informs the manner in which I examine how digital technologies and the everyday meet. Practical questions, however, remain.

3.3 – Framing the solution

To return briefly again to van Dijck, the issue remains as to how we may develop a way of understanding the processes that facilitate this ongoing negotiation between space and place, strategies and tactics, and the networked knowledge generated through the digital conversion of individual actions into data points that occur alongside this negotiation. De Certeau, once again, provides a useful means by which to think through such a task:

> even if it is drawn into the oceanic rumble of the ordinary, the task consists not in substituting a representation of the ordinary or covering it up with mere words, but in showing how it introduces itself into our techniques…and how it can reorganise the place from which discourse is produced (1984:5)

In other words, we may ask what happens if we attempt to replicate the processes that are driving our questions. What happens if we follow a research process that explores the ways in which engaging with data and computational processing may provide us with the answers to our questions? What this means is a shift in focus. Rather than looking only towards ideas of results derived from data – whether in the form of 'insights' in the commercial setting, or 'findings' in the research setting – we may instead look to what we can learn from engaging critically with the processes that help deliver those outcomes. What, in other words, can a research process that exists in what Liu calls 'polluting proximity' to its object of study tell us about the processes, issues and questions related to data-derived, networked knowledge? To understand this distinction we can look once again towards De Certeau, who argues that science constitutes the whole as its remainder,

and that the remainder is what we call culture. In other words, as far as science is concerned – and here I specifically mean the kind of data science performed by engineers at Spotify and in other entities engaged in datafication processes – culture is reduced to that which fits into the receptacles of procedure. It is precisely this conflict that we see in the practices of data-driven analytical techniques, each of which are engaged (directly or otherwise) in producing what Berry calls "destabilising amounts of knowledge" and meanings that are potentially far removed from the sum of their constituent parts, but are read nevertheless as objective truths due to the apparent robustness and validity of the procedures behind their construction. Their remainder is what we call culture.

It may follow, then, that an examination of those procedures as they relate to popular music culture is able to inform us about the growing relationship between the two. The crucial point here is that that which can be reduced to a quantifiable unit, and then measured and analysed before being deployed as a form of cultural knowledge (such as in a recommendation playlist, for instance) ultimately has an effect upon the culture it measures. As such an exploration of the conflict between the artificial language of scientific explanation (the empiricist position observed by Kitchin) and the natural language of the social world (in other words, the cultural, the tactical/spatial) is a potentially useful route. It is through such processes that De Certeau's expert "blots out" the philosopher, formerly the specialist of the universal (1984:7) – which echoes with Berry's (2011) observation that data-intensive systems produce "destabilising amounts of knowledge and information that lack the regulating force of philosophy". Raymond Williams (2011) defined culture as the expression of the ordinary by human societies through their activity. That we may perhaps be moving beyond this, to a point where culture (in part at least) is generated by data-derived knowledge, expands such a definition even when the production of the systems capable of such 'intelligence' fall under Williams' definition: we can recall again here Danietz's observations, or Ananny's conception of NIAs, and observe that systems of this kind are still designed and operated by humans (as indeed are research

processes). That human ingenuity may no longer be the sole driving factor in the creation of culture (or knowledge, in the case of research) is a prospect worthy of consideration. Given that, it seems reasonable to suggest an understanding of the techniques, ideologies and motivations behind such a shift would be a useful thing to consider. Such a consideration, however, recalls and builds upon the issues discussed earlier in this chapter about the difficulty of undertaking research of this kind, particularly in the humanities.

Although processes of data-derived knowledge are largely scientific in nature (i.e. mathematic, computational) it is not always the case that they are transparent, or available to scrutiny. The notes of the laboratory procedures, so to speak, are more often than not never entered into the public record, and as such the experiments they produce are not entirely replicable. As Morris (2015) points out, the difficulty in studying (and thus potentially holding to account) such processes, is made doubly difficult by not only their private and proprietary conditions, but also by their complex and fluid nature. In other words, and as was discussed above, there are often a rapidly moving target. Lin and Ryaboy (2013), in a paper describing their working practices at Twitter, where both are employed in technical design capacities, describe for instance the importance of plumbing in their day-to-day operations and system design. Plumbing (which we can observe is yet another natural metaphor) can be understood as the recognition that new analytical and procedural solutions must be built around existing and legacy processes, and often the combination of 'old' and 'new' systems produces unforeseen consequences (and benefits). Local knowledge such as this is fundamental, but local knowledge is often subject to rapid change: the volatility and hugely competitive nature of the music streaming space is a case in point.

Further to this, many of the steps in such processes, or the tools and instruments required, and indeed much of the raw material used (i.e. the data), are more often than not obfuscated behind a combination of commercial, technical and legal firewalls. Once again, these barriers to entry have been conceptualised by Boyd and Crawford (2012) more broadly as a

widening "big data divide", where the lines of power and influence are drawn between the data "haves" and the data "have nots". To challenge the outcomes of such processes is, in effect, to run the risk of relying on so much guesswork about the motivations and techniques involved (although Prey and Webster et al's work is useful here), or else to attempt a process of reverse engineering when presented with the real-world results. Here we can consider De Certeau's argument that the expert ultimately trades his expertise for authority, before eventually realising that his discourse has been "no more than the ordinary language of tactical games between economic powers and symbolic authorities" (1984:8), or, to put this another way, the result of the analytical process deployed across the aggregated mass of abstracted everyday experiences may not necessarily be utopian. However, if we are to believe the pronouncements of the experts and proponents in the field of data science, which van Dijck argues are redolent of a belief system she calls "Dataism", the utopian is at least at this stage a residual possibility. A critical position invites us to ask what the risks are instead of outcomes where expertise is traded for authority and, ultimately ideas of diversity meet with economically driven systems in a manner suggested by Straw's work discussed in Chapter 2.

What this means is considering the possibility and raising questions about whether it is as much a closing off of (cultural) possibilities, rather than utopian or even syntopian scenarios, that will ultimately be the result of digital technologies, rather than more potentially liberating and democratic outcomes. Such a position further underlines the need for an enquiry into the relationship between digital technologies and everyday experiences, regardless of (or perhaps precisely because of) how difficult such a task may be. The processes and motivations that reside within the mechanisms of data collection, storage, categorisation and analysis, are thus worthy of our attention, and the optimistic position presents itself of whether a practical, critical engagement with the processes involved could be a useful one. Could, in other words, a critical, reflexive use of data collection procedures and computational analytical techniques produce either a different type of

knowledge (about popular music, about the process of research) to that which may be produced by and through either commercial processes and/or existing modes of academic enquiry? Could a practice-led process allow us to peer past the obfuscation and opacity inherent in many of socio-technical assemblages, and could this help us examine more closely what Morris (2015) points out as the "recursive loop" between interaction with and recommendation of cultural goods under the gaze of datafication? Could it also enable us to consider in new ways the on-going tactical/strategic negotiation between space and place that occurs when everyday life and digital monitoring technologies meet? These are the inter-related issues of debate and method that drive my own enquiry, and my reading of the extant and emerging literature leads me to speculate that the most productive route towards addressing the many issues raised in these chapters so far, is a practice-led study that looks at the relationship between cultural practices and how they are digitally monitored and computationally analysed.

3.4 – A new way of working

> Computation forces us to rethink our current disciplinary practices in the humanities from the ground up. What counts as evidence? What is the relationship between theory and practice? How do we account for the technological mediations of our critique? But culture too impinges upon computation. It challenges the universalism and the neutrality implicit in many computational applications. It reminds us that knowledge is always situated, somewhere, at some time, by someone. Putting culture into computation cautions us to remember where we are when we think we know something. (Piper, 2016)[12]

"At what point did it became necessary, in the sense of unavoidable, to use computation to study culture", asks Andrew Piper (2016), who also points out that there are some who consider it "profoundly unnecessary. It

[12] Piper's article is published online by the journal of Cultural Analytics. As such, no page numbers are listed next to his quotes. The full article is available here: http://culturalanalytics.org/2016/05/there-will-be-numbers/

consumes resources, it is politically and technologically expedient, i.e. it fails to *resist*." (ibid) But we will get to that in due course.

Piper here is describing the emerging field and discipline of cultural analytics, which can be broadly understood as the ongoing dialogue between fields of study related to culture and computation. The question for Piper as the discipline evolves, is one of 'how computation will change the study of culture, and in turn how culture will change the study of computation.' What does this mean for my attempt to situate my work in a position where fields such cultural analytics (and digital humanities and social computing) intersect with and perhaps help form a new variant of popular music studies? Piper identifies four key developmental issues with cultural analytics that are worthy of further exploration here in terms of framing my own attempt, and these will be considered alongside Kitchin's observations of the epistemological positions underpinning empiricist and data-intensive approaches described above.

Beginning with Auerach's Mimesis (2013), Piper argues that all cultural criticism views examples as representative of the whole, and that this is flawed because cultural critics are "exiled from an understanding of the representativeness of their own evidence." The cultural analytics practitioner, on the other hand, is self-conscious about their position and their role in the creation of knowledge. A research project (of the kind I am engaged in, and perhaps of any kind) is thus acknowledged as a construction and a creation, and so it follows that any data collected and the means by which it is collected, has to be considered in these terms. It also follows that any argument for the demonstration of an objective truth is difficult from such a position. Data and any knowledge derived from it are constructions rather than objective facts. Thus, for the cultural analytic, the critical object becomes the process through which the construction occurs. It is as much a study of self-reflexivity and process as it is a study of the representativeness of any data collected. This speaks to my earlier point regarding De Certeau's expert, and of culture being that which fits into the receptacles of proper

scientific procedure, from which I suggested that the useful 'results' of such an approach may be the knowledge gained from critically engaging with the process, rather than the results of the process itself.

Piper then moves on to argue that a cultural analytic approach can be viewed as more democratic than traditional 'elite' knowledge generation, but that, equally, computational processes are not a short cut to democratic knowledge:

> Computation in this sense isn't fast – it slows us down and forces us to be more self-reflective. Cultural analytics makes the study of culture more architectonic rather than agonistic, more social and collective. There is a basic politics to this practice that has largely been overlooked. (2016)

There is resonance here with my project: the crowd-sourced nature of the data gathered contains precisely the social and collective qualities that Piper describes. In addition to this, the dissemination of practical elements of my work (discussed in more detail in Chapter 4), including the provision of 'raw' data and the code involved with computationally processing it, speaks further to this. Work of this kind is intended to be replicable and verifiable, but given the conditions of its creation (and it is a creation), it can and should not make claims to elevated, 'elite' status. Piper's position has resonance here, mainly because it is true in my own research experience that there is a certain amount of democratic knowledge dissemination within the wider culture of data science, but his claim to democracy (and indeed the opposite status of 'elite' conferred on 'traditional' knowledge) elides the fact that, to recall Kitchin above, research does not come from nowhere. Cultural analytics, however, is a relatively new field and it must therefore stake and defend its claim to uniqueness and legitimacy, which may explain Piper's stance. The comparatively more established field of digital humanities takes a more conciliatory stance towards 'traditional' knowledge, although by no means has an easy relationship with it (this is discussed in more detail later in the chapter).

Related to the above, Piper states that there is "epistemological humility" running through cultural analytics: it knows what it knows, and that which it does not know, and this is because it cannot take the part for the whole. For Piper, the conception of methods as models is key to this move, in that models are fluid and can be adapted, and as such they relate to Lynch's argument that networked knowledge has a more fluid constitution than knowledge measured in stopping points. As Piper argues, 'the more we think about our methods as models the further we will move from the confident claims of empiricism to the contingent ones of representation'. This, in terms of my work, is a key point and consideration.

Projects of this kind are not and cannot be methodologically robust in the 'traditional' sense, because this would sacrifice the fluidity required of modular methods. Research projects in cultural analytics are thus always necessarily works in progress but – as Piper's previous point argues, and my caveats notwithstanding – potentially more democratic as a consequence since woven into their very fabric is the possibility that any element (small or large, from data collection to knowledge creation) can be shared, changed, adapted, or repurposed. In other words, there is a greater potential to adapt to the fluidity of their objects of study, which – as we have seen – are given to volatility, rapid change, and so on. In this way, research methods become modular in a manner that calls to mind the vehicle Johnny Cash built "one piece at a time"[13] – open access data, open source code and other building blocks are made available[14] to ensure that building upon (or overhauling) work is a more accessible, democratic process. It is thus liberating in the sense that any knowledge produced (and it is produced) can be overhauled, debunked, repurposed, celebrated, and so on, because it is as much the process and the raw materials used within that process that represent the contribution to knowledge (in the PhD sense) as it is the findings that may be

[13] This 1976 release by the country singer Johnny Cash told the story of a car factory worker who assembled his own car from parts stolen over time from the factory. The song can be heard here: https://www.youtube.com/watch?v=rWHniL8MyMM

[14] I mean freely available here in the sense that they are not placed behind the 'firewalls' of academic journals or other prohibitions to access, but this does not mean access to such elements are entirely 'free'. We should consider here Boyd and Crawford's observations regarding data haves and have nots; it may well be that a lack of technical knowledge and/or analytical tools, or data storage, act as more significant barriers that access to data and code.

derived from the data collected. All of that being said, those familiar with the work of Johnny Cash will recall that his 'personalised' car was built from assembly-lined parts[15]; likewise much of the data collection and analysis in cultural analytic scholarship is built from parts made up from commercial APIs and off-the-shelf algorithms, and this requires an additional level of reflexivity, critique and objectivity not mentioned by Piper. I address some of these questions in chapters 4 and 5. Nevertheless, and regardless of the origins of the constituent parts, Piper demonstrates that it is (much like Johnny Cash's car) in the configuration of them and the creativity employed within the research process that uniqueness – or contributions to knowledge – may be found. This is the site of the creativity in the research process alluded to earlier, but the creative limitations are worth bearing in mind.

Piper's fourth and final point, which I will touch on briefly and return to in my concluding chapter, relates to the realm of institutional research systems and funding mechanisms that many academic projects (mine included) are firmly embedded within. Here he considers how making cultural analytic work accessible, verifiable, democratic, can be thought of in terms of impact, the catch-all term for the various metrics by which academic success is presently measured. For Piper, impact is not 'a dirty word' in the sense that the necessary proximity to and immersion within commercial technologies (and institutional funding mechanisms) is a route towards positive change. This is a point I will pick up on in discussing Alan Liu's work below. Piper concludes that, "cultural analytics thus encompasses the variety of human (and non-human) cultures and media that have gotten us to where we are today. What unites it is a belief that computation can show us things about culture that previous media and their metonymical impasses could not. Likewise, culture is able to 'speak to' computation in the manner that academic knowledge, particularly in the social sciences, has traditionally attempted to speak to power."

[15] It should also be remembered that Johnny Cash did in fact steal the parts he used in the creation of his research project. This should serve as a reminder regarding our responsibilities to research ethics, particularly when data and other components can be – so to speak – liberated from the factory. I discuss my own position on research ethics in Chapter 4

Piper here provides a useful introduction to the epistemological underpinnings of cultural analytical work in much the same what that Kitchin's examination of epistemological positions that inform approaches in the digital humanities did, and indeed at first glance there appears to be large similarities between the two fields. But there are crucial differences, and a closer engagement of the differences between them, and of the related field of social computing, is required at this stage.

The term cultural analytics was developed by Lev Manovich in 2005 (Manovich, 2007), who has more recently summarised it as "the analysis of massive cultural datasets and flows using computational and visualization techniques"(2016)[16]. He offers further definition by pointing out that, since 2005, a great deal of work has occurred in terms of examining UGC (user generated content) posted to social media platforms, explorations of the professional practices that take place in fields related to digital media and culture (e.g. web design), and computational examinations of historical periods. He argues, however, that this work has taken place within the fields of social computing and digital humanities. Both, he argues, are defined by the data they analyse. Digital humanities applies computational techniques to historical texts and is an updated version of existing modes of social science; social computing meanwhile examines content at scale produced by humans online post-2004 and is primarily a discipline of computer sciences and not of the humanities – it has a focus on the technical, commercial and computational ahead of the cultural. Although digital humanities, cultural analytics and social computing all use computational techniques to gather and analyse cultural data, social computing has a computational ontology and the epistemology of the computational as its goal: it strives for the efficient collection, processing and analysis of data, regardless of the data. The deployment of knowledge it produces, however, has cultural consequences, as we saw in chapter 2. These manifest in the interfaces of services and

[16] Manovich's quotes here are taken from his cultural analytics website and so will not have page numbers: http://lab.softwarestudies.com/p/cultural-analytics.html

platforms that are increasingly central to everyday life. It is considered here precisely because of its proximity to the commercial and technical practices that impact on culture.

The advantage of the large cultural datasets available to all three fields, according to Manovich, is that the larger the dataset the more likely/possible it is to construct a representative sample. He cites the Internet Movie Database's (IMDB) repository of data on 3.4 million TV shows and films as an example here. Related to this, from a representative data set it is possible to discover trends and patterns in much the same way that empiricist and data-intensive science epistemologies do. The issue with representative samples, however, is they cannot reveal what Manovich describes as "content islands". These are "types of coherent content with particular semantic and/or aesthetic characteristics shared in moderate numbers" (2016). We can consider and contrast here once again the issues discussed in chapter 2 regarding the removal of marginal outliers and the drive for a statistical consensus that Kitchin, Vanderbilt, Van Dicjk and others observe in empiricist epistemologies, and we can observe that there is a very important difference in approach suggested when the various "content islands" or "outliers" within a dataset are foregrounded and examined, rather than statistically smoothed over. This foregrounding and examination of outliers is more closely associated with cultural analytics than with digital humanities or social computing.

This leads Manovich to the question, then, of whether we can study everything: a representative sample drawn from a massive dataset still takes De Certeau's part for the whole, and still fits culture into the receptacles of proper procedure, a key part of the problematic with systems of this kind. For Manovich, however, representative samples based on hitherto unprecedented amounts of data provide us for the first time with an "opportunity to describe, model, and simulate the global cultural universe, while questioning and rethinking basic concepts and tools of humanities that were developed to analyze 'small cultural data'". This is an approach that

Manovich claims is more akin to the linguist or the biologist – analysts who take every element possible to construct their studies.

Once again, we can observe here that the object of study shifts from the dataset to the research process and, recalling the problem raised by De Certeau, we can see that Manovich tempers his apparent advocacy by acknowledging that there are indeed limitations to his approach. He acknowledges that not everyone in the world is online, that the sources of content and data (Tweets, scrobbles, etc) have what Gerlitz and Hellmond (2013: 14) call the "medium specific infrastructure[s]" that shape, restrict and to an extent direct human actions within them, and that the commercial APIs making such study possible do not provide "everything by everybody". There are, in other words, numerous limitations (which are further reductions and abstractions to those discussed previously), many of which have been raised throughout this and the previous chapter.(see: Boyd and Crawford, 2012; Piper, 2016; Tufekci, 2015; Van Dijck, 2014b; van Dijck, 2013)

Despite these important limitations, however, the cultural analytic position is that the availability of a hitherto unprecedented amount and variety of data nevertheless presents an opportunity, as Manovich states, to rethink what he describes as the "inevitable" move towards a sociology of culture that large datasets move us in the direction of. It seems natural, he argues, to take large cultural data sets and to begin organising them according to "social, cultural and economic indicators". Whilst this is of interest to the cultural analytic approach, there is also the opportunity to seek out the individual and the particular (the "content islands"), as well as the general and the regular. The practical key to this, beyond the computational techniques and the availability of large data sets, is a developmental difference in the anatomy of the data sets available.

For Manovich the traditional "long" data set has given way to the "wide". From datasets with many cases and few variables – such as in Kitchin's

example of Census data – we now have to take into consideration (as our research objects, as our research tools) datasets with many more cases and many more variables, and which are dynamic in the sense that more data and variables can be added, in real-time, all of the time. We can recall from Lynch that the Internet does not have stopping points, and also Cheney-Lippold's fluid, measureable types. What Manovich means here is that long data with large sample sizes and a limited number of variables (e.g. age, income and gender), exemplified by census data, previously enabled the creation of clusters and patterns but were ultimately limited in scope. Wide (fluid, dynamic) data sets have a ratio between cases and variables that is more evenly matched, and potentially even reversed. This means not only the possibility of more, and more nuanced clusters and affinities (such as in Spotify's taste profiles), but also the opportunity to question many of the taken-for-granted dimensions of categorisation, such as age and gender, exactly as Webster et al (2016) suggest. Borrowing a term from Viktor Shklovsky, Manovich conceptualises this as estrangement: "making strange our basic cultural concepts and ways of organising and understanding cultural datasets. Using data and techniques for manipulating it to question how we think, see, and ultimately act on our knowledge."

The tantalizing prospect for Manovich is the development of a science of culture, one that learns from the developments in scientific discovery through deterministic then probabilistic paradigms, and eventually arrives at computationally aided model simulation. In other words, it is a question of what the possibilities may be if we harness Hey's 4th paradigm in undertaking critical social research. The aim here, according to Manovich, is not to eradicate free will (we can recall the discussion of predicative modeling and inferential statistics in Chapter 2), but instead, and following Gilbert and Troitzsch (2005), to reach a point where "thinking of how cultural processes can be simulated can help us to develop more explicit and detailed theories than we use normally", and it is perhaps in this way that we may develop Nissenbaum's (2009) contextual norms for the use of data whose uses have been "radically, unpredictably decontextualised", or perhaps consider the

ways in which the expert may not necessarily blot out the philosopher, or – following Piper's claim that culture and computation may reciprocally benefit – how the expert may trade his expertise for something other than authority and instead lead towards outcomes that open up rather than closes off cultural possibilities.

3.5 – Critique and response

In much the same way that Piper points out that some argue cultural analytics "fails to resist", Kitchin observes that digital humanities also has its detractors who question the efficacy of using machines in the derivation of ideas of meaning when addressing cultural texts and phenomena. Summing up and paraphrasing the various critiques, Kitchin states that it is ultimately seen negatively because it is a mode of academic enquiry that:

> foster[s] weak, surface analysis, rather than deep, penetrating insight. It is overly reductionist and crude in its techniques, sacrificing complexity, specificity, context, depth and critique for scale, breadth, automation, descriptive patterns and the impression that interpretation does not require deep contextual knowledge. (2014:8)

Likewise, a data-intensive approach to social science receives similar criticism for being "mechanistic, atomizing, and parochial, reducing diverse individuals and complex, multidimensional social structures to mere data points[17]", which are then used to explain complex, human phenomena according to correlation, patterns and other insights revealed by data. "Reducing this complexity", observes Kitchin, "to the abstract subjects that populate universal models does symbolic violence to how we create knowledge" (2014:8). Meanwhile, the cultural analytics approach suggested by Manovich has been criticized by Caplan (2016)[18] for being exemplary of "a conflation of map with territory".

[17] Kitchin is quoting an article in press at the time of his writing: Wyly E (in press) Automated (post)positivism. UrbanGeography.
[18] Caplan's article was published online and so subsequent quotes will not have page numbers. Full article is available here: http://www.e-flux.com/journal/72/60492/method-without-methodology-data-and-the-digital-humanities/

Caplan builds her critique from a 1946 short story by Jorge Luis Borges, "On the exactitude of science" (Borges, 2002), in which a fictional imperial project to create a territorial map of the exact size and scale of the territory itself is ultimately revealed to be futile; "a warning against confusing a thing with its representation" (Caplan, 2016). For Caplan there is a contradiction at the heart of 'big data' that is found in its simultaneous division of publics into ever smaller bits of individual information and then its endless aggregation of them into larger, more complex undifferentiated masses in a manner that "instantiates Borges's oscillation between map and territory as a permanent feature of society" (Caplan, 2016). We can recall here both Prey's observation, derived from Williams, that there are no individuals, only ways of seeing people as individuals, and Cheney-Lippold's idea of measureable types. Caplan argues that what she perceives as the lack of a methodology in Manovich's flagship 'SelfieCity' project[19] exemplifies a key problem with work of this kind through the way it "tacitly constructs both analytical value and an image of the social that demands further analysis". SelfieCity and the related Manovich project, 'Phototrails', for Caplan "shuffle between offering concrete findings and taking a more exploratory approach that refuses to nail down conclusions in favor of trying out different data visualizations to no foreseeable end". It is in this way that such work, according to Caplan, "conflates the map with the data, and the data for the territory" (2016). Driving this confusion, she argues, is a longing for an understanding of the social, but in the process there is the avoidance of a crucial category of knowledge: the social fact. For Durkheim, Caplan argues that the social fact "consist[s] of manners of acting, thinking and feeling external to the individual, [and] which are invested with a coercive power by virtue of which they exercise control over him". Statistical sociology revealed to Durkheim social facts that in turn revealed "the significance of more social facts than he could see", whereas SelfieCity (and thus digital humanities/cultural analytics) employ statistics to "supplant the social fact rather than assert it",

[19] Caplan describes SelfieCity as exemplorary of the digital humanities, whereas Manovich claims to be a cultural analytic. This highlights the degree of confusion and overlap between the two disciplines.

thus disfiguring the social within the aggregated facts. Once again, we can consider Van Dijck's warnings, and the work of Prey and Kitchin from Chapter 2 and earlier in this chapter, about the various difficulties and issues with work that seeks to utilise the very technologies it attempts to critique.

Further to Caplan's critique, and recalling Piper's position that impact is not a dirty word, we may also consider Allington, Brouillette and Golumbia's (2016)[20] more wide-ranging and institutionally specific critique of digital humanities, one that could equally be applied to cultural analytics. The authors make parallels between the development (and continually heavy funding for) digital humanities projects and neoliberal ideologies:

> For enthusiasts, digital humanities is "open" and "collaborative" and committed to making the "traditional" humanities rethink "outdated" tendencies: all Silicon Valley buzzwords that cast other approaches in a remarkably negative light, much as does the venture capitalist's contempt for "incumbents." (2016)

Ultimately they argue that digital humanities stands in opposition to the idea that academic work should be critical. It fails, in other words, to resist. Alan Liu (2016)[21], however, argues that digital humanities can 'contribute to ameliorating the very idea of "development"–technological, socioeconomic, and cultural–today'. In doing so he sheds further light on critiques of digital humanities[22] by arguing that these can be understood in the reductive terms of 'hack v yack'[23]. This is a question of whether researchers should be engaged primarily in making, coding and other 'hacking' activities, or instead focussed on a more critical interpretation and theorisation of digital media. We return, once again, to van Dijck's observation from the beginning of the chapter regarding the need to think about the type of critical questions we

[20] As with Caplan's piece, Allington et al's critique was published online. In this case by the Los Angeles Review of Books. Full article available here: https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/

[21] Liu's quotations here are taken from a book in progress, which he has posted on his website: http://liu.english.ucsb.edu/drafts-for-against-the-cultural-singularity/

[22] These critques extend to cultural analytics as well as to digital humanities. From the outside, both appear one and the same. It is only practitioners of either that insist on the difference.

[23] A good overview of the short history of the Hack v Yack debate is provided by Bethany Nowviskie here: http://dhdebates.gc.cuny.edu/debates/text/58

need to be asking. Liu's work is useful here in terms of that, particularly in light of the various critiques of a computational approach discussed above.

Indeed, Liu contends that the debate is (or, at least, should be) much more nuanced: 'hack' aligns with more concrete, pragmatic, post-structuralist theorisation (including actor-network theory, assemblage theory), whilst 'yack' can be understood in terms of the critical traditions of the Frankfurt School, Foucauldian archaeology, Gender and Race theory. The question is not about the absence of theory, but rather, exactly as van Dijck suggests, the type of critique that is appropriate.

> What critique, in other words, not only allows the field to assist mainstream humanities critique but could not be conducted except through digital humanities methods that use technology self-reflexively as part of the very condition, and not just facility, of critically knowing and acting on culture today? (Liu, 2016)

Liu suggests the answer lies in the development of a mode of critical infrastructure studies that is possible and useful since infrastructures now have much the "same scale, complexity, and general cultural impact as the idea of "culture" itself". The centrality of streaming services, social media networks, Google, and other DDBMs in many aspects of the cultural everyday would support this claim. In this way, the data collected that makes such digital humanities/cultural analytic studies possible are understood to have been created within, and gathered from, infrastructural bodies that "at once [enable] the fulfillment of human experience and [enforce] constraints on that experience" (2016), in a manner that speaks further to De Certeau's ideas of the negotiations over control of place, space, tactics and strategies, and to Lefebvre's conflict between abstract and social space. As we have seen, digital interfaces for experiencing popular music are organised along similar conceptual lines. For Liu, "it may be that..the experience of infrastructure at institutional scales..*is* operationally the experience of "culture."" As such the word infrastructure, "can now give us the same kind of general purchase on social complexity that Stuart Hall, Raymond Williams, and others sought when they reached for their all-purpose word, "culture.""

In other words, understanding the conditions and processes (which, for Liu, are infrastructures) through which culture is experienced, produced, monitored, quantified, analysed, and so on is (and I use this word deliberately) critical. Liu offers two potential approaches to this, the first of which speaks directly to my own work. It is what he refers to as a "lightly-antifoundationalist" approach. The logical process of the anti-foundationalist critique on which this is built occurs through three steps: Anti-foundationalism begins from an acknowledgment that the groundwork (infrastructure) for anything that is important to people is fundamentally ungrounded, and thus it relies on the support of other factors (discursive, cultural, institutional) to function (again, we can consider this in terms of De Certeau's place/space, strategies and tactics). It follows that any critique must question its position within those functions that prop up ungrounded groundworks, i.e. its place from within infrastructures, or – as van Dijck, Kitchin and others suggest – the consideration of the issues at hand when we harness the same technologies we are attempting to critique. From there, the anti-foundationalist critique can then "seek to turn…complicity to advantage", in the sense of Foucault's specific intellectuals, by operating within a system with the hope and intention of steering such systems towards social good. We can consider here Prey and Webster et al's attempts to understand the role of recommendation algorithms and digital interfaces within streaming services, or Piper's vision for computation and culture to enter into mutually beneficial collaboration, in these terms. Liu then moves to suggest that the digital humanities 'hacks' only the 3rd element from this logical process. He cites Smithies' postfoundationalism (Smithies, 2014) as an example of such an approach, which he calls "an intellectual position that balances a distrust of grand narrative with an acceptance that methods honed over centuries and supported by independently verified evidence can lead, if not to truth itself, then closer to it than we were before"(2016). In other words, the route to understanding music reception that is in turn to varying degrees governed, directed, shaped within strategic operations of digital infrastructures/places, is to be found through "producing" that

understanding through similar techniques and processes to those that used by infrastructures. The digital humanities/cultural analytic approach of continually taking incremental methodological, technical, theoretical and reflexive steps, rather than producing a "fixed" (Lynch) or even "elite" (Piper) pieces of knowledge, is well suited to such a project. This is particularly so when we consider the specific infrastructures concerned – in my case, Spotify, Twitter, Discogs, Google, Facebook, and so on – which are themselves fluid, incrementally developed, and continually honed and worked upon (see Amatriain, 2013; Lin and Ryaboy, 2013). Examining a complex, continually moving target, in other words, requires a philosophical and epistemological commitment to a complex and modular method that is able to attempt to keep pace with its object of study.

Ultimately, the potential of the lightly-antifoundationalist approach for Liu, and indeed for my own work, "arises from [a] polluting proximity to, and sometimes even partnership with, their objects of critique". This "polluting proximity" is thus key to my approach, but we should be aware that this has the potential to operate in both directions: either in the hopeful sense of helping to influence the practice of infrastructures and of participants, and in the potentially dangerous sense of producing a critique that is unable to resist. This is why the self-reflexivity often cited as a central tenet of both cultural analytics and digital humanities research are so crucial, and this will be in evidence throughout the proceeding chapters.

3.6 – The Harkive Project

The various issues, critical positions and disciplines discussed in this chapter are intended as means by which I can consider and frame my methodological choices in terms of the "philosophical stance, or worldview" (Sapsford, 2006:1) I have developed in these opening chapters. As we will see in the chapters that follow, the data I have gathered is rich and complex and the intention is to continually update it over time through both further data

collection activities and the augmentation of existing data through methods of computational analysis and data generation that will render my data as both wide (rather than long) and fluid/dynamic (rather than static). Similarly, the methods of analysis are fluid and modular and will continue to be so according to my development as a researcher, analyst, coder, and so on, and through attempts at forging modes of collaborative work in my post-doctoral phases (discussed in chapter 8). Thus, the Harkive project will never be 'finished' in much the same way that neither the digital infrastructures I study or the cultural practices of the individuals contributing their data will also never be fixed. As such Harkive represents – recalling Piper – more of a model than a method, and is one that is able to be significantly more dexterous than a fixed mode of working. The Internet, we can recall once again from Lynch, does not do stopping points. I am thus proposing a fluid, incremental mode of work, investigation and critique that is better able to move with as much dexterity (or almost as much) as the fluid infrastructural systems and individual and collective experiences with music reception that it attempts to study.

Further to this, the data I have gathered, the means through which I have gathered it, and the methods of analysis I will employ, mean that Harkive respondents can be categorised and segmented in numerous different ways – we can recall once again here the observations of Prey and Cheney-Lippold. This will be facilitated by the creation of variables derived from inferential statistical analyses of the language that respondents use (or don't use), and their explicitly and implicitly recorded preferences for both older formats (CDs, vinyl, radio) and newer modes of engagement, including streaming and automated recommendation. Respondents can also be potentially categorised and segmented in numerous other ways, each potentially deployed in a vast number of configurations. As such individuals and groups can appear in one, more, or no segments at all, depending on the parameters of analysis. The analysis will be able to reveal both the broad trends of empiricist approaches and the content islands suggested by Manovich. The observations gleaned from subsequent analyses of segments (and content islands) thus

simultaneously foreground and negate the individual in much the same way that commercial practices do. And it is here also where the creative act of methodology is foregrounded and can be considered reflexively alongside the reductions inherent in the data collection process and the apparent neutrality of platforms that facilitate such collection.

Harkive becomes, then, an experiment through which I am able to explore the issues raised in my opening chapters, and one which has its own impacts on the production of knowledge represented by my findings. It also, I should point out, enables me to answer my central research question: What can an analysis of the data generated by The Harkive Project reveal about the music reception practices of respondents?

The practice-led experiment of creating Harkive leads to an online social space where respondents reflect upon their music reception practices, whilst also behaves in a manner similar to a commercial online place because the respondents are, as Michel and Lupton describe, simultaneously "the informant, the informed and the information" (Michael and Lupton, 2015). I am thus making a deliberate distinction between cultural 'space' and commercial 'place' here in direct reference to De Certeau's model. I conceive of the activities related to music reception recorded through The Harkive Project as both tactical expressions of cultural practice and manifestations of participation in wider, strategic systems of music commerce. The tension between the two, and particularly in terms of how digital technologies and data collection relate to this, lies at the heart of what I am attempting to investigate and as such is woven into the very conception of Harkive as a research process: it is the means through which these tensions can play out within a research environment, and simultaneously the method by which I am able to investigate those tensions, all of which is contained within a modular, reflexive process that foregrounds the roles of researcher, interface, tools and analytical methods in the process of knowledge creation.

All of which is to say that data can be used to tell many different stories, and it is through exploring the manner in which these stories may be told that we can potentially find routes to critical engagement with the infrastructural systems that chapter 2 demonstrated as increasingly coming to bear upon our everyday experiences. Ultimately, we arrive at questions of epistemology: How do we 'know' through such systems and processes? How do we think through the impacts upon epistemology of systems of data-derived knowledge creation? What happens when data-related systems, technologies and practices are simultaneously the object and the method of study? These are the various issues I will attempt to address through the development of the work started in this project.

In this chapter I have surveyed a wide variety of scholarly work that concerns itself with debates about how to approach the growing role of digital technologies in everyday life. Through my engagement with this work I suggest that thinking through those technologies can provide a route towards a greater understanding of contemporary popular music cultures. Considered in epistemological terms, I suggest that a productive route forwards is for scholars to find a way of harnessing the same technologies that sit at the centre of many of the issues of debate discussed in my opening chapter, and that such work – although difficult – is nevertheless possible. These new modes of working provide a difficult but potentially useful route towards new knowledge and understanding, but only if reflexivity and modularity are built in to the fabric of our methods and approaches.

I have demonstrated that this is both a difficult but also a hugely important task, and move to suggest that a reflexive, experimental and exploratory methodology is a useful route forward. It is time, then, for popular music scholars such as myself to roll up our sleeves and to begin to forge the new methods that may help us better understand contemporary popular music cultures.

CHAPTER 4

An experimental approach

In this chapter I will discuss in detail the means through which I will tackle my central research question: what can an analysis of the data generated by The Harkive Project reveal about the music reception practices of respondents? In doing so I will also investigate the various issues of debate I have engaged with in the opening chapters, whilst also working through the numerous issues of method that my engagement has highlighted, and which have been further revealed through the development of The Harkive Project.

4.1 - Overview

Harkive is an online project running annually on a single day in July of each year that encourages people to provide detail and reflection on their experiences with music across the course of a single day. Since the project first ran in 2013, it has gathered 7582 individual stories. The shortest story contains only two words, the longest almost 4,000. Each represents a snapshot of individuals' engagement with music, and the technologies, locations and everyday situations variously involved. Some stories are inspired by memories, others detail engagement with technologies, many show how experiences with music and technology are deeply woven into the rhythms and routines of everyday life (Hesmondhalgh, 2002).

The central methodological challenge of this project has been to devise a means by which I can understand this collection of texts. By developing a method that harnesses data-derived techniques associated with the computational turn in the humanities (Berry, 2011; Hall, 2013), such as exploratory data visualisations, clustering, and natural language processing, the aim has been to develop a different kind of insight than would have been possible through manual readings alone. Using these tools and analytical methods enables me to examine the role that digital monitoring and data technologies play in contemporary modes of engagement with music. As my opening chapters showed, these now have a

central role in the distribution and consumption of music. An equally important and related aim of this route has been to examine and reflect upon the means by which these tools can help produce that understanding. I have sought to understand what we, as researchers, can learn from reflexively using these techniques, and whether an understanding of those processes help further knowledge in the field of popular music studies. The process of exploring these related issues of method and debate reveals more, and more nuanced issues, as it unfolds. The method I have developed enables me to address these issues. It has been conceived, following Kitchin (2014), as a modular method that comprises a variety of separate tasks and functions that together form a workflow. It is my suggestion that some or all of these tasks and functions, and indeed the workflow as a whole, may be of use to other researchers. As such, I envisage my method as a substantial element of my contribution to knowledge. This is a methodology comprised of many parts and as such it is useful to consider each of its components in turn, and how they combine to create an analytical workflow. We can begin by considering the data that has been gathered and how this facilitates the overall workflow.

The reflections and detail contained within the Harkive stories in the main come from posts made to social media platforms, from participants who have emailed the project directly, or from people who have completed an online form. During the 2016 instance of the project a survey was made available to participants intended to gather additional, quantitative and demographic data that would help supplement the qualitative data of the story-gathering process. Importantly, all information gathered is in a digital format, and thus reducible to data points that can be counted, analysed, and otherwise computationally processed at scale. Stripping away the unique, individual detail contained within each story, the 'raw material' I am working with can thus be understood primarily as data. For the purposes of this project, this is an important distinction.

Using a number of automated collection methods, data has been gathered into a single database containing not only the text-based stories and quantitative

survey responses, but also metadata gathered during the collection processes and additional variables generated through computational analyses. Metadata here can be understood to include information such as time/date stamps, or detail on which platforms each story was collected from. Additional variables generated via computational analysis can be understood as numeric abstractions of processed text and other variables. As such, the reflections of individuals (the texts) can be considered alongside not only the numeric abstractions produced through computational analysis, but also in terms of the processes that generated those abstractions. They can, of course, also be considered in their original format of texts that contain detail about individual experiences with popular music. The findings chapters that follow will explore these different vantage points, beginning with a reading of the abstracted data produced by machine analysis in chapter 5. This will focus on broad trends, clusters and themes within the dataset in a manner that is closely related to ideas of the computational turn, and to empiricist and data-intensive positions discussed by Kitchin (2014) in chapter 3. In chapter 6 a hybrid approach with be used, combining both distant (computational) and close (manual) readings of the texts. The process will end, in chapter 7, with a manual reading of the original texts. All three chapters thus arrive at useful answers to my central research question by approaching the same data set in different ways.

Each chapter begins with a specific, additional research question related to the arc described above. Chapter 5 asks what an unsupervised, computational analysis can reveal about Harkive respondents' engagement with different formats, technologies and services. Chapter 6 investigates what a combination of 'distant' and 'close' reading of the Harkive texts can reveal about engagement with streaming services. Finally, Chapter 7 explores how respondents describe their relationships between everyday experiences of music and digital monitoring and data technologies.

In describing the methodology that will enable me to explore these questions I combine technical detail with reflection that enables me to engage with both issues of debate and method. This begins with a discussion of developmental

work I undertook that pre-dates this research, and which provided a solid but ultimately limited foundation for my analysis. I then describe in detail the methods of data collection and analysis I have developed, and demonstrate how the techniques and methods (and issues they raise) are common in much of the work emerging from the disciplines and approaches discussed in chapter 3. Following an overview of the research ethics that have underpinned this project, I then discuss a number of 'by-products' and non-text outputs of my work, including a data and code repository and API, that form part of a suite of outcomes that on-going activity around The Harkive Project in the years to come will further contribute towards. I then close with a brief discussion that relates the specifics of this method to the wider arc of my research.

4.2 – The Harkive Project – development

The Harkive Project was initially developed as my final project during my MA Music Industries at BCU, in 2013. I was interested in the pace of technological, commercial and cultural change in popular music. Much of this I felt was linked to the widespread adoption of digital and online technologies, but I was simultaneously somewhat dismayed by the often reductive discourse around music consumption, particularly in media outlets and commercial literature (a theme subsequently explored by Edwards at al (2015)). Because of this I sought to develop a means by which the new ways in which people were engaging with music could be documented in a more nuanced manner. The original intention was that any information gathered by the project could potentially be made available to researchers who, in future years, may wish to study the period. To be clear, at this point, an analysis of the data I hoped to gather was not a primary consideration. This is an important point to make in relation to Kitchin's (2014a) work discussed in the previous chapter – the data here was collected without specific research questions in mind and can thus be understood instead primarily as a resource for question formulation through exploration.

Following early research into large-scale observational and annual music

projects[1], I settled on an approach that would seek to gather text-based reflections from people about the detail of their experiences with music across the course of a single day. Deciding on the name Harkive, a unique conflation of the words 'hark' (to listen) and 'archive' (a collection of records), and deciding on the date of 9th July 2013 as my collection period, I began the process of promoting my project via online platforms and through the dissemination of key messages via media releases. To give the project a longitudinal dimension, I positioned it in promotional material and communication as a new, annual event[2].

My communication to potential participants asked them to 'tell the story' of 'how, where and why' they listened on the day. Seeking to make involvement in the project as easy as possible for potentially interested parties, I decided that rather than use a single method of data collection, such as a survey, I would instead collect participants' contributions from a variety of 3rd party online platforms, including Twitter, Facebook and Tumblr. My rationale here was that regular users of these platforms may be more comfortable with communicating within their preferred platform, and that this could potentially lead to more useful responses. For those who perhaps would not wish to have their reflections published on openly available social media platforms, I also accepted stories via email and through a form on the project website. The project did not impose a word limit on submissions.

Having communicated in the promotional material that participants needed to include the hashtag #harkive in their social media posts, I was able to utilise several free, off-the-shelf services that facilitated the collection of data from 3rd party social networks via their Application Programming Interfaces (APIs). This enabled me to automate the collection of responses from the various places they were made online. To my immense surprise and satisfaction, the inaugural Harkive day was a huge success, gathering thousands of responses and receiving

[1] These included: the annual Rajar survey into radio listening habits, Record Store Day; the One Day in History project; and various Mass Observation projects.
[2] The 'About' page on the project website provides an example of how the project is framed to potential participants - http://harkive.org/about/

positive coverage in numerous media outlets[3]. My PhD project began in October 2014, by which time Harkive had already run on two occasions, in July 2013 and 2014. It ran again during the early phase of my doctoral research, in July 2015. Once engaged in the PhD process, however, several key limitations with the project design and my technical research skills were revealed.

4.3 –"I am not a coder" – learning R and data science

In chapter 3 I developed the argument for an analytical approach derived from the fields of cultural analytics and digital humanities. Lacking sufficient technical skills to undertake such work, however, I have devoted a considerable amount of my research energy to developing skills in order to follow that route. In particular I have gained an understanding and practical knowledge of the R programming language and its various, task-specific packages.

R is a free, open source programming language and software package that is extremely powerful and efficient in terms of processing data. It was initially developed by Ihaka and Gentleman (1996) at the University of Auckland during the 1990s as a piece of statistical software, but through the efforts of the community of developers that has coalesced around it, R can now also be used for data collection, textual analysis, interactive online visualisation, and a host of other functions. More importantly for my purposes is that the culture around R is one of collaboration and sharing. This has not only helped me learn how to use the software for my own research (essentially from scratch – I had no prior experience with coding, or programming languages) but it has also opened my eyes to the possibilities of enhancing, improving and disseminating work through sharing and collaboration. What is interesting is that this commitment to a culture of collaboration and sharing is something that is also intrinsic to emerging scholarly approaches in the humanities and social sciences that are attempting to harness, understand and critique data-related technologies (see,

[3] This included: written features in music media outlets such as NME and numerous radio features, most prominently on BBC Radio 4's 'Today' programme.

for example, the recent work of Housley et al (2014), Sandvig & Hargattai (2015)). I see my own work as part of that wider movement.

Learning the rudiments of the R programming language involved taking online courses in R programming basics, working through textbooks specific to the R language (for example, Haider, (2015)) and a general immersion and engagement in the world of data science through podcasts, conferences and other materials[4]. This initial, self-directed learning activity led me to more formal training[5], which enabled me to gain a better working knowledge of importing, organising and analysing data using the R package. This period, which represented an extremely steep learning curve, was a pivotal phase in the direction of my research. In terms of the practicalities of my methodological challenge, the first fruits of this learning period were a major and important rethink in terms of how Harkive data was collected.

4.4 – Data collection

It has been acknowledged by data analysts that a significant proportion of the time and effort involved with their projects is spent preparing and cleaning data to make it ready for analysis (Dasu and Johnson, 2003; Wickham, 2014), and my own experience with the 2013-15 Harkive data would indeed support this. Data collected from the various online platforms resided in separate spreadsheets, each organised according to different schema, and each containing both common data (e.g. the Harkive texts) and also data specific to the platform concerned. In addition, formats for common data across spreadsheets (particularly time and date stamps) were often different. The process of manually collating this data, standardising formats, and so on, was hugely time-consuming and error prone. I was aware also that my project would not be able to grow in scale if manual processes such as this were not automated. Subsequent data collection phases

[4] Some examples here: The Partially Derivate podcast (http://partiallyderivative.com) is a magazine-style show dedicated to data science. It discusses developments in the field, interviews practitioners, and also provides resources for people wishing to learn new skills. Another example of the activity engaged in here was my attendance at the 2015 London leg of the annual Big Data Week conference (http://bigdataweek.com), an industry-focused event where practitioners presented cutting-edge work and technologies.

[5] I enrolled on and completed an intensive, 4-day training course at Nottingham Trent University in the summer of 2016. Details of the course are here: https://www.ntu.ac.uk/study-and-courses/courses/find-your-course/science-technology/short-course/2017-18/intensive-r-course

have thus attempted to eradicate or else reduce the processing required concerning these issues. This has been achieved by introducing a three-step collection and pre-processing phase facilitated by the Zapier service. Following an exploration of the data available via each 3rd party APIs, I created a schema[6] that collects data into a central database organised according to the principles of tidy data (Wickham, 2014). These principles hold that in order for a dataset to be tidy (as opposed to messy) it must be arranged such that: each variable forms a column; each observation forms a row; each type of observational unit forms a table. (Wickham, 2014:5)

In order to populate this database according to those principles and with the correct data from the different 3rd party APIs, a process of matching differently named elements with their relative content was undertaken. For example, from the Twitter API the content of a tweet is held within the text element and the date/time is held within the created at element. Meanwhile, the Tumblr API holds its text content in the body element, and the date/time in the date element. Further to this, both Twitter and Tumblr return date and time information in different formats:

Twitter API date/time format: Fri Jul 09 09:51:53 +0000 2013

Tumblr API date/time format: 2013-07-09 09:51:53 GMT

The central Harkive database, then, needed to be arranged so that elements collected from the text (Twitter) and body (Tumblr) elements were situated as observations within the same variable (column), and that their sources (i.e. Twitter or Tumblr) were recorded as observations along the relevant row, but as different variables. The same process was applied to the date/time elements described above, but only once an intermediary step had been introduced that converted the different date and time formats into a common one. This rationale[7]

[6] This schema is available to view in Appendix A
[7] The process is explained in more practical detail in a blog post on the Harkive project website, along with a walk-through video and further discussion of the affordances and limitations. See: http://harkive.org/datcolzap/. This is an example of the type of 'benchwork' Sandvig & Hargatti (2015)describe, discussed in more detail later in this chapter.

was applied across all methods of data collection so that responses ultimately resided in a central, 'tidy' dataset.[8]. A by-product of this process is that the Harkive data can now be made available via its own API[9], since all data is arranged according to the same schema: e.g. in the 'story' variable there are rows containing only the original stories, regardless of their source, whilst information about what that source was, the year it was collected, the data and time, etc., are held as observations within their own variables. Despite the obvious benefits of this more logical and robust approach to data collection and management, the need for a process of data cleaning and sense checking was not eradicated entirely. By exploring some of those processes below, I reveal some of the issues with an approach such as this and how researchers may mitigate against them.

The central database necessarily contains some blank or NA observations (cells), because certain data available from the API of one service is not available, or else does not exist, in another. For example, Tumblr provides information regarding the hashtags users have attached to their posts in a separate tags element, whereas Twitter and other services do not. Consequently the variable 'tags' in the Harkive dataset contains blank observations in the rows associated with Twitter. Whilst this particular issue could be resolved by extracting any text from the 'story' variable that begins with the character '#', and then using an R command to write it to the blank observation in the corresponding tags variable, there are nevertheless some blank observations that cannot be resolved. As such, any analysis based on variables needs to consider the extent to which they are populated within the dataset as a whole.

Further to the above, the collection methods for 3rd party platforms are based on the existence of the word 'harkive' in the body of the text of posts. As such, any post to social media platforms containing that word was collected[10]. Because of the promotional activity undertaken by the project, the press coverage that it generated, and also general online conversations around the project itself – in

[8] Following this, data from 2013-15 was re-processed from the original, separate spreadsheets according to the same rationale, and then added to the central database
[9] The Harkive API can be viewed here: http://developer.harkive.com
[10] There are certain limitations to this, discussed in terms of boyd and Crawford's (2012) work in the following paragraph

other words, posts that mentioned harkive but which were not 'stories' – the collection methods captured these posts also. An additional problem was so-called Twitter bot accounts (Chu et al., 2012), which attempt to piggyback on certain trending topics (which #harkive achieved on each of the four occasions it has run so far) to promote products, services or online content. Whilst retaining this data for further and potentially different kinds of analysis at a later date is prudent, it nevertheless constitutes 'noise' in the present process of analysis. As such, a process of data cleansing was necessary to remove promotional posts originating from the Harkive project accounts, retweets and other non-story elements, and automated promotional tweets from bot accounts[11]. The bulk of this activity was relatively straightforward and achieved by, for instance, removing anything originating from Twitter that began with the characters 'RT' (signifying a Retweet of an existing post). However, the existence of a small amount of data within the corpus not removed by such automated filtering was unavoidable. Weeding out these erroneous observations was thus an iterative process undertaken as and when they were foregrounded during the analysis stage.

With the exception of data gathered via email and the form on the Harkive website, the process described above contains within it the potential for fragility because it relies on the availability of data from the owners of 3rd party platforms. As Boyd and Crawford observe, "data companies have no responsibility to make their data available, and they have total control over who gets to see them" (2012:14). The data available via APIs is thus limited and subject to change at any moment. Boyd and Crawford also observe that it is not clear, and indeed largely impossible to discover, whether the process described above captures everything, or just a sample. Tweets, for instance, from accounts where users have instigated privacy settings (known as 'protected tweets' in the language of Twitter) are excluded from searches of this kind and so will not be present in my dataset. Therefore, my data can only ever be described as a

[11] An interesting observation from the 2017 period of data collection, not covered in this analysis, was the increased number of Twitter bot accounts related to so-called 'Fake News' stories, and to US President Donald Trump in particular. This may be an interesting subject for reflection and analysis on the process of using social media data for research purposes at some later stage.

sample. This issue, however, needs to be acknowledged and highlights the criticism of an approach such as this in terms of confusing "the map for the territory" (Caplan, 2016), or indeed the inherent reductions involved with data generated through digital platforms ((Berry, 2011; Kitchin, 2014) that were discussed in Chapter 3.

In addition to the fragility of the API-derived data, there is also an issue with the manner in which the availability of data through 3rd parties directs the methods by which people can contribute their stories. The Facebook API is a case in point. This currently only allows collection of data that is posted by a user to a specific page. Anything posted to a user's own timeline is not collected. As such, the 'How To Contribute' documentation on the Harkive website asked that Facebook users post their stories to the Harkive page, and in order to do that they must first 'Like' the page. This differs from the main means by which users engage with Facebook – i.e. through their own feed, and not the pages of others. As such any Facebook user who misread, was not aware, or else ignored the request to post on the Harkive page, and instead posted to their own timeline, will not be present in the dataset, even if their post included the #harkive tag.

Despite the issues outlined above that are revealed through this reflection of the approach to data collection, it nevertheless produces a useful dataset that has – crucially – been automatically organised through the process above in a manner that made it ready for computational analyses. However, for a computational process to be able to recognise the difference between someone describing, for instance, their playing of a vinyl record, or the streaming of a song via Spotify, each description must first be reduced to a mathematical variable that can be used to denote the difference in the two descriptions. The sorting, analytical and visualisation processes facilitated by my method are – in blunt terms – culturally ignorant and it is thus the job of the analyst to inform the process. It is in this way that Berry describes the "regulating force of philosophy" (Berry, 2011) that plays a role in processes of this kind. We may also consider Dunietz's[12]

[12] http://nautil.us/blog/the-fundamental-limits-of-machine-learning?utm_content=buffer9ffc0&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer

observations regarding processes of machine learning from Chapter2. Despite these limitations, however, we can understand that if a description of playing a vinyl record can be understood as (or reduced to) being equal to *x*, and streaming a song can be understood as being equal to *y*, then both observations can be processed mathematically as distinct entities. Once other coded variables are introduced to this process, further, deeper analysis can be facilitated. For instance, if certain 'positive' adjectives[13] are contained within a new variable *a* and 'negative' adjectives are contained within another, *b*, the process of discovering the differences, similarities, etc., in the ways in which people describe their experiences with x (vinyl) or y (streaming) becomes possible. As was demonstrated via Manovich and Kitchin in Chapter 3, the addition of further variables that renders datasets as incrementally 'wide' affords the possibility of exploring greater degrees of nuance and complexity. We may recall here also the discussion of Spotify's Discover Weekly service from Chapter 2, and how that process generates a large number of variables based on both listener activity and musical content in order to produce 'taste profiles' and recommendations.

There is, however, a disconnection inherent in the outline above that can be understood more fully in terms of the relationship between qualitative data and quantitative process. Data analysis of this kind ultimately resides in the field of mathematics, statistics and computational code, whereas the stories gathered by Harkive – although rendered as data – are nevertheless representations of detailed, nuanced texts that are themselves representations of real world social actions. We have previously observed via Prey (2015), for instance, that the qualitative (for which we can read, Harkive stories) can never be fully absorbed by the quantitative (i.e. the numeric/categorical abstractions of those stories).

In other words, some additional context is required in order to help better understand the abstractions produced, and how these relate to individual respondents and their attitudes, opinions and activities. To that end, and in addition to the story gathering element of the 2016 instance of The Harkive

[13] Positive and negative adjectives are revealed through an algorithmic analysis referred to as sentiment analysis. This will be performed in Chapter 5 and explained in greater detail there.

Project, a survey was devised that invited participants to answer a series of questions about their music reception activities and their attitudes towards certain statements related to music and technology[14]. As demonstrated by Krause et al (2015) both demographic information and the level of engagement with music a listener reports is often associated with selection behaviors, and as such this was information gathered by the survey. Sloboda (1992) has suggested also that control over music listening choice produces positive effects, which is particularly relevant in the context of this project when one considers machine-derived curation and Manovich's conception of the "programmable user". Further to this, Heye and Lamount (2010) have posited the distinction between the "technology user" and the "technology consumer", where the former exercises greater choice over their music listening through self-created playlists, and the latter relies on more automated selection and/or frequently falls back on small selections of favourite songs. While much of this work is associated with the field of music psychology and explicitly explores the relationship between music listening and well-being, the observations drawn provided useful pointers for the lines of enquiry my own survey took in terms of generating additional, contextual data points. Similarly, the idea of "selection stress" suggested by Straw (1997a) was informative in terms of questions posed about automated recommendation, as was Barnes' (2006) idea of the "privacy paradox" in questions regarding data collection, privacy, and the means by which people discover and recommend music[15]. The responses to these various prompts and provocations enabled the creation of further, quantitative variables that could be considered alongside the abstractions of the original texts.

Once respondents had provided their informed consent (discussed in the ethics section below), the survey began by asking certain demographic questions (age; gender; location), and whether respondents had participated in the story-

[14] The survey was created using the JotForm service and then hosted on the main Harkive project website. Invitations to complete the survey were publicised via the project website, social media channels, and through the project mailing list. In addition to this, respondents who had submitted to the story gathering process via email or the form on the project site, who were required to supply an email address in order to do so, were contacted directly and invited to participate in the survey. Finally, those participants who had contributed more than 20 different entries via social media channels (in particular Twitter) were contacted directly via that platform's direct message function and asked to participate. The survey was made available shortly before the 2016 occurrence of Harkive and entries were closed on in late 2016. In total there were 190 responses to the Music Listening Survey.

[15] A complete list of the questions in the Harkive Music Listening Survey is available in Appendix C

gathering element of Harkive in the years 2013-16. Where respondents indicated that they had participated, they were then asked to provide details of that participation[16]. This element of the process captured usernames on services such as Twitter, Tumblr and so on, which crucially enabled the matching of responses to the survey with stories in the main database. These basic, introductory sections of the survey then gave way to more specific questions regarding the respondents' music reception activities and their attitudes towards certain statements arranged in four sections. Participants were asked to respond to Likert scale statements about their everyday music listening, their engagement with particular formats and technologies, their attitudes and opinions regarding recommendation and curation services, and finally towards data collection and privacy.

4.5 – Survey data cleaning and processing

As with the story database the survey database also required some further processing to render it 'tidy'. Because respondents were asked whether they have previously participated in the story-gathering element of the project and, if so, to provide a means by which their stories could be identified (e.g. by providing their Twitter username, email address, Tumblr/website), an explicit link between respondents to both could be established between the Survey and Story databases.

In total there were 90 questions in the survey. With the exception of 6 survey questions, all produced responses along Likert scales. For example, Likert items ranging from 'Strongly Agree' to 'Strongly Disagree' were recorded as corresponding text strings. Whilst the R package is capable of processing text as 'strings', it requires numerical values for mathematical solutions for the purposes of visualization and measurement, and for the creation of either clusters of respondents or the foregrounding of Manovich's "content islands" (2016). It was necessary to convert text-based responses to numeric values.

[16] Respondents who indicated that they had not participated in the story-gathering element were automatically directed to Section 3 of the survey. There participation in the rest of the survey would contribute data that became useful in the survey analysis stage.

Brown (2011) introduces a note of caution that should be considered at this stage regarding the confusion amongst researchers between ordinal and interval data, and the means by which results of both should be interpreted. In rendering the responses numerically, we foreground the assumption that the interval difference between 5 and 4 (Strongly Agree and Agree) is identical to the interval difference between 4 and 3 (Agree and Somewhat Agree), which itself is built on the assumption that person *a* means the same as person *b* when both select Agree. Once again, the reduction of complex social action to that which fits into De Certeau's problematic proper procedure is highlighted. As such, any claims that may arise from an analysis of this data must be considered reflexively[17]. Again, we can consider here Kitchin's modular conception of methods. We can see that should the survey element and/or its processing need to be changed, removed, or iteratively improved, this can be done without altering the overall methodology of the project, or the linear progression of the analysis. In other words, the collection of texts that occurs before the survey element, and the computational processing of stories that follows, can remain intact.

Brown's issues notwithstanding, the general purpose of rendering responses to the Harkive survey numerically is to provide a means of displaying groups of respondents according to their responses to certain questions. The ultimate aim of this is not to draw conclusions from the survey data itself – although it does reveal some interesting observations – but instead to inform and facilitate the segmentation of the story data.

4.6 – Cluster identification, text mining and exploration

The primary purpose of the survey data was to identify clusters of respondents within the story-gathering element of the data collection process. Because an explicit link has been created between the survey and story data sets, an iterative, exploratory process of visualising different combinations of survey data based on the themes of the survey questions becomes possible. This in turn

[17] Brown also suggests a number of mathematical checks and balances, such as Cronbach's Alpha test, in an attempt to ensure that the products of analysis of numerically rendered survey data are as robust as possible

enables the extraction of stories that originate from respondents in a given cluster. Once isolated, the unique themes can then be explored in further detail based on either machine analysis or a closer, manual reading of the texts. Thus we edge closer toward a means by which we can take the large and complex data set of Harkive stories and begin to extract useful information from it, pursuant to the aims of this methodology in providing a means of answering my central research question.

The means by which clustering of responses and subsequent text mining is achieved based on story and survey data is best illustrated by an example. In the example instance presented here the questions used from the survey data and the subsequent clustering they reveal are not suggested as conclusions or findings, they instead used to demonstrate the flexibility of the data exploration process facilitated by the R package and the rationale of data collection and organisation I have employed. When similar techniques are used in the formal analysis stages in chapters 5 and 6, a rationale will be provided for the selection of methods and variables in each instance. For the time being, we can select two variables from the survey data to produce clusters of respondents by way of demonstration. The key point here is that once survey responses are numerically encoded they can be visualised in isolation, or can be chained together to revealing smaller, more specific clusters in a manner that recalls Manovich's 'content islands'.

In the example shown here (*fig 1),* responses to the questions regarding the importance of cost (n17)[18] and exclusivity (n22) as everyday motivations for the use of formats and services are compared. In both cases respondents were asked to rate the importance of these factors, producing a range of responses from -3 (not important at all) to 3 (very important), with a neutral response (neither important nor unimportant) being recorded as a 0 value. Data points on the graph are then coloured according to additional data from the survey regarding

[18] Motivations for selecting formats for listening were rated from Not Important to Very Important along a 7-point scale; n17 and n22 were respectively the motivations of Cost and Exclusivity. For more information on the survey questions, see Appendix C

the age of respondents (ranging from 18-22 to 65-70), and shaped according to whether or not respondents had provided Harkive stories (Yes/No to q5). A key to these additional visual cues is provided to the right of the main graph in *Fig1*.

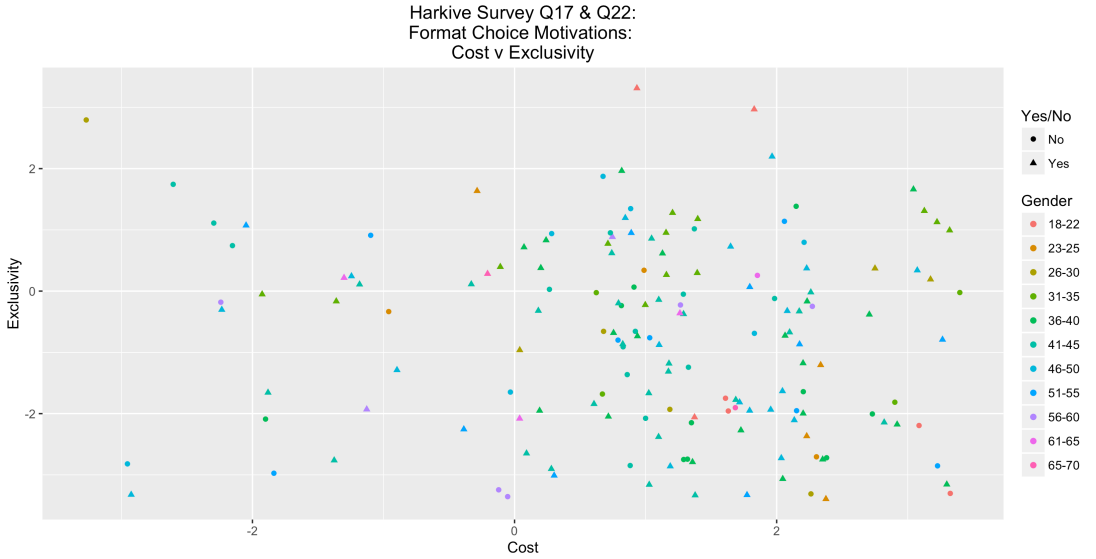

Fig 1: Cost v Exclusivity as motivations for format/service selection

The visualisation suggests that for the majority of respondents cost is an important factor, represented by the clustering of data points over on the right-hand size of the x axis (Cost). The visualisation also shows that within this majority group there is also something of a split between those who consider exclusivity to be an important factor, and those who do not – this can be seen by observing data points above and below the 0 line of the y-axis (Exclusivity). Further to this, based on the existence of data points rendered as triangles, we can observe that many of the respondents who fall within those two groups are also present in the Harkive story database. Clearly, however, the size of clusters produced by this process is not an indicator of the usefulness of story extraction. Depending on the question posed, it may well be the case that outliers and unique cases produce more interesting results. Again we can consider here Vanderbilt's (2016) observation regarding the removal of outliers and the arrival at a statistical consensus that many commercial processes similar to the above engage in, and simultaneously question this in terms of the possibilities offered

by Manovich's (2016) idea of these processes foregrounding instead 'content islands'.

The visualisation produced above could be tailored further before proceeding by introducing new variables (e.g. gender, favourite format, and so on), or by filtering out elements of certain variables already shown here (e.g. by showing only responses from people over 40 years of age). With 5 demographic data points and a further 87 Likert scale responses in the survey database, each of which contain up to 7 levels of response, it is possible therefore to explore and visualise the survey data according to a large number of possible configurations. The process thus has the capability of revealing both large-scale patterns – i.e. Vanderbilt's statistical consensus – and also the outliers, or 'content islands' Manovich describes. Once put into practice, as in the example above, it is clear to see that the data enables numerous potential routes to knowledge creation[19].

Once a subset of survey respondents of potential interest for further exploration have been identified, the next stage is to extract from the Harkive story database the responses from those who fall within the subsets of interest. In the example we are following above, this creates two subsets of Harkive respondents: those from people for whom cost is important but exclusivity is not; and those from people for whom both cost and exclusivity are important. For brevity these groups will be referred to from this point onwards as group a (cost but not exclusivity) and group b (cost and exclusivity). In the case of this example, a filter based on group a returned 574 stories from the Harkive database, and 260 from group b.

Once stories are extracted, data about them can too be visualised and explored further according to additional variables present in the stories database, thus affording the possibility of further sub-setting if required. Stories from each group can be interrogated according to the method by which they were told, or

[19] Given the amount of possible configurations of survey data, however, it would be inefficient to manually alter code in the R script for every possible permutation and combination of data at this exploratory stage. In order to make the exploration process more efficient, during my post-doctoral work survey data will be added to a dashboard that automatically creates customised survey data visualisations when the user selects survey questions and alters parameters. This tool will be made available to respondents. The rationale for this is explored in Chapter 7.

by the years in which they were collected. Based on an interpretation of the visualisations created at this stage, further filtering of stories based on particular criteria becomes possible based on other variables within the story database that will be created through computational analysis of the text (see section 4.7 below). We can thus consider here Webster et al's (2016) idea of how automated recommendation systems move beyond 'traditional' categories of taste, such as age, class or gender, and also Prey (2015)and Cheney-Lippold's (2017) work around the manner in which individuals are subject to processes of individuation through constantly shifting categorical formations. Harkive respondents can similarly be seen to have the potential to appear in one or more categories without the content of their original contribution materially changing. For the purposes of this example we will continue with the complete data sets as they were extracted, without employing additional filters at this stage.

Once we have a potentially useful configuration of stories from the database, the next step is to generate information regarding their content. At this stage it would of course be possible to undertake a manual reading of the stories, and in some cases this would be extremely useful. In the case of this example workflow, the intention is to demonstrate how the processing of datasets of this kind can be used, as Kitchin (2014) observed, to reveal possibilities for further enquiry, model building and theory creation. With several hundred stories in each group revealed by the survey analysis, a manual exploration of the associated stories would be time-consuming whilst carrying no guarantee that a useful result will be arrived at following undertaking that work. Instead, a combination of R packages related to text mining, cleaning and visualisation can be employed to enable initial explorations of the texts. Through this, if a given segment of the database reveals potential for useful, further analysis, it can proceed. If not, the modular workflow is able to retreat a few steps to generate a new configuration of respondents and stories.

Proceeding with groups A and B, in the first instance stories are isolated from the main database and collected into a corpus. Each corpus is then cleaned to remove characters that are to be excluded from analysis (for example, @

characters from Twitter names, or : // characters from urls and other links), along with white space, punctuation, and capitalisation. At this stage 'stopwords' are also removed from the corpus. Stopwords can be understood as extremely common words, such as 'the' and numerous prepositions (at, in, on, etc) that would not only provide little insight during an exploration of the wider themes of the corpus, but would also show up prominently in visualisations and data due to their frequency. Additional stopwords specific to the project are also introduced at this stage: the word 'harkive' appears in almost every post, as do the words 'music' and 'listening' – and it is almost entirely given that the stories are concerned with listening to music, so removing them at this stage will provide space for the words related to them to be foregrounded. Here we can consider not only the inherent reduction involved with systems of data collection and processing, but also a further stage of reduction and abstraction that is introduced by the role and actions of the researcher. These have productive/reductive potentialities of their own in that they can be as equally informed by specific domain knowledge as they can be limited by questionable assumptions. To continue with the example, however, a quick exploration of the words found in each group's stories is shown in the two R-generated wordcloud visualisations shown in figs 2 and 3 below

fig 2: Wordcloud based on Group A stories

fig3: Wordcloud based on Group B stories

Quick visualisations of this kind can provide an initial insight into the most commonly occurring words in each set (radio, work, album) that may provide a route to specific analysis of certain words, stories, or respondents. For instance, in this example the word 'like' shows relatively prominently in group A but is not foregrounded in Group B. Ultimately, however, the visualisations above reveal more about the commonality between the datasets than their differences. In order to explore how the descriptions of experiences from the two groups may differ (if that were our question), an extraction of the words unique to each is potentially more revealing. The following visualisations (figs 4 and 5) show the words from both sets of stories that are unique to each, and which occur more than 4 times across each corpus. These words are organised according to their

level of frequency across all stories within the group, with the most frequent appearing at the top of the list[20].

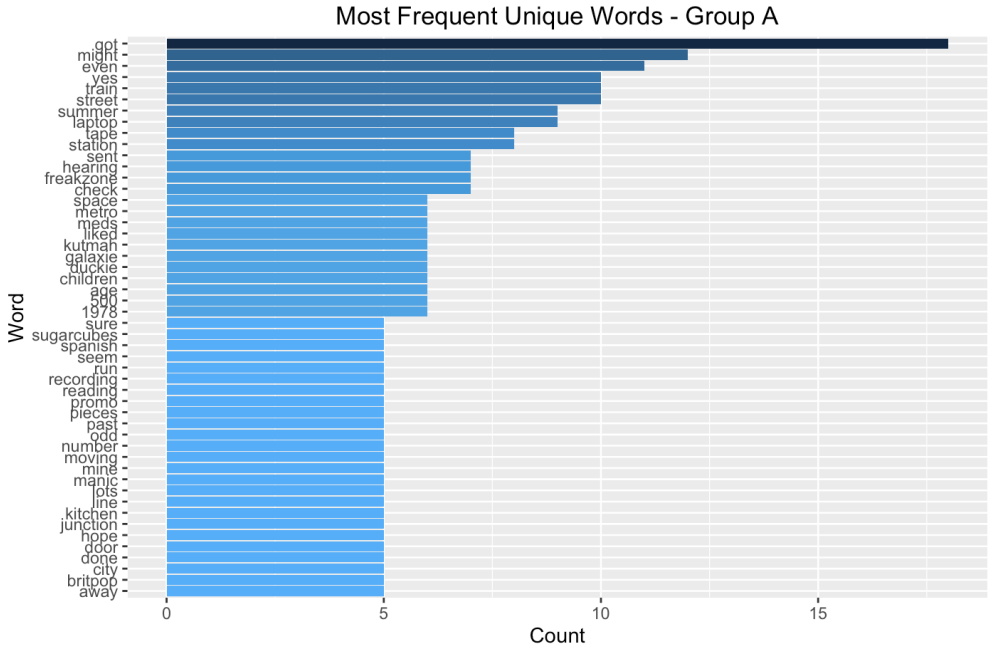

fig 4: Most Frequent Unique Words, Group A

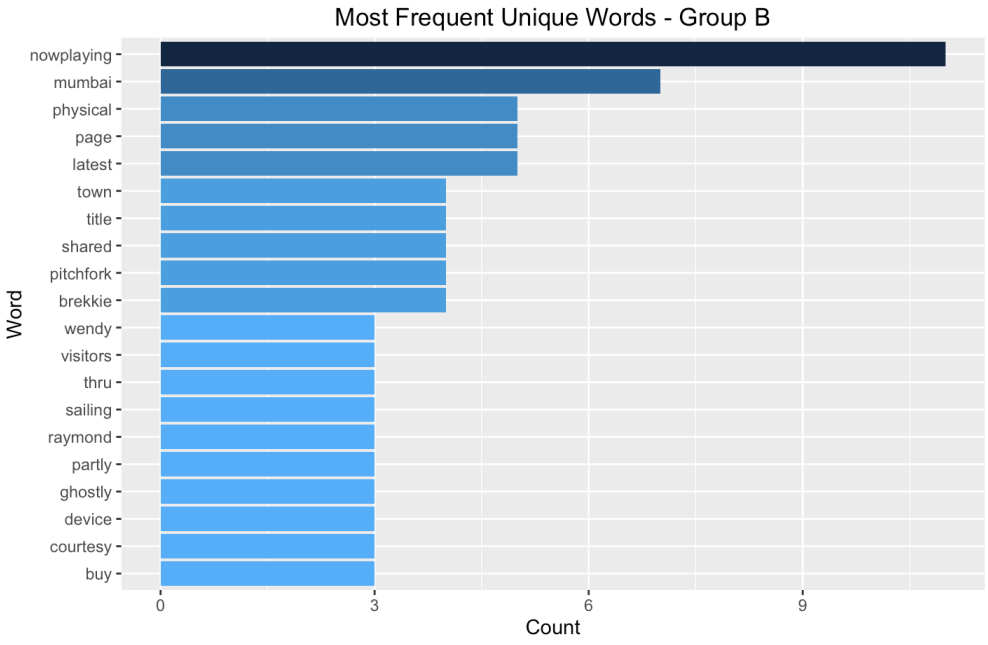

fig 5 – Most Frequent Unique Words – Group B

[20] We may recall here that Group A contained almost twice as many stories as Group B, which explains why there are more words occurring more that 4 times in the first visualisation.

These visualisations reveal three things that can inform approaches at this stage of the modular process:

- The stopwords filter introduced earlier needs to be more robust, which can be achieved through continuous iteration: words such as 'got', 'might', and 'even', which appear with a high degree of frequency in group A, would not reveal insights in this particular instance, and it is unlikely that they will do so in others. As with the removal of words such as 'harkive', 'music' and 'listening' earlier, adding to the stopwords stage with each iterative exploration would help quickly begin to foreground words that are perhaps more likely to lead to insights.

- Equally, however, caution must be exercised when considering removal of words and assumptions should not be made about words that seen in isolation that do not immediately suggest a route to potential insight. A degree of what data scientists call domain knowledge, or what Berry may consider 'the regulating force' of the field specific knowledge of the researcher should thus be introduced alongside such caution. The appearance in this dataset of the words 'galaxie' and 'sugarcubes', for example, are likely to refer to bands, whilst 'freakzone' is the name of a popular BBC 6Music show. Therefore they should not be removed until the context of their appearance can be fully understood.

- Despite, however, the somewhat 'noisy' nature of the data as it presented at this stage of the approach, there are nevertheless kernels of thematic elements beginning to emerge. Words such as 'physical', 'pitchfork' and 'buy' uniquely appearing in group B, for whom exclusivity is an important factor, could be understood as potentially significant in that they may be linked to purchasing preferences (buying physical goods) and preferred recommendation channels (Pitchfork being a well known music media brand). A picture thus begins to emerge at this stage, and as it has at all stages above, it is possible to pause and take a closer look at individual or

small groups of cases that may be of interest. It is also possible for the modular workflow to retreat and reconfigure the creation of clusters.

4.7 – Unsupervised machine learning

As useful as the above process is for an initial exploration of the story dataset, it remains very much that: an exploration. Any further analysis would only produce group-level insight following an extraction and reading on a case-by-case basis, something that may be fruitful and should not be discounted as a possible method of enquiry. Equally, however, this offers no guarantee that useful insight may be arrived at. More importantly, for this process to be able to scale alongside the Harkive story database as it grows year-on-year, and for it to function within the context of the intellectual project of this thesis to examine issues regarding data-derived knowledge in popular music and the research processes associated with its examination, I have developed an approach that utilises unsupervised machine learning and natural language processing techniques. This, however, as Piper (2016) observes, is not "fast" (either in and of itself, or in comparison to manual reading) because such a process "slows us down' in a different way: it 'forces us to be more self-reflective"[21].

Unsupervised machine learning is a case in point. We may recall here the work of Kitchin (2014), discussed in the previous chapter, regarding empiricist and data-intensive epistemologies, as it is helpful in considering the two broad strands of machine learning that are available to researchers. Through methods of unsupervised machine learning, patterns, correlations and other potentially interesting insights are derived "born from the data" through computational techniques that are largely guided by their own internal mathematical logics. This differs from supervised machine learning, a much more common approach in statistical analyses, which usually has the aim of predicting outcomes based on the presence or absence within a data set of feature and response variables. This is the basis of concepts such as inferential statistics and latent trait inference

[21] Piper's article was published online by the Journal of Cultural Analytics and as such no page numbers can be attributed to his quote. The full article is available here: http://culturalanalytics.org/2016/05/there-will-be-numbers/

discussed through Tufecki's (2015) work in Chapter 2. A very basic example of this may look to predict the fuel efficiency of a car – which we can call a response variable – through analysis of combinations of feature variables, including engine capacity, the weight of vehicles, or whether a vehicle uses petrol or diesel fuel. Given a certain amount of existing training data, that contains both feature and response variables, a statistical model can be generated with a good degree of accuracy that is be able to suggest the fuel efficiency of a car based on its weight, engine size, and so on. This model can then be iteratively improved via test data before it is then deployed on a data set that contains only feature variables and from which the model enables the generation of predicted response variables[22]. The goal of unsupervised learning, on the other hand, is primarily exploratory, rather than predictive or explanatory, and usually facilitates instead the creation of sub-sets of a large dataset, as is the case in my own work.

The concepts of machine learning are similar to activities around automated recommendation in popular music and other areas of engagement with cultural content, products and services that were discussed in chapter 2. The goal is to identify and categorise groups and individuals who may have a likely affinity for a particular item. Unlike a car, however, which produces data (weight, engine size, etc) that is common across other cars and – crucially – together can be understood to have the same potential impact upon a response (fuel efficiency), the motivations for consumers of cultural goods contain far too many unknown variables that may impact on their decisions in selecting particular items. Instead, groups and clusters are generated that are based on known variables (e.g. demographics, devices used, whether a song has been played, skipped, or shared, and so on), and then further iterations of data collection can occur (a song is recommended, which is either played or not). It is in this way, as Webster et al showed, that recommendation systems reconfigure existing categories of taste, and also how categories and individuals' allocation within them constantly shift (see: Prey (2015), Cheney-Lippold (2017)). As such, exploration and analysis continues to develop and is often not based on the starting point of a

[22] Pradeep Menon provides a useful overview of this process here: https://medium.com/towards-data-science/data-science-simplified-key-concepts-of-statistical-learning-45648049709e

hypothesis (car *x* will be less fuel efficient than car *y*, person *a* will play song *1*, and so on), rather models of this kind are designed to iteratively improve as more data is collected and the knowledge derived from it is deployed via dynamic interfaces, in real-time, all the time (see the discussion in chapter 1 regarding dynamic interfaces and the work of Tkacz)

Exploring ideas behind such processes is one of the aims of this project and is thus reflected in my method: although the data collected in the form of stories was initially gathered without specific research questions in mind, they nevertheless have enabled the creation of a question. Through following and examining computational processes of analysis in order to answer that research question my suggestion is that we may discover new knowledge about both the data (i.e. the stories) and the processes that lead us to that knowledge (i.e. the modular elements of the method). This process is facilitated by both the quantitative data gathered during the survey element, and also additional variables generated through computational analyses of the text contained within qualitative stories, which also produces quantitative data. As James et al (2013) argue, however, the unsupervised learning approach I am taking is often a more challenging undertaking than supervised learning:

> The exercise tends to be more subjective, and there is no simple goal for the analysis, such as prediction or response…Furthermore, it can be hard to assess the results obtained from unsupervised learning methods since there is no universally accepted mechanism for performing cross-validation or validating results on an independent data set (2013:270)

In other words, there is no 'training' data containing responses that would enable prediction, even if that were the goal, and nor is their likely to be, certainly in terms of the scope of this research. The application of a degree of domain knowledge, which in this case can tell us that – for instance – cultural practices related to engaging with music through new and old technologies are complex (see: Nowak, 2014), leads us to the very reasonable assumption that the response of person A (who may, for instance, describe their vinyl use in particular terms) is no real indicator of whether, why, where, or when person B

may do the same. Thus the goal becomes rather to identify groups of potential interest through unsupervised learning that will facilitate a close, manual reading of texts that is – in turn – interpretative and which can, as Kitchin observes, employ "critical social theory to frame how the research is conducted, how sense is made of the findings, and the knowledge employed" (2014:9). As such an examination of how abstracted variables cluster individual respondents, and what type of insights or interpretations these suggest, is a productive line of enquiry given my broader aim of attempting to understand the role and influence of data-derived knowledge on cultural practices and the research process. This is the main thrust of Chapter 5, which explores the results of a purely computational analysis.

In order to facilitate such a clustering of groups or respondents, the text gathered from their Harkive stories is processed to generate several additional and different numerical and categorical variables. This will be achieved via the two unsupervised machine-learning processes of topic modelling and sentiment analysis. The specific details of these processes as they have been applied to the Harkive corpus will be explored more fully in Chapter 5, but in essence each provide a means of producing at-scale mathematical abstractions of the text gathered. Topic modelling processes text according to the frequency and collocation of words in the separate documents within a corpus, from which groups of documents can be organised according to 'themes'. Sentiment analysis, meanwhile, scores documents based on the appearance of words that are allocated with 'positive', 'negative' or 'neutral' scores. Each process produces several new variables within the main dataset that can be used, along with quantitative data from the survey elements, to facilitate sub-setting and clustering for further exploratory, interpretive analysis.

These new variables – since they are numeric and/or categorical – can be analysed for statistically significant/similar groupings, relationships, or else according to trends within the data, and once again we can recall Prey, Cheney-Lippold and Webster et al's work here in terms of how computational data analysis techniques shift existing categorical variables of audience construction.

In terms of my research process, the computational techniques of topic modelling and sentiment analysis can be used to generate different types of clusters, the generation of additional exploratory visualisations, or instead analysed according to the extent to which changes in one or more variable may be seen as related (or not) to changes in others. To put this into the context of my own research aims, an example here may be that stories gathered via a particular method exhibit a greater degree of higher sentiment scores, or the presence of words related to a particular technology, either of which may be seen to grow or reduce over the course of the collection period of the project. This is work not carried out with prediction in mind, but rather as an exploration of how such processes produce a form of knowledge and how – through an interpretation of that knowledge production – the issues with its potential role in both cultural practices and scholarly work can be examined. Both processes thus align with Meyer-Schoenberger and Cukier's definition of datafication (2014) through the digital conversion of individual actions into data points, highlight the reductions inherent in data collection processes, the apparent neutrality of platforms that facilitate such collection and analysis. The computational techniques used in the pursuit of this are the result of peer-reviewed work by academics and developers working in a number of wide-ranging disciplines, from English literature to computer science. These processes have been formalised as workflows contained within R packages specific to their respective purposes. Whilst these packages have not been developed with the specific task of processing and analysing text data about popular music in mind, they are nevertheless extremely useful for that purpose, as will be shown. By recourse to developer community blogs and forums, I have been able to adapt existing coding scripts and packages developed for the R statistical software package in order to process my own data and produce the results discussed in subsequent chapters[23].

An interesting point raised by this process is related to the role of the human researcher in undertaking the computational analyses that produce the kind of

[23] I have provided an example R script in Appendix B. In addition, the Harkive website contains a number of posts that specifically detail the processes discussed in this chapter. For an example of the use of Topic Modelling and Sentiment Analysis, see: http://harkive.org/h17-text-analysis/

new variables discussed here. This role is often limited to decisions regarding changing a small number of parameters and other settings within complex algorithmic processes that he/she may not necessarily fully understand. There is a gap, then, between the results achieved and the method by which they were produced, which adds further complexity to the levels of reduction and abstraction involved with work of this kind. The complex individual practices involved with engaging with music have, ultimately, been reduced to numeric and categorical variables within a uniform dataset. The construction of knowledge (in the form of findings) proceeds from the point of the results of such abstraction, which could be understood as an approach that compounds the problems of abstraction and reduction. However, and as I have shown in my opening chapters, such activity when engaged in by commercial organisations has a growing influence on the everyday experience of millions of people. In the case of music, such activity can be understood in terms of automated recommendation services, or the foregrounding of certain content within digital interfaces, each of which in turn impact upon real-world experiences. Chapter 3, meanwhile, demonstrated that there is much to be learned from a reflexive use of these technologies through practice-based research that can lead to a critical analysis of them. Thus the innovative and experimental method I have devised provides me with the means of not only answering my central research question, but also a way of engaging with the various issues of debate and method raised in this and previous chapters.

4.8 – Data and research ethics

Given the nature of my data gathering procedures, and in particular a heavy reliance on contributors posting 'stories' and other information to 3rd party platforms, several issues are raised about the ethical use of these platforms for academic research purposes. Indeed, as Townsend, Wallace et al (2016) point out, the relative infancy of social media platforms is such that "there is as yet no clear ethical framework for researchers entering this field" (2016:4), something the authors attempt to address in their work, which I shall return to shortly. To

perhaps illustrate their point, the ethical guidelines published by my host institution, Birmingham City University, are at the time of writing going through a review, one of the aims of which being the provision of clauses related to research of this nature. In creating my ethical statement for Harkive I was grateful, therefore, to my faculties' Research Ethics Convenor, Prof. Paul Long, for his assistance and guidance in ensuring my project met with the standards expected by the University.

I have made best efforts to ensure that the research ethics of my project are robust and that my data collection methods are justified. There remain, however, several ethical questions regarding the nature of my methods of analysis, dissemination, and for my future work, that should be reflected upon. As discussed above, Harkive stories posted to 3rd party social media platforms remain publically available until such time as the user deletes them, and/or their account, or the service concerned, ceases to exist. In terms of dissemination of findings, then, quoting from stories of this kind means it is relatively easy for readers and other interested parties to locate the original post (Narayanan & Shmatikov 2008, 2009). The debates around this particular issue pivot on the nature of a lot of social media research, particularly as it relates to anonymity and consent (see Zimmer (2015), for example). Mining social media platforms for keywords, or around discussions of a particular issue, is increasingly easy to perform and as such large amounts of data can be gathered around a particular topic without the authors of the posts being aware, and certainly without what would be usually understood as their informed consent. Harkive differs in one very important way from research that utilises the affordances of online data-mining in that way, and this difference is to be found in the word Harkive itself. This is a word that did not exist until I invented it, in 2012, and relates directly and only to this research. Variants of it (for instance, the hashtag #harkive) are inserted into social media posts by people wishing to participate in the project, and the automated data collection methods utilised by the project will only look for posts that contain this keyword. My project, then, is not mining keywords from social media users who may be entirely unaware of the research activity (one could mine the popular #nowplaying hashtag in this way, for example, and

gather tens of thousands of tweets about music listening daily), but instead only from people who insert the unique keyword related to the project. A link to a page displaying detailed information regarding the research ethics of the project is prominently displayed on the home page of the project website[24], and at various other points across the site that a related to data collection[25].

Clearly, though, it would be an assumption to suggest that everyone posting to social network sites with the word Harkive has also visited the project site and has read the research ethics statement, but I would argue that it is nevertheless a reasonable assumption to suggest that the overwhelming majority of posts containing the word Harkive have come from people who are telling their stories willingly. In addition, and as outlined in the research ethics statement on the project website, participants have the right to request that their data be withdrawn from the project at any time[26].

I am mindful, however, that any publishing of additional data such as that gathered via the survey (for instance, gender, age, and so on) in conjunction with publically available social media posts, is problematic and should be avoided. Townsend and Wallace suggest that a possible solution here is that quotes are paraphrased rather than reproduced verbatim, thus greatly reducing the possibility that readers can find a route back the original post via search engines. Of potentially wider interest and impact is the response from UK government, the Higher Education Funding Council for England (HEFCE) and Research Councils UK (RCUK) to the Finch Report (2012), that post-2014 REF all published, peer-reviewed journal and conference papers, and (crucially for my purposes) theses that are funded by one of RCUK bodies, must be made available via Open Access methods. Under this edict, some or more of the 'raw' data generated by research projects must also be made available so that the research can be verifiable and replicable.

[24] The Research Ethics Statement is available here: http://harkive.org/research-ethics-informed-consent/
[25] For examples of where edited versions of the Research Statement in Footnote 24 are displayed see: http://harkive.org/how-to-contribute/ & http://harkive.org/submit/. Contributions gathered via email and the online form on the project website generated automatic replies to contributors that also provide information regarding research ethics.
[26] Since 2013 I have only ever received one email on this subject. The respondent was happy for her story to remain in the database, but requested that her name be obscured in any dissemination activity.

To conclude this section on the ethics underpinning my project, I state that I have to the best of my ability attempted to ensure that the rights of participants have been protected and that I have operated within the guidelines of Birmingham City University's Research Ethics statement. The nature of my research raises several issues that scholars in a number of different fields are currently wrestling with regarding performing internet and social media-based research ethically: public/private data; protecting anonymity; reducing risk of harm; Open Access data. This presents an opportunity as I progress to post-doctoral work based upon this data and project, to contribute to the development of ethical research practice.

4.9 – Non-text research outputs

The methodology described above produces a number of non-text outputs both during and at the completion of the process. Therefore, alongside existing routes to output, such as journal papers, conference presentations, and so on, I also have the opportunity to disseminate my work through new channels. Of particular interest here is my research being chosen as a case study in a current project by The British Library, investigating the manner it which it can make non-text outputs of doctoral theses available via their EThOS web service[27]. This project is allied closely to an AHRC-funded initiative, The Academic Book of the Future (https://academicbookfuture.org), which is investigating how "new possibilities [that] are being opened up for researchers in the arts and humanities as a result of technological developments, and how might these impact on the way research is conducted and shared in future" (AHRC, 2016). As such, thinking beyond the traditional format of written theses, journal articles and so on, the nature of my research and the methodology I have developed are ideally placed to take advantage of and contribute towards discussions around the direction of arts and humanities research. A brief introduction to the some of the additional materials produced by my work that may facilitate this are detailed below.

[27] The British Library case study on my work can be read here: https://www.bl.uk/case-studies/craig-hamilton

4.91 – Code and data repository

Each stage of the process of data cleaning, visualisation, and so on, described in this chapter requires a bespoke R script to perform its function. Because even seemingly basic operations, such as re-arranging a data frame into a 'tidy' format, can often require lengthy scripts, it is common practice amongst R programmers to punctuate their scripts with regular comments that describe each step. Such comments have a primary function of reminding the programmer of where they are in the process, but are also useful in collaborative work disseminated via platforms such as Github, or when sharing solutions to coding community platforms via platforms such as Stack Overflow.

When coding scripts are shared for these purposes it is also common practice for sample data sets to accompany them. This enables collaborators to replicate the work of the original coder, either for the purpose of improving on the original code (known as a 'fork' in GitHub) or else when adapting code for their own projects. This speaks to calls from the likes of Housley et al (2014) for the fostering of collaborative work, Piper's (2016) claim that work of this kind is more "democratic", and also the observations of Sandvig and Hargatti (2015) regarding the publishing of digital humanties "benchwork" – in their words, where "researchers can reveal the messy details of what they are actually doing, aiming towards mutual reflection, creativity, and learning that advances the state of the art" (2015:5). As such, my intention is to make the coding scripts I have generated available in a similar manner. Where the ethical constraints of the project allow, I will also make sample data sets of the real project data available alongside them. Where privacy and research ethics prevent this, dummy data sets organised according to the same schema will be posted.

Making these elements of my work available as freely available resources has a number of advantages:

- It provides a mechanism for experienced programmers to improve upon the code, the benefits of which can be fed back into my project.

- Each new release of code and data can be accompanied by complimentary blog posts and/or tutorial videos, thus helping disseminate the findings of my work in ways beyond traditional academic routes.
- The work can be built upon by other academic researchers, particularly in the field of popular music studies, leading to greater potential for not only the impact of my work, but also for the possibility of cross disciplinary collaborative work in my post-doctoral career.
- The repository of code, data and tutorials becomes a teaching resource that can assist me in working with students during my teaching career, but also be available to others who may find either the code, data, method or findings useful in their own.

4.92 – API

Because of the work undertaken during the creation of this methodology, and in particular the elements of data collection, cleaning and organising according to 'tidy data' principles, elements of the 'raw' data collected by the project have been made available via an application programming interface (API) that could develop into a useful resource for fellow researchers in the field of popular music studies.

The function of an API is to allow access to data in a structured, reliable way, so that applications, visualisations, and other tools can be created by making use of the data held within it. The crucial point is that although the data held within an API may change over time, the structure the data is held within remains constant. This means that anything built upon an API is able to change dynamically in line with changes in and to the data, without necessarily having to change its own structural dynamics. As the Harkive project progresses year on year with annual story collection days, the API will grow with that data. In addition to that, results from analysis can be contained within 'buckets' that are also accessible. For example, key themes extracted from text can become additional elements within the API, allowing for searching, application construction, and so on.

4.10 - Discussion

In earlier chapters I have demonstrated how over the last two decades digital and Internet technologies have emerged as an important and influential factor in how popular music is produced, distributed and consumed. These technologies, allied to practices of data collection and computational analysis, now play a significant role both in how audiences engage with music, and how those audiences are now conceived of at the level of the individual through the real-time collection and analysis of data. This is a significant development from earlier categorisations that divided audiences according to class, gender, age, and so on (although, of course, such divisions remains possible). Importantly, as Prey, Webster et al and Cheney-Lippold all show, individuals may now be abstracted into many different versions, and all of the time, depending on the location (in both the real and virtual worlds) and the type of activity they are engaged in. This route towards digital monitoring and multi-faceted abstraction is a process that has run concurrently with over a century of developments within the specific field of popular music that has seen music reception develop in line with commercial models and technological innovation. The primacy of newer modes of listening based on access models that sees engagement with popular music occur through digital and online interfaces intrinsically links these areas of development and brings popular music into some wider societal debates. However, the continued existence of popular music practices that pre-date data and connected technologies – see, for instance, the recent resurgence of vinyl record sales – complicates our understanding of the popular music experience and raises issues related to how people are negotiating the new conditions at both individual and collective levels. This leads to further debates around what happens when individual and collective questions of agency and choice that occur throughout everyday life – something particularly evident in the ways in which we engage with music – meet with the rationalising logic of computational techniques. Ultimately this becomes a debate that considers variously what happens once knowledge is generated by assemblages of data-

driven business models (Hartmann et al., 2014), becomes networked (Lynch, 2016), and plays a role in our everyday experiences.

By engaging with recent academic work from a variety of disciplines that is harnessing and attempting to understand the relationship between data-related systems and society, I have situated my own work (and particularly my methodology) within wider debates and research associated with the "computational turn" (Berry, 2011) in the humanities. It is through exploring the advantages and limitations of these methods that I will provide a new means of understanding the changes, issues, challenges and opportunities for popular music scholars discussed in the opening chapters. Broadly, then, this will be achieved through an examination of the contemporary engagement with music, with a particular focus on the manner in which people responding to The Harkive Project reflect upon their experiences when posting in online environments, and how those environments and the computational analyses they help facilitate inform such an examination. In short, it is through the method I have devised that I am able to approach my central research question: what can an analysis of the data generated by The Harkive Project reveal about the music reception practices of respondents?

The complexity of everyday experience cannot be adequately reduced to a collection of data points, yet this is precisely what occurs both in my project and in online services where the activities associated with individual and collective engagements with music now predominantly occur. My methodology has been constructed in a way that enables me to consider both the role of the socio-technical systems and the role of researcher/analyst. This needs to be considered in terms of the broader framework I have developed in chapter 2 from the work of De Certeau regarding the ongoing negotiation between ideas of strategic place and tactical space. Following De Certeau, but also Kitchin, Mejia and Manovich, we can thus recognise that as researchers we are too engaged in a process of transforming/reducing complex (social) action into that which fits into a mathematical (and thus potentially programmable) procedure. This "polluting proximity" (Liu, 2016) is the crux of the various issues of debate

discussed in Chapter 3, and it is here also where De Certeau's expert may not only "blot out" the philosopher, but also problematically trade his expertise for authority. It is thus here where, to recall Piper, such a mode of working may "fail to resist" (2016). By using such an approach, we are thus compelled to acknowledge the limitations and assumptions contained within it. However – as Liu and Kitchin in particular show us – by being reflexive in our use of such techniques, we can begin to critique the knowledge such processes generate.

As with reductive, 'generate' is a word also used here deliberately – and reflexively – in the sense that, although a project methodology may aim to be robust in and of itself (specifically as it relates to empirical data and a research question), any claims to the arrival at an objective truth through it must be set alongside a critique of the process of generation. This is where Kitchin's conception of methods as models is helpful. Not only are modular methods replicable, verifiable, and so on, they also contain the potential to be iteratively improved upon (through the removal/replacement of discrete elements) without disturbing the overall aims of a given project. This is illustrated by the elements of data collection and analysis in my own method, which can be understood as modular elements of the methodological process as a whole.

Reflecting this back to my discussion in Chapter 3, it's worth returning to Piper, who asks, "at what point did it become necessary, in the sense of unavoidable, to use computation to study culture?" (2016). This is a fair question: what does code have to do with cultural studies? An alternate way to look at this question as it relates to popular music studies is to ask: given the rapid emergence and increasingly central role of computation and code on the ways people engage in acts of music reception, to what extent can we hope to understand popular music cultures without an understanding of code, data, and computation? As Piper ultimately argues, not only can computation help us to develop a better understanding of culture, but the knowledge of cultural scholars can have a beneficial impact on the nature of code.

Whether the field in which this work takes place is digital humanities, or cultural

analytics, or, as I'm attempting in my own work, to operate in and forge a new form of popular music studies that draws upon computational techniques, there is a fascinating tension between the empiricism of numbers, data and scientific process, and the reflexive realisation that the methods used in pursuit of this are ultimately creative acts. Collins et al (Collins et al., 2007) call this a "trading zone" between disciplines. Much is still up for grabs they say since, "no one is sure exactly what it is, what it is called or should be called, who should do it, or how exactly it ought to be done"[28]. Sandvig and Hargattai, meanwhile, argue that – apart from in ethnographic work – there is very little notion of "bench science" in the humanities and social sciences, but that there should be. Their point is that the "workaday" practice of our research processes need to be highlighted, particularly in areas of work that look at digital media and the Internet, because these are producing the "new methods, new opportunities, and new challenges for understanding human behavior and society." (2015:5) As the authors state, the desired outcome is a space where "researchers can reveal the messy details of what they are actually doing, aiming towards mutual reflection, creativity, and learning that advances the state of the art" (2015:5).

As I develop my own analysis across the next three chapters it should be understood as an analysis that has been constructed through a period of practical learning that has enabled me to perform the analysis, but also an engagement with extant literature and theory that underpins why such an analysis is necessary. The inter-related issues of debate and method explored through that engagement are thus woven into the fabric of my modular method. As such it is not only the written work that interprets the data Harkive has collected where I make my contribution to the field, but also in the detail of the step-by-step 'benchwork' that comprises my method. This is a new form of popular music studies, and it is one that I suggest opens up exciting new possibilities for the field. In the following chapters, I apply the modular method I have developed in order to address my primary research question, and engage in reflection upon the results of this.

[28] This quote was reproduced in the Sandvig and Hargattai (2015) article.

CHAPTER 5

Unsupervised machine learning: exploring experiences with formats and technologies

In this chapter I will work through a data analysis of the corpus of Harkive stories gathered between 2013 and 2016. The analytical approach taken will be based on a computational analysis of the stories collected, and of numeric and categorical variables produced through the use of two unsupervised algorithmic processes. These processes are linked closely to analytical techniques associated with the computational turn in humanities research (Berry, 2011; Hall, 2013), and also with empiricist and data-intensive analytical methods that Kitchin (Kitchin, 2014a) has highlighted as being instructive in terms of formulating reflexive computational research techniques in the humanities. The particular techniques used in this analysis are similar also to the type of natural language processing used by Spotify and Echonest that were discussed in Chapter 2, and are emblematic of more widely used commercial processes that take 'freely' available online text-based data from social media channels, websites and blogs in order to produce actionable results as part of data-driven business models (Hartmann et al., 2014).

5.1 – Computational analysis and popular music studies

The findings of this chapter show that alongside numerous benefits to deploying techniques of this kind in humanities research, and particularly in terms of processing large, unstructured text-based datasets, there are equally a number of issues with such an approach. I demonstrate, however, that the computational analytical tools deployed here provide the possibility of highlighting patterns, correlations and trends in a text-based dataset that may not be visible to a manual, or close reading. This allows me to approach and answer my central research question of what an analysis of the data generated by The Harkive

Project reveal about the music reception practices of respondents, whilst simultaneously unfolding an exploration of questions regarding the type and efficacy of knowledge that is produced through computational analytical techniques when they are applied to cultural datasets in humanities research. In taking this approach I will answer the following sub-questions: what can an unsupervised, computational analysis reveal about Harkive respondents' engagement with popular music? What are the advantages, problems and questions raised by such an approach? The aim of this chapter is thus twofold: to explore what types of insight an unsupervised computational analysis can produce from the Harkive dataset, and secondly to explore and critique the manner in which algorithmic processes function when applied to that dataset.

My exploration of the type and efficacy of knowledge produced through the use of computational tools is related specifically to my central research question when the initial corpus-wide analysis facilitates a secondary and more specific reading of the Harkive texts. This secondary analysis examines specifically the technologies, formats and services reported as being used by people responding to Harkive. Unlike recent popular music studies work that take a particular service, format or technology as its focus, for example Hagen, Webster et al, and Prey's work in terms of streaming services (2015; 2015; 2016), or Bartmanski and Woodward's (2014) study of vinyl records, this study instead takes as its starting point what Nowak (2016) calls the "fractured and heterogeneous" ways in which people listen to music. These often include the use of many different technologies, old and new, in what Maguagga (2011) calls a "circuit of practices" that include not just formats and technologies as they have hitherto been understood (i.e. CDs, vinyl records, radio), but also digital interfaces, streaming media, smartphones and other recently developed technologies. Considering this complex landscape of music reception, and through a small-scale case study of the two particular technologies of the iPod and of vinyl records, I demonstrate what a computational analysis can reveal about engagement with music in a variety of everyday contexts – including in commuting, working, social and other situations. These everyday situations are what Felski calls the "mundane activities that frame our forays into more esoteric or exotic worlds" (1999: 5),

and here the focus is on how music and use of various technologies provide the means through which these forays occur.

In order to understand both the patterns and trends in the full corpus and the specifics of the manner in which technologies and methods of engagement are described in the secondary analysis, I have used two different types of unsupervised machine learning algorithms to process the 7582 texts collected by Harkive between 2013 and 2016. These analytical processes help produce a range of new numeric values that I then use in order to foreground texts, or groups of respondents/texts, for further analysis in the manner described in my previous chapter. What this chapter demonstrates is that although what can be understood as the empirical, mathematical methods used are able to facilitate a process of sub-division and categorical allocation that on the one hand has the appearance of being unequivocal and robust (highlighting the empiricist position explained by Kitchin), a closer engagement with the results, and the processes that produce those results, ultimately raises more questions than the procedures themselves appear to help answer.

5.2 – Overview of analytical process

The rationale for taking this approach, explored in Chapters 2, 3 and 4, is grounded in a desire to explore the nature of cultural knowledge creation via computational processing and to relate this specifically to how popular music studies may engage more critically with the role data technologies play in the production, distribution and consumption of music. As my opening chapters have demonstrated, computational techniques are increasingly seen in the commercial recorded music industries, manifesting particularly in the type of automated recommendation and related systems detailed in Chapter 2 (Prey, 2015; Webster et al., 2016), but also within other online and digital environments, such as social networks and media outlets (see for example (Amatriain, 2013; Tufekci, 2015). As we saw in chapter 3, scholars have numerous concerns related to the powerful positions such platforms and

processes now occupy (see: van Dijck (2014; 2013), Housley et al (2014), Savage and Burrows (2007)).

I reiterate here the words of caution and advice found in the work of Kitchin, Liu and others (2014b; 2016) that were discussed also in Chapter 3: processes and techniques such as those used in this chapter should be used reflexively and critically. To that end, in presenting the results of my computational analysis of Harkive stories I also highlight and reflect upon the detail, benefits and limitations of taking such an approach. As such, my contribution is to our understanding of not only how research of this kind may be undertaken, but also how the computational analysis of cultural data can be usefully, critically and reflexively interpreted. We can recall also Kitchin's suggestion from Chapter 3, that methods for work of this kind should be conceived of as modular processes. In terms of the progression of the overall analysis (from data collection to analysis, to findings, for example), the two unsupervised machine-learning techniques have to be understood as discrete yet component parts of the modular method I have devised. Understanding the role and function of those components parts, and how they fit into the trajectory of the research process, is key, and this is because the peculiarities and affordances of each process will be shown to have a direct impact on the type of knowledge produced. As such, an exploration of and reflection on the two unsupervised learning algorithms deployed here represent further findings of my research for popular music scholars who may themselves wish to use similar techniques.

5.21 – Topic modelling

The first of the two processes used is topic modelling. David Blei (2012) defines this as a process that "provides a suite of algorithms to discover hidden thematic structure in large collections of texts. The results of topic modeling algorithms can be used to summarize, visualize, explore, and theorize about a corpus"[1].

[1] Blei's article appears online in the Journal of Digital Humanities, and as such no page numbers are attributed to his quotes. The full articles is available here: http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/

Topics can be understood as recurring data points (in this case, words) across a larger dataset (a corpus of text documents). The model, meanwhile, is a mathematical representation of the extent to which each individual entry in a dataset – i.e. the documents within a given corpus – contain data points – i.e. topics/words. Applying this to the case of my own research and the data I have collected, the immediate potential in a process of this kind is that the larger a collection of documents is, the more difficult and labour intensive it becomes to manually explore, encode and reveal common themes within it. Beyond simply saving time and effort, however, a further potential advantage lies in the possibility that there may be themes or topics within the dataset that are not immediately apparent. In other words, they may be hidden or otherwise made 'latent' by the complexity and scale of the corpus. The rationale behind topic modelling is that these latent themes may be revealed by a method that is capable of processing large and complex text-based datasets at scale. We can recall from the discussion of unsupervised learning in the previous chapter that such an approach is characteristic of exploratory, unsupervised learning methods, in that the process of automated topic modelling can reveal, as Blei suggests, what can be understood as the 'latent' or hidden themes within the Harkive dataset.

I have employed the most commonly used approach in automated topic modelling, latent dirichlet allocation (LDA), which was originally developed in 2003 by Blei and colleagues at Berkley (Blei et al., 2003)[2]. He has since argued that automated modelling of this kind can help humanities scholars "build a statistical lens that encodes..specific knowledge, theories, and assumptions about texts" (2012). Blei reveals that LDA in particular can be understood as being based on two assumptions: that there are a finite number of patterns of words, or groups of terms, that occur together within a corpus, and, secondly, that each document within a corpus exhibits these to a varying degree. Although the documents and words are observable to a manual reading, the thematic structure – in other words, the topics – may be hidden, and will be more likely to

[2] Perhaps an indication of the extent to which LDA has been adopted as an analytical technique is the fact that the original article has over 16,000 citations, according to Google Scholar.

be hidden as the scale and complexity of a corpus increases. LDA infers these hidden structures based on what it can computationally process (i.e. by counting the frequency with which unique words occur across documents), and represents these in terms of the probabilistic likelihood that a document belongs to a given topic.

The process of revealing this hidden structure is mathematical and iterative and begins by assigning words randomly to a given a number of topics (k)[3]. Once (k) has been assigned a value the LDA process begins by providing an initial structure of word distributions across topics and the extent to which each document exhibits a probabilistic relationship to each of the (k) topics. The results at this early stage would be largely useless since the initial distribution is entirely random. It is from these initial results and structure, however, that further and numerous iterations can begin that gradually improve the results[4]. The LDA algorithms calculate the probability of each word appearing in particular topics and documents whilst assuming all other allocations are correct. This either leads to a word being assigned to a different topic, or (less likely, at least in the initial stages) confirmation that the original allocation was 'correct'. This process is repeated several thousand times across each word within the corpus, and the model thus improves iteratively with each step. Eventually each topic is understood in terms of the extent to which it contains words, and each document is understood in terms of it containing varying degrees of different topic exhibition. This is based on both the words a document contains and their frequency relative to their appearance in other documents. These results are represented numerically, with each document being allocated to the topic it most closely aligns with. The extent to which each document displays a relationship to all topics (n = k) is provided as a series of additional variables that, when added together, produce a total of 1. Each document is thus shown to display a degree of 'belonging' to all (k) topics. The most closely associated topic in this sequence of (k) numbers is represented by the number

[3] A commonly identified drawback with this approach is that (k) must be assigned before beginning the process. As such, the researcher assigns a value to (k) and observes the results before being in a position to decide that (k) is correct. This necessarily means that several runs have to be made before the 'correct', or most useful (k) value is alighted upon.
[4] One of the parameters the research sets in this process is the 'Burn Rate', which discards (n) initial results based on the premise that the first several hundred (and sometimes thousand) iterations will produce incomplete or unsatisfactory results.

that accounts for the highest proportion of the total score of 1, and this is the topic to which the document is allocated.

5.22 – Sentiment analysis

Sentiment analysis, sometimes referred to as opinion mining, has been defined by Liu (Liu, 2010:3) as the "computational study of opinions, sentiments and emotions expressed in text". Much like the process of topic modelling described above, sentiment analysis affords the possibility of at-scale, automated processing of complex, unstructured text documents in order to reveal often latent or hidden expressions of sentiment or opinion. Like topic modelling it is also an exploratory process that produces numeric values based on text and from which individual documents can subsequently be grouped together according to numeric similarities, differences, or statistical relationships. Sentiment analysis algorithms search documents for the appearance of certain words that are then individually scored, producing an overall value that marks a document as either exhibiting a positive, negative, or neutral sentiment.

In order to achieve this form of analysis on the Harkive corpus, I used the syuzhet package developed by Matthew Jockers (2015)[5]. His aim in developing this was to attempt to reveal "fluctuations in sentiment" as a means of understanding narrative structures within works of literature. The R package Jockers developed references four different dictionaries of terms (bing, affin, nrc and syuzhet), each of which scores the appearance of words according to slightly different scales that together result in an overall score for documents along a positive/negative scale. In addition to this, the nrc dictionary component produces several separate scores for seven different components the algorithm groups according to what its developers consider to resemble themes of anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Because these sub-sections are recorded as distinct variables, documents interpreted as exhibiting these various properties may be separately foregrounded through visualisation

[5] Jockers describes the development of this package on his website: http://www.matthewjockers.net/2015/02/02/syuzhet/

or clustering.

Swafford (2015), however, has identified a number of issues with this approach that should be considered. Firstly, words are processed in isolation and only once, meaning that the presence of modifiers, negations or repetitions of target words are not considered in the allocation of overall sentiment scores. As Swafford explains, 'I am happy' will attract the same positive score as both 'I am very happy' and 'I am not very happy', based on the process scoring only the word 'happy'. In addition, Swafford argues that the lexicons are poor at interpreting nuance within positive or negative language. For instance, the words 'good' and 'wonderful' are scored equally as positive, when arguably the second has a stronger positive emphasis. A further problem revealed by using this process on the Harkive corpus in particular is that stories differ in length, from short tweets to longer responses gathered via email and other sources, and as such longer stories produce larger numbers based purely on the presence of more words. I have made some attempt to address this by generating additional variables that produce values based on sentiment scores in relation to word counts, but the disparity and other issues highlighted by Swafford nevertheless remain and should thus be considered when interpreting results.

These limitations are clearly a barrier to a defensible position regarding the emotional content of a story, and add a further layer of complexity to what Berry refers to as the inherent reduction and abstraction involved in the process of converting social action into data points. Given that, I would exercise extreme caution in terms of making claims at either corpus or single entry level. Nevertheless, sentiment analysis (and, indeed, topic modelling) are processes widely used in commercial settings (Lim and Buntine, 2014) and thus in terms of an initial exploratory process that is able to both reveal broad trends or clusters for further analysis and simultaneously reflexively critique those processes concerned in terms of their efficacy as research tools, it is a useful starting point.

5.3 – Applying LDA topic modelling and sentiment analysis

The Harkive dataset contains 7927 entries, each of which represents an individual response to the Harkive project in the form of a text-based personal music story. Each entry in the dataset also contains additional variables such as the year in which stories were told (2013 to 2016) and the method through which they were collected (including Twitter, Facebook, email, and others). As was shown in Chapter 4, these have been collected and organised according to the principles of tidy data (Wickham, 2014), which allows for more efficient computational analysis. For the purposes of this stage of the analysis, however, we need only be concerned with the variable within the tidy dataset that contains the text from the stories collected. With almost 8,000 entries, the Harkive dataset not only meets Blei's definition of a large collection of texts, but also, and given the highly individualised nature of the stories, it potentially contains themes (topics) and sentiments that may not be immediately visible to a manual reading.

In terms of the context of this research project, we can recall the discussion in Chapter 2 of Echonest's collection of online text-based data related to conversations about artists, genres and songs, and how the results from processes such as this are deployed via interfaces where capture, analysis and output are integrated (see: Rieder, 2016). This in turn enables digital and online interfaces to become both "consequential gatekeepers" (Tufekci, 2015) and "producers of sociality" (van Dijck, 2013). The results from the computational processing of Harkive texts should thus be considered in terms of being redolent of the potential outcomes of processes of this kind, not limited to the manner in which respondents may be classified and grouped according to results. Prey, Webster el al and Cheney-Lippold (2017; 2015; 2016) have all variously demonstrated that these results have the potential to change existing categorises and conceptions of taste and identity. By working through a version of this process, Harkive further acts as an experimental place/space that is able to explore the mechanisms, processes and outcomes of the data-derived abstraction of the cultural practises associated with music reception. Again,

however, and recalling Kitchin's (2014a)consideration for modular methods, and also Berry's "inherent reductions and abstractions" (2011) en route to knowledge, we should consider the detail involved in each step of the process, because these reveal further issues of debate and method. Once again, the exploration and reflection of the detail of working through such processes are presented here as additional findings and context that may assist popular music scholars approaching texts (e.g. reviews, lyrics, etc) in their own work.

Before the LDA Topic Modelling process can begin, the corpus of text is pre-processed so that certain elements are removed. This includes:

- Removal of all punctuation and other extraneous characters (e.g. @, #, //)
- Removal of words that occur with very high frequency in written text, commonly known as 'StopWords'. (e.g. *the*, *it*, *at*, *were*)
- All text is converted to lower case (i.e. to avoid the counting of *'Vinyl'* and *'vinyl'* as separate entities)
- Removal of 'whitespace', such as that which occurs between paragraphs
- All words are 'stemmed' to their roots (i.e. to avoid *'played'*, *'play'*, *'player'* and other derivations being counted a separate entries)
- Removal of specific additional words that occurred with very high frequency in this particular corpus – for example: 'harkive', 'music', 'listening', 'project'[6]

The reduced corpus is then converted to a document-term matrix (DTM), which provides a mathematical overview of the distribution of words within each document. The rows in this matrix represent the individual documents (7929), and the columns represent the unique terms within the corpus (17517). The resulting values in each cell represent the number of occurrences of terms to documents. It is this process of reduction and abstraction that facilitates the algorithmic processes of topic modelling and sentiment analysis. Following the

[6] It may appear counter-intuitive to remove the words 'Music' and 'Listening' from an analysis of music listening stories, but the stories are all accounts of music listening – in other words, this much can be safely assumed – and given the high frequency with which these terms occur, their presence in the corpus is likely to skew results in that all other words used would be considered by LDA in relationship to them.

removal of sparse terms (i.e. those with zero values), the terms present in the corpus to be processed via LDA topic modelling is reduced to 1,935 and the document count is reduced to 7528[7]. It should be noted at this stage that although this pre-processing represents a significant reduction from the texts contained in the original dataset, it has not altered the size or structure of that original dataset. Any observations gleaned from the process can thus be tested against the original corpus. Each story has a unique identifier that can lead to a close reading of the original text, which retains all elements removed at the processing stage, including punctuation, stopwords, sparse terms, etc.

Having processed the corpus via the LDA topic-modelling algorithm, the stories are then processed separately by the sentiment analysis algorithms to create sentiment scores for each of the 7528 entries. This results in the creation of several new numeric variables containing values that denote positive, negative or neutral scores. Neutral values denote stories where either positive and negative values cancel each other out, or else where stories do not contain words present in the dictionaries provided by the R package. At this stage, then, following the collection of data according to principles of tidy data, some pre-processing and reductions, and the application of two commonly used machine learning algorithms, the Harkive stories in the database – each of which represents a specific and individual cultural practice associated with music – have been converted into a dataset of unified, numeric abstractions. Now in this format, further exploration through clustering and visualisation becomes possible and the central research question can be approached.

5.4 – Exploring results

Using the topicmodels package with the R software, and based on iterative runs that identified a k-value of 5 as the most productive number of topics to base the analysis on, the LDA process revealed the following words associated with each

[7] Because of the high number of unique terms (> 17,000) in the Harkive corpus, most terms do not appear in most documents – i.e. infrequently. The LDA process can reduce the DTM based on relative document frequency, which takes place on a sliding scale of 0 to 1, where the lowest parameter will retain all terms/documents with the corpus, and the highest only those with frequency across the entire corpus. For further information on this, see the following discussion: https://stackoverflow.com/questions/28763389/how-does-the-removesparseterms-in-r-work

of the 5 topics. These are organised in the table below (fig 6) according to frequency, with the top 20 words associated with each topic displayed in descending order.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
1	morn	time	play	album	song
2	home	post	sound	radio	love
3	nowplay	tweet	record	spotifi	play
4	ipod	video	track	listen	time
5	offic	head	tri	playlist	vinyl
6	tune	youtub	day	track	boy
7	shuffl	hour	album	car	live
8	walk	look	peopl	bit	sing
9	thank	project	actual	iphon	hear
10	soundtrack	write	friend	mix	band
11	headphon	juli	hear	stream	start
12	train	stori	time	start	summer
13	pop	watch	world	bbc	sun
14	commut	night	found	get	version
15	bus	blog	feel	itun	favourit
16	whilst	stuff	heard	phone	beauti
17	drive	check	lot	week	girl
18	hot	bit	decid	even	littl
19	singl	final	band	enjoy	blue
20	danc	late	piec	soundcloud	make

Fig 6: Top 20 frequently occurring words in each of the 5 LDA topics

It is here that a degree of domain knowledge and familiarity with the corpus becomes essential in terms of interpreting results. Clearly, the LDA process has no knowledge or context about the contents of the documents in the corpus and, as we can recall from the outline above based on Blei's work, has processed the text contained within the corpus purely according to the relative statistical probability of words appearing together (or not) in documents. Despite this, however, the five topics identified by the process can be interpreted as being relatively distinct and related to ideas of:

- Time, motion and place
- Contributing to the Harkive Project
- Listening, hearing and choice
- Formats and technologies

- Emotions and experience

In fig 7 below, the words that can be understood to closely relate to these themes have been highlighted, and the column names have been amended and colour-coded to reflect the different topics. The high frequency of words such as 'morning', 'home', 'office', 'walk', 'train', for instance, within the same topic suggested that, according to the algorithmic thematic analysis at least, topic 1 contained themes of time, motion and place in the stories allocated to this group. Likewise, the grouping together of words such as 'album', 'radio', 'spotify', 'playlist', 'iPhone', 'soundcloud' would suggest that technologies and formats of listening are recurring themes in another.

	B	C	D	E	F
	Time	Harkive	Listening	Formats	Emotion
1	morn	time	play	album	song
2	home	post	sound	radio	love
3	nowplay	tweet	record	spotifi	play
4	ipod	video	track	listen	time
5	offic	head	tri	playlist	vinyl
6	tune	youtub	day	track	boy
7	shuffl	hour	album	car	live
8	walk	look	peopl	bit	sing
9	thank	project	actual	iphon	hear
10	soundtrack	write	friend	mix	band
11	headphon	juli	hear	stream	start
12	train	stori	time	start	summer
13	pop	watch	world	bbc	sun
14	commut	night	found	get	version
15	bus	blog	feel	itun	favourit
16	whilst	stuff	heard	phone	beauti
17	drive	check	lot	week	girl
18	hot	bit	decid	even	littl
19	singl	final	band	enjoy	blue
20	danc	late	piec	soundcloud	make

fig 7: Top 20 frequently occurring words in each topic. Colour-coded words are those which have been interpreted as closely aligned to the interpreted topics

Moving now to the initial results of sentiment analysis, we can recall that one of the four dictionaries used provided a breakdown of sentiment into 7 different sections. Fig 8 below shows that the algorithmically produced results suggest there is a broad trend towards positive, rather than negative sentiment within

the Harkive stories. Sections such as joy, trust, and anticipation produce much higher scores than those with negative connotations, such as anger, disgust and fear.

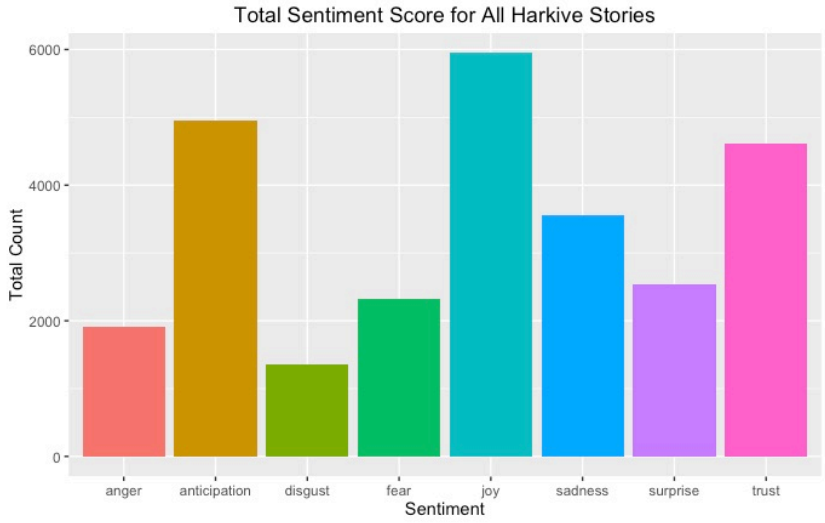

fig 8: Results of nrc sentiment analysis of Harkive stories, split according to 7 categories.

Combining the results of the four different sentiment analysis dictionaries to divide stories into three distinct groups of positive, negative and neutral sentiment further illustrates that the overall trend is replicated by the other algorithms used in the process. That all algorithms should produce similar results is not surprising since – for example – the word 'happy' will attract a positive score when processed by each.

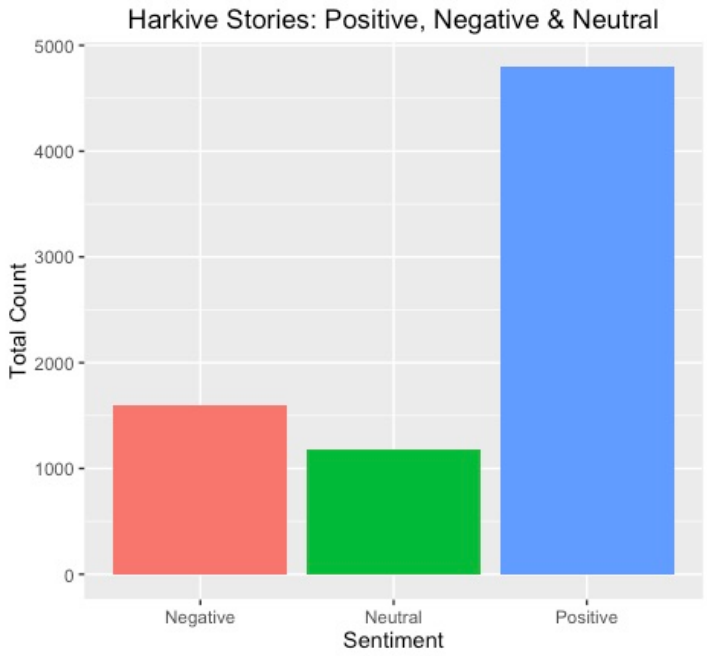

Fig 9: Average Sentiment of Harkive stories, scored either Positive, Negative or Neutral

That the results on the corpus suggest that stories are overall positive is also not surprising given the nature of the Harkive project and the manner in which it communicates its intentions to potential respondents. It would seem a reasonable assumption that people may focus somewhat more on the positive aspects of their experiences with music, and indeed the overall tone of the promotional material accentuates the positive aspects of the popular music experiences. However, and as we'll see here and in later chapters, a Harkive story recounting a 'positive' experience is not always the case and once again we can consider how, and how reflexively, the results of algorithmic processing could and should be deployed.

5.5 – Interpreting results – anatomy of a Harkive story

An interesting and immediate finding that can be identified from creating and running the topic modelling process is that across the entire corpus there is what appears to be a thematic manifestation of what I would call the anatomy of an idealised Harkive response. The project asked people to describe how, where and why they listened to music, and stressed that it aimed to capture

experiences with formats and technologies[8]. Based solely on the results of an unsupervised computational process and subsequent interpretation of the results, it can be said that the corpus as a whole appears to demonstrate that the project has indeed captured that which it aimed to. This is linked to what Blei described above when "the results of topic modeling algorithms can be used to summarize, visualize, explore, and theorize about a corpus" (2012) – in this case, the process provides a degree of validation that the corpus contains that which it aimed to collect. As for Blei's claim that LDA enables the discovery of "hidden thematic structure[s] in large collections of texts" (2012) we are less clear, however, since the initial results appear only to confirm what we may have expected to see. In other words, themes such as time, motion and place, or formats and technologies are not necessarily hidden. That being said, at this stage of Piper's modular process we can at least proceed to the reasonable assumption that the data collected contains information that warrants further enquiry. Researchers exploring other data sets (for example, historic news articles, or song lyrics) may enter that process without expectations of what they may find, and so the results of LDA processing may reveal useful starting points. When dealing with large cultural data sets, and particularly those that may have been collected without specific research questions in mind, or where the initial goal is exploratory, topic modelling thus appears to be a useful tool, but it is clear that further and more detailed enquiry is required.

A route towards such further enquiry can be shown from another way in which the broad overview produced by the LDA topic modelling process can be interpreted and explored, and by concentrating instead on the extent to which each topic is represented within the corpus as a whole. Documents, we can recall, have been individually scored in terms of their relationship to each topic, and thus have been assigned to the topic they most closely align with. From the data visualised in Fig 10 below, it would appear that 'time, motion and place' is the

[8] From the Harkive website: *"Harkive will return for its fourth year to once again collect stories online from people about the detail of their music listening experience"* & *"The project asks people to tell the tale of **How**, **Where** and **Why** they listen to music on a single day each year, with the aim of capturing for posterity a snapshot of the way in which we interact with the sounds and technology of today."*

most highly populated topic, followed by those related to 'contributing to Harkive'.

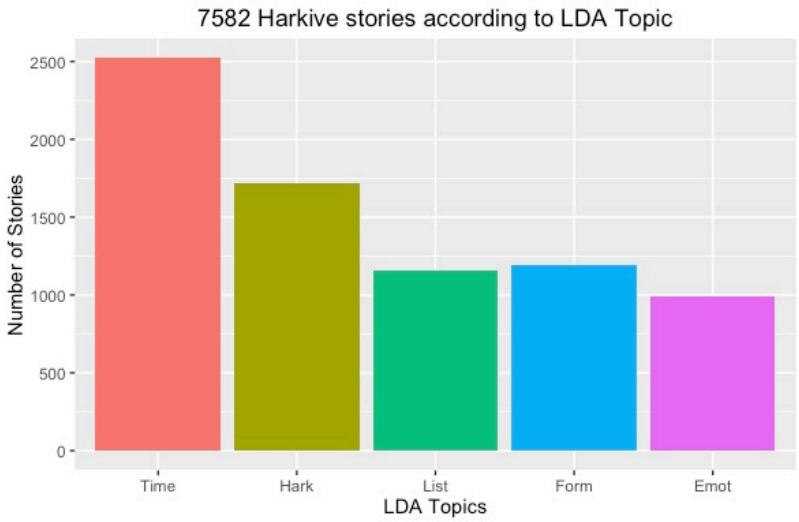

fig 10: Topic allocation across entire corpus

A closer look at the numbers involved here, however, reveals another finding from an engagement with this process: the differences between documents represented by their numeric alignment with discrete topics is not as unequivocal as the visual overview above may suggest. This in turn has consequences for the manner in which the results (or insight) may be deployed. We can recall from the description of the topic modeling process above that documents within a corpus exhibit relationships to all topics in varying degrees. The figures in fig 11 below show those relationships to the different topics in terms of a total score that is equal to 1. We can observe that differences between documents and thus their relationships to topics are extremely marginal and suggest that further enquiry is necessary[9] before drawing conclusions based on these figures.

[9] Not all document scores are as close as those shown in the table here. Certain documents score in the regional of 50% for one topic and can thus be said to closely exhibit properties more aligned with one topic.

	Time	Hark	List	Form	Emot
1	0.18867925	0.20754717	0.20754717	0.18867925	0.20754717
2	0.18867925	0.18867925	0.18867925	0.22641509	0.20754717
3	0.21428571	0.17857143	0.23214286	0.17857143	0.19642857
4	0.18518519	0.2037037	0.18518519	0.24074074	0.18518519
5	0.18867925	0.20754717	0.18867925	0.18867925	0.22641509
6	0.19230769	0.19230769	0.23076923	0.19230769	0.19230769
7	0.20754717	0.18867925	0.18867925	0.18867925	0.22641509
8	0.19642857	0.21428571	0.19642857	0.19642857	0.19642857
9	0.17857143	0.17857143	0.23214286	0.19642857	0.21428571
10	0.22222222	0.22222222	0.18518519	0.18518519	0.18518519

fig 11: Topic Allocation scores for the first 10 documents in the corpus. Each row produces 5 values, each associated with a topic, that are divisions of 1.

Based on the marginal nature of the difference between topic allocations, a closer examination of the results of the LDA process would therefore be useful. What we can consider at this stage, however, is the fact that despite the very marginal numeric differences, the documents (and thus the respondents who produced these documents) have nevertheless been placed into particular and distinct categories, or De Certeau's receptacles of proper procedure. This invites us to consider questions of how algorithmic/computational systems produce actionable results about users, content and other cultural data. The commercial systems discussed in earlier chapters, such as automated recommendation, content targeting, and the foregrounding of particular types of content to certain users through dynamic interfaces that iteratively improve/produce models of categorisation, are in part informed by processes exactly such as the ones used in this chapter. In terms of music, Webster et al (2016) show these systems have an influential cultural role. Bucher (2016), Van Dijck (2013) and O'Neill (2017) meanwhile show how similar systems are used elsewhere, in various aspects of everyday life, from retail to social life to government. Furthermore, Manovich (2016), Piper (2016) and Kitchin (2014b) show how similar techniques are now central to many research processes, some of which aim to study the questions raised by concerns over the same systems they deploy. In all cases, the raw materials involved are computational reductions and abstractions of often complex, real world activities. Based on the marginal nature of the results

produced, through a process I have shown to be based on a hugely complex, modular chain of assumptions, reductions and abstractions, we may ask what consequences are for the listener, the artist, genre, or person, of appearing in one marginal category and not another. These questions are further compounded when we consider that the abstracted, numeric results of the topic modelling and sentiment analysis procedures can be analysed in combination and in terms of existing variables within the dataset. Topic allocations can, for instance, be considered alongside the method by which the original story was gathered. Sentiment can be examined in terms of the different years in which the stories were told. Both hypothetical processes facilitate different formulations of clusters of stories and respondents that highlight Nissembaum's (2009) observation regarding radically decontextualised data. Indeed, this layered decontextualisation can be demonstrated with the Harkive data as new variables can be created based on combinations of the values generated by the automated analysis and existing variables.

To illustrate this I have generated values for the standard deviation between the topic allocation scores in order to provide an indication of how strongly a story aligns with the topic allocated to it through the LDA process. Word and character counts have also been created to produce additional numeric variables that can be used in conjunction with sentiment analysis scores. Both processes, facilitated by the choices of the researcher and the fact that complex social action has been reduced to ordered, numeric data points that can be statistically processed, demonstrate that variables can be combined to facilitate clustering of stories and respondents in ways that can occur at several abstracted removes from the original text (or action). The visualisations below demonstrate how automated processing has converted the single column of text-based data in the original Harkive stories into a number of different numeric and categorical variables that can be used to both explore the 'hidden' structure of the corpus and to cluster and organise stories and respondents according to a number of different configurations. Each of the visualisations below demonstrates one such configuration of variables and thus each requires a brief explanation. Fig 12 highlights that the topic interpreted as relating to 'time, motion and place'

contains a higher proportion of negatively scored stories compared to other topics, and that stories related to the 'listening' topic contain a higher proportion of positively scored stories than other topics[10]. Meanwhile, fig 13 shows that stories collected from Blogs contain a higher proportion of stories related to the 'contributing to harkive' topic, and that almost 50% of stories gathered from Instagram have been allocated to the 'time, motion and place'. In fig 14 we can see that stories gathered in 2013 and 2015 contain a slightly higher proportion of positively scored stories. Meanwhile, Fig 15 combines several of these variables to show the variance in sentiment scores across topics.

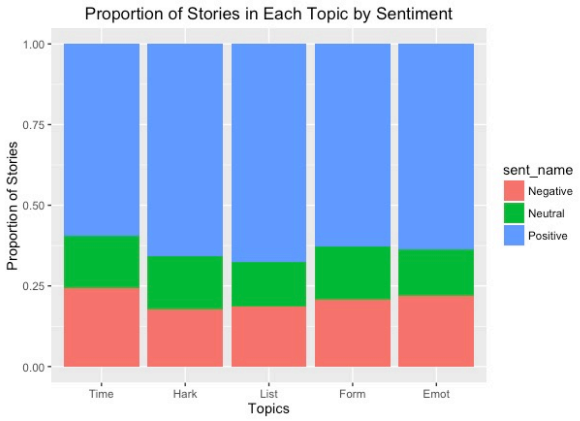

fig 12: Proportion of stories in each topic according to sentiment

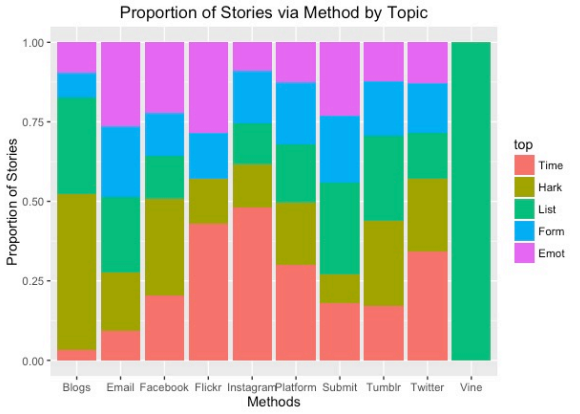

fig 13: Proportion of Stories via each method assigned to 5 topics

[10] There may be a number of possible reasons for this that could be further explored; for instance, the frequency of terms related to commuting in the Time, Motion and Place topic may account for the higher proportion of negative stories, and should that be something the researcher may wish to explore he/she could look for statistical correlations between negative sentiment scores and the appearance of such words.

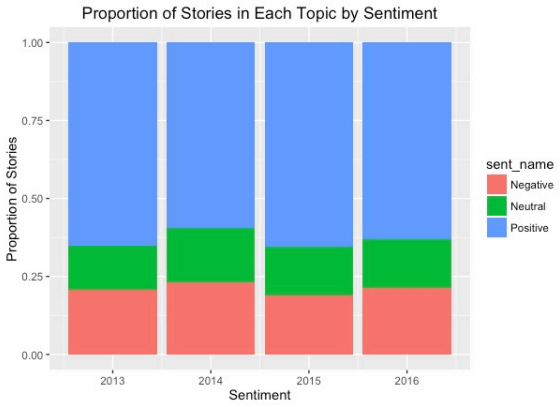

fig 14 Proportion of stories per year according to sentiment

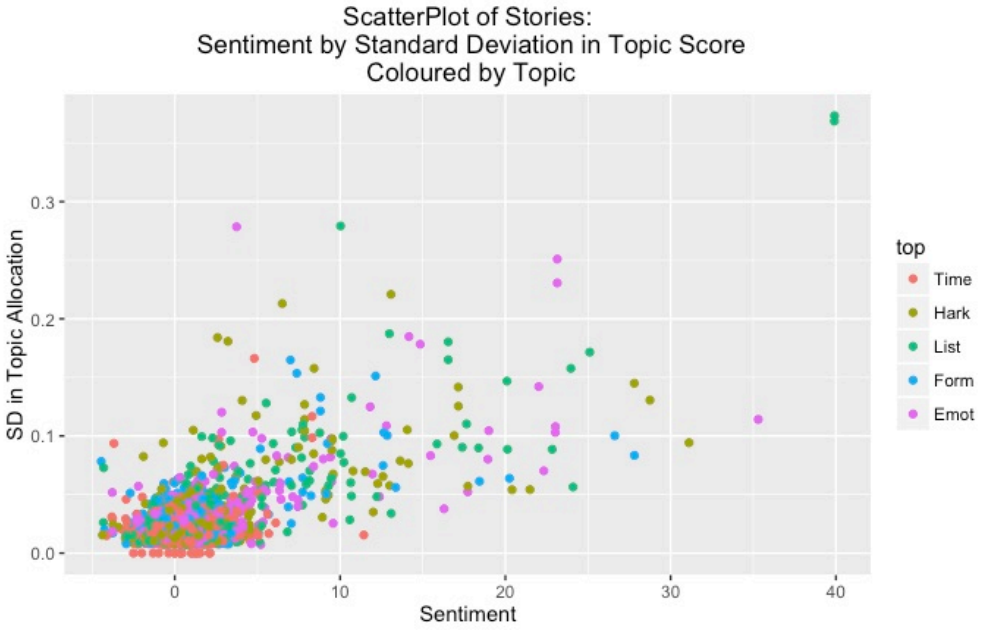

fig15: Scatterplot of Sentiment by Standard Deviation in Topic Allocation, by Topic

All of these visualisations demonstrate a means by which the researcher can begin exploring the wider corpus, potentially highlighting individuals, or groups of stories/individuals that on the one hand may raise questions for further analysis, but on the other demonstrate the relative ease with which respondents or other numerically abstracted elements may be grouped in different categories depending on the type (and rationale) of analysis used. Here we can recall Prey (2015) and Cheney-Lippold's (2017, 2011) observations from Chapter 2, regarding the issue of how automated processes of this kind subject individuals to destabilising processes of 'individuation', and where the efficacy of recommendation systems in particular should be considered not only in terms of

whether the recommendations are useful to users, but also in terms of the extent to which the way individuals are understood and their choices subsequently influenced are a result of the processes concerned. We can consider this in more detail by returning to the corpus level results of the topic modelling process, and the allocation of words associated with the formats and technologies topic, to explore how the computational processing used may produce results about respondents' use of particular playback technologies.

5.6 – Vinyl as an 'emotional' format; iPod as 'functional' device

We can recall that the LDA procedure produced 5 topics that were interpreted as containing to varying degrees elements of the following themes: 'time, motion and place'; 'contributing to the Harkive project'; 'listening, hearing and choice'; 'formats and technologies'; 'emotions and experience'. Fig 7 above highlighted the words that were interpreted as being closely associated with the themes that appeared to be revealed by each topic. By extension, respondents providing those stories can also be clustered and organised according to those topics, in a manner that is redolent of the issues raised by Prey and Cheney-Lippold above, and by the various theorists discussed in chapter 2 (eg: Bucher, 2016; Tufekci, 2015; Van Dijck, 2014; van Dijck, 2013). To illustrate some issues with this, and as Fig 16 demonstrates, there are words that we can interpret as being related to particular topics that instead appear in others. For example, with regard to the 'formats and technologies', the words iPod, shuffle and headphones appear in the 'time, motion and place' topic. The word 'vinyl', meanwhile, is present in 'emotions and experience' topic.

At first glance observing this appears to speak to and perhaps confirm certain common assumptions about those particular technologies. Such an observation also supports the arguments of certain popular music scholars. The analogue format of vinyl is often described in terms of being an authentic, somehow more valuable experience than digital formats, see for example Barthamski and Woodward (2014). The allocation of the word vinyl to a topic by the LDA process alongside words such as love and beautiful, rather than work or commute for

instance, lends support to such claims. It could be argued that – according to the interpretation suggested by the LDA process, at least – vinyl records can be understood as being discussed by Harkive respondents in more experiential and emotional terms, particularly if we follow the empiricist position that 'data can speak for themselves'. Meanwhile the word iPod appears in a topic alongside words related to commuting, the office and time. Here we can see echoes of, for example, the work De Nora (2000) and Bull (2006), who have closely linked mobile listening devices with the management of exactly those everyday scenarios. Given what I have already shown to be the marginal nature of these topic allocations, however, such a conclusion requires further examination.

	Time	Harkive	Listening	Formats	Emotion
1	morn	time	play	album	song
2	home	post	sound	radio	love
3	nowplay	tweet	record	spotifi	play
4	ipod	video	track	listen	time
5	offic	head	tri	playlist	vinyl
6	tune	youtub	day	track	boy
7	shuffl	hour	album	car	live
8	walk	look	peopl	bit	sing
9	thank	project	actual	iphon	hear
10	soundtrack	write	friend	mix	band
11	headphon	juli	hear	stream	start
12	train	stori	time	start	summer
13	pop	watch	world	bbc	sun
14	commut	night	found	get	version
15	bus	blog	feel	itun	favourit
16	whilst	stuff	heard	phone	beauti
17	drive	check	lot	week	girl
18	hot	bit	decid	even	littl
19	singl	final	band	enjoy	blue
20	danc	late	piec	soundcloud	make

fig16: Words associated with Formats and Technology topic present across all 5 topics.

By isolating from the corpus only those documents containing key words related to formats and technologies but that instead appear in different topics, we can begin to examine the potential conclusions and findings drawn from an interpretation of a corpus-wide computational analysis. Do the assumptions we carry about the technologies of vinyl records and the iPod, for instance, that appear to have been foregrounded and validated by the results of the computational reading facilitated by the LDA process, carry over into a manual

reading of the texts concerned? In other words, how far does Blei's 'statistical lens' take us in terms of theorising about a corpus?

On a practical note here, and with consideration of Piper's focus on modular methods, it is important to show here that because of the manner in which the original dataset was organised according to principles of tidy data, as outlined in Chapter 4, the results of the various computational analyses undertaken in this chapter can be bound to the original corpus. This in turn can facilitate sub-division of that corpus based on the variables generated through processing. Clusters of stories in their original format can thus be extracted and further sub-divided based on the presence of words, which in this case are vinyl and iPod. Despite the issues raised above, then, regarding the clustering of individuals (and other elements) based on often decontexualising combinations of computationally generated numeric and categorical abstractions, there is also a potential benefit. In the case of the Harkive database of stories, original texts can be retained for closer readings that enable the researcher to critique the results and processes of their abstraction. Through the example of vinyl and iPod stories, I demonstrate how this can be achieved. Once again, the functionality of my method is presented here as an auxiliary finding related to the method I have devised being an element of my contribution to the field.

Isolating from the corpus those stories that contain mentions of the word vinyl returned n=139 entries, which represented 1.83% of the total corpus. From the following visualisations we can observe two things. Fig 17 reveals that not only are stories mentioning vinyl present in all topics, but that a similar number (n=40) appear in the 'time, motion and place' topic to which appear in the 'emotion and experience' topic (n=33) that the LDA process assigned the word vinyl to. Further to that, Fig18 suggests that there are higher and more varied sentiment scores and stronger topic associations to be found in the 'contributing to harkive', 'listening' and 'formats and technologies' topics than in the 'time, motion and place' or 'emotions and experience' topics. We may ask ourselves, then, what is going on here because, almost immediately, these results appear to challenge the conclusions suggested by the corpus wide analysis discussed in the

previous section. We can see that vinyl is discussed in many other topics than the one we have interpreted as emotional and experiential, and that stories in other topics have an apparent relation to positive sentiment that exists outside of emotional and experiential accounts. To what extent, then, do stories containing the word vinyl allocated to particular topics exhibit the characteristics our interpretation has assigned to those topics. A closer reading of the 139 stories containing the word vinyl taken in terms of their allocation to each of the 5 LDA topics reveals some interesting results.

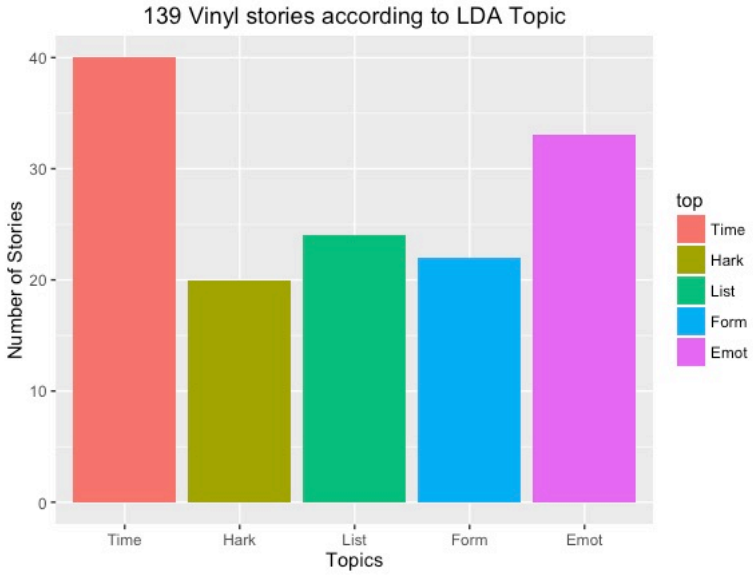

Fig 17– LDA topic allocation of stories containing the word vinyl

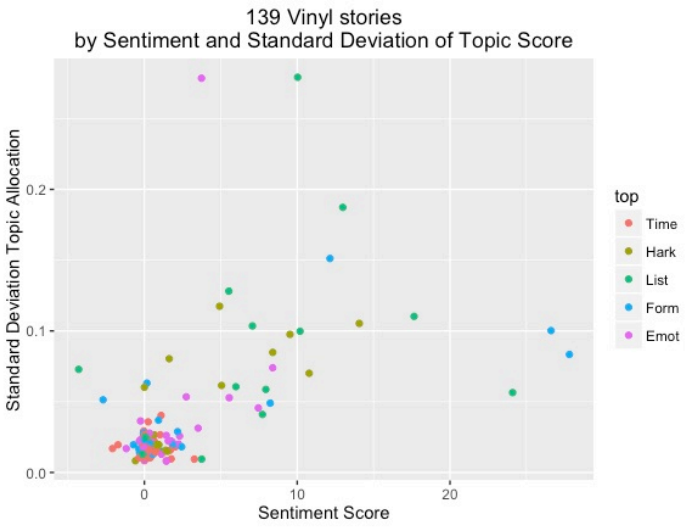

Fig 18: Sentiment scores and SD of topic allocation of stories containing the word vinyl, colour-coded by topic allocation

Of the (n=40) vinyl stories contained within the 'time, motion and place' topic, 20% (n =8) contain only information about what has been played. These are stories where the respondent mentions only the name of an artist or record they were listening to at the time, and have mentioned they were using vinyl, but do not provide any additional context. Of the remaining 32 stories, 69% (n=22) explicitly reference listening whilst working, travelling or in domestic spaces, and here we see vinyl records being used in mundane, everyday situations, including ironing school uniforms and hanging out washing. Considering first of all the topic allocation, the activities accompanying vinyl listening here are emblematic of what we may expect, yet appear to challenge the conclusion suggested by the corpus-level analysis.

> Quick coffee before next set of jobs / chores. Ornette Coleman - Twins, vinyl (#295)[11]

> Decided the Marvin's too uptempo for the heat. Playing 'One On One' by Bob James & Earl Klugh on vinyl LP, hanging out the washing (#7040)

> Working from home today so will mainly be listening to vinyl (#5764)

> Ironing school uniforms "Toussaint" by Allen Toussaint on vinyl and what a cracker this is (#2377)

In the emotions and experience topic, meanwhile, we can perhaps begin to see support for the conclusion suggested by the corpus-level analysis. Although, as in the 'time, motion and place' topic, some stories (33% (n=11)) contain only information about what has been played and provide no further context, of the remaining 22 stories, however, 45% (n=10) make specific reference to the aesthetics and physicality of the vinyl record as an object. These features appear to play a central role in its descriptions of its use, as evidenced by the examples below:

[11] All quotes and extracts from Harkive stories presented in this and subsequent findings chapters and will be identified only by their unique story number allocated by the collection process outlined in chapter 4.

> I really want to get this Bosconi Stallions set on vinyl as it comes in a box that looks like this [photo attached] (#3180)
>
> It's my Vertigo copy of Autobahn, I love this sleeve almost as much as the record (#330)
>
> On to vinyl (must remember to take it off at the end, can't risk warping!) (#7643)
>
> I listen to music in all formats, CD, Cassette, MP3 but my favourite is still vinyl. I just love the ceremony of playing a record (#4159)

27% (n=6) stories make explicit reference to vinyl in terms of memory ("I'd forgotten how good this sounded!"). In the first example below, a longer-form story about playing a record, the respondent recalls going to a Bruce Springsteen show. Meanwhile, 50% (n=11) of the stories in this topic also discussed listening to vinyl in other experiential terms, referencing physical and emotional reactions to the music they heard on vinyl.

> I remember being lent Born To Run, The River, and Greetings From Asbury Park by our very musically wise youth leader with a "you MUST listen to these", and I did, and loved them too. But for some reason I stopped listening to Springsteen at all a few years after that, my vinyl got put away in a cupboard. (#2155)
>
> Mrs R out again so loud vinyl time. This one really shakes the walls (#1364)
>
> Darn That Dream' from 'Undercurrent' by Bill Evans & Jim Hall on vinyl. Music that doesn't break a sweat whilst the rest of us do (#7039)
>
> Little Richard & His Band Pt 2 on 7" #vinyl (London 1957). Imagine hearing Tutti Frutti for the 1st time. (#71)

In the formats & technology topic cluster, which we may recall contained almost all the other common methods through which people listen (i.e. radio, Spotify, iTunes, and so on), we find 22 stories, 9% (n=2) that contain only information about what has been played and provide no further context. Of the remaining 20 stories, however, 75% (n=15) discuss vinyl within the context of other technologies, services and formats. Here we can see much more evidence of the

type of “fractured and heterogeneous” listening suggested by Nowak (2014), where meaning is derived not so much from the properties, affordances or perceptions of a particular format, but rather from the “circuit of practices” Maguadda (2011) describes that together combine to constitute cultural practices around the reception of music. Vinyl listening is variously informed by use of streaming services, or else discussed in terms of the download codes that now accompany new vinyl releases.

> I guess I make most of my musical discoveries in the evenings and then do further investigation/research during the day via Spotify (before deciding to either download the album from eMusic or buy on CD/vinyl) (#5822)
>
> I got their new album on LP from my local indie vinyl place last week and the handy digital download means the songs live in all my devices immediately (#5560)
>
> Finally streaming "Electric" by Pet Shop Boys after ten annoying minutes of having to download Pandora on a new phone and resetting my password. If it's good I'm pre ordering the vinyl. 11:34 Well that was fantastic. Aside from the ads (#162)

Taken together this brief analysis of the small number of stories containing the word vinyl demonstrates that this particular format can be seen as being described by respondents in many other ways than the initial corpus-wide computational analysis and subsequent clustering and visualisation suggested. As well as evidence of vinyl providing the types of visceral, aesthetic and emotional experiences that Barthamski and Woodward suggest are defining factors of vinyl use, we can also see it being used as a background accompaniment to more mundane, everyday activities – something perhaps more commonly associated with digital technologies (see, for example, Bull, De Nora, Sterne (ibid)). We also see vinyl being used as part of more complex listener practices that make use of different formats and technologies as part of individualised cultural practices. An interesting thing to observe also is the apparent absence of any qualifying statements regarding the ‘better’ or ‘more authentic’ qualities of vinyl listening in comparison to other formats and technologies. Again, this is something we may have expected to see based

particularly on commercial rhetorics of vinyl use in comparison to other modes of engagement. The idea that metaphors of technology use are replicated and maintained through dialogues between consumers and producers is something that will be explored in more detail in Chapter 6.

Turning our attention now to stories from the Harkive corpus that contain mentions of the iPod, a similar extraction process to that described above with stories mentioning vinyl returns n=202 entries from the dataset that contain the word iPod. The following two visualisations reveal some interesting results, particularly when compared with those produced through extraction of vinyl stories in the previous section. Fig 19 shows that, unlike the more even manner in which vinyl stories were distributed across topics, here we can see that 54% (n = 110) of iPod stories are contained within the topic of 'time, motion and place', which we can recall that the LDA topic modelling process situated the word iPod within. The remaining four topics each contribute comparatively fewer numbers of stories. The second observation we can make, based on Fig 20, is that compared to stories about vinyl, iPod stories appear to have both higher sentiment scores and much more variability. Indeed, the mean sentiment score for iPod stories (2.71) is higher than that of vinyl stories (2.24), and the standard deviation from those means is also higher in iPod stories (5.46) than for vinyl stories (4.92). In addition to this, the standard deviation between topic allocation scores is also slightly larger for iPod stories (0.039) than for vinyl stories (0.0035). Together this suggests that iPod stories are more strongly aligned with the topic they are allocated to than vinyl stories, an observation supported by the fact that more than half of iPod stories appear in the 'time, motion and place' topic. Based on our observations of the vinyl stories we saw in the 'time, motion and place' topic, where vinyl was shown being discussed as an accompaniment to mundane, everyday tasks, we would therefore expect to see this replicated in stories related to the iPod, yet the higher and variable sentiment scores also suggests that we may discover unexpected things.

Indeed, extracting stories from the database of stories containing the word iPod, as with the vinyl stories above, revealed some interesting insights that perhaps

challenge the results and observations that may have initially been suggested from the corpus wide analysis facilitated by topic modelling.

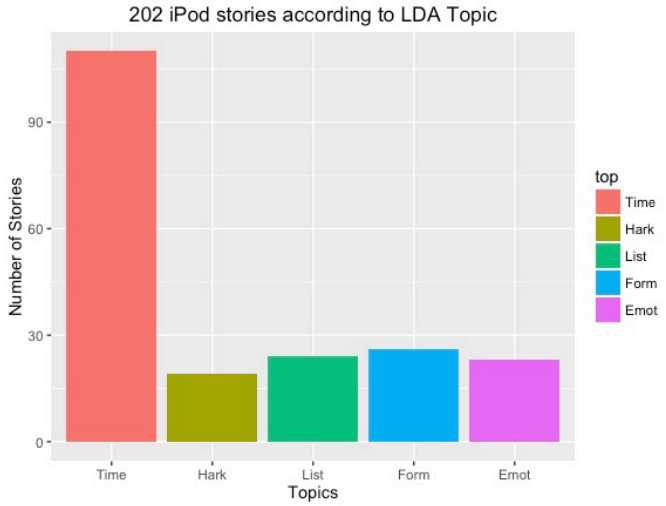

Fig 19: LDA topic allocation of stories containing the word iPod

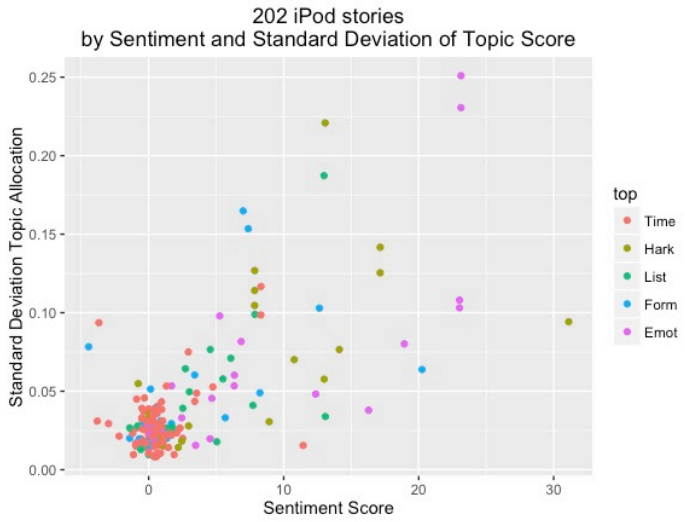

Fig 20: Sentiment scores and SD of topic allocation of stories containing the word iPod, colour-coded by topic allocation

In the examples below we can see the 'functional' iPod closely associated with experiences that can be linked to ideas around identity and authenticity. In the first, the iPod is intrinsically linked to the respondent's identity as a fan of Hardcore music, and acts as a proxy for the lack of his experience of a 'real'

scene, as yet undiscovered in a new city. Meanwhile the second respondent describes an experience with music that could be interpreted as emotional, visceral even, and one in which the iPod plays a central role in what is clearly a thrilling reaction to music.

> I moved here from London and haven't been able to find a proper scene for hardcore music, my ipod classic has me covered there. I spend a lot of time walking around the city or on the metro and I always have my headphones in listening to something…hopefully I'll manage to find some people who share my taste out here but no luck so far! (#2956)

> So there I was walking through central London in the heat and sun, singing along..as I turned onto my street, "What In The World" began. I was so overcome by the awesomeness of the guitar tone and the melody that I rather loudly said "This is fucking awesome!" (#1088)

The examples below, meanwhile, provide some insight into more complex layers of engagement with audio fidelity that exist across digital devices. We can observe instances where earlier generations of iPods have been retained and chosen ahead of newer devices that provide similar functions. This is because of a perceived increase in audio quality afforded by the older models. Once again, this may lead us to challenge assumptions around how people engage with digital devices and files, and we can recall here Sterne's observation regarding the inherently 'distracted' nature of listening that digital files assume and exploit. Respondents here are clearly paying attention, at least in terms of audio fidelity, and this further suggests a relationship with digital modes of listening that goes beyond the merely functional level suggested by the corpus-level results.

> Listening to the new Morrissey album using a 'classic' iPod (much better sound than most mobile phones) on the train to work.(#925)

> Listening to Hubba Bubba by Damaged Bug on my iPod classic now, I think it sounds better than my phone so I use it much more (#1202)

The picture, then, for both vinyl an iPod use, is more complex than the results of computational analysis may lead us toward. This short section has demonstrated that a combination of computational analysis and a close reading of cultural texts provide us with a means to question certain assumptions that accompany these particular formats. For instance, the idea that both vinyl records or the iPod produce or enable particular, generalisable modes of engagement with music. The section has also, and perhaps more interestingly, shown that computational analyses when used uncritically and unreflexively could generate results that may help reinforce or otherwise appear to offer support to existing positions, when in actuality the detail beneath the abstraction suggests the opposite. A key reflexive observation enabled by the analysis undertaken in this chapter, then, is that computational processes have both numerous benefits and limitations when used in humanities research and with cultural texts. The responsibility for ascertaining the difference between these two extremes falls to the critical researcher, who is required to make careful decisions at each stage of a modular research process. Just as music listeners and the various agencies involved with the kinds of digital monitoring are involved in a co-production of sorts, a similar thing can be said regarding the researcher and their tools. This too can be considered in terms of De Certeau's ongoing dialogue between strategic place and tactical space. The results of empirical algorithmic processing being at various times complicated, maintained, challenged, replicated and dismantled by the rationales of the human researcher. In both cases, the area of interest is the very location of the tension; in other words, it is in the unfolding detail of the negotiation process.

5.7 – Discussion

This chapter has demonstrated how a small selection of different computational analytical process can be used in an unsupervised, exploratory analysis of a large, text-based dataset describing cultural practices associated with music in everyday situations. It thus provides the first of my answers to the central research question of this thesis: what can an analysis of the data generated by The Harkive Project reveal about the music reception practices of respondents?

It has shown that an unsupervised approach can reveal broad trends and patterns within a text-based corpus in a manner that is highly efficient, but only once the human analyst involved in the process has deployed a degree of knowledge about the corpus in interpreting the results. More importantly, I have shown that although the results from an analysis of this kind can produce numbers that on the surface appear unequivocal – validated by the use of complex, robust mathematical processes – a closer examination of the data (in this case, the stories) behind such claims to efficacy has revealed that differences are often marginal and that the discovery and understanding of nuance within cultural texts requires more than recourse to patterns, clusters and correlations. Although these processes are efficient tools for exploratory analysis, they require the closer, grounded attention of Berry's (2011) "regulating force of philosophy" before defendable positions can be taken.

Despite these important caveats, however, the patterns revealed by the automated processes used here were demonstrated as useful in helping point towards potentially fruitful lines of enquiry that could be facilitated by recourse to further use of computational techniques, such as the automated extraction of elements based on keyword searches, data visualisations, and other types of exploration. It was, however, in the minute detail of the individual cases that the truly interesting detail was found. This leads to the consideration of further issues about both the content of the dataset itself and the processes that have facilitated the clustering and segmentation.

The findings in this chapter primarily demonstrate how the computational processes used can invite us to simultaneously reinforce and challenge widely held popular assumptions regarding attitudes to particular formats and technologies. Vinyl records are often positioned as a more valid or authentic mode of listening, particularly when placed in opposition to the less legitimate facsimile of digital devices and services. This analysis has shown that vinyl is also often used as an accompaniment to mundane activities, acting as background noise to work or other everyday chores. Conversely, digital modes of listening

have been shown to provide respondents with routes to deeply emotional experiences.

In this chapter I have taken what Chapter 3 demonstrated as being the approach suggested by Liu (2016), Kitchin (2014) and Piper (2016), and through the work above argue that a practical understanding of technologies related to data can help popular music scholars gain not only a better understanding of the role and function of the data-related technologies, but also provide them with a method for (carefully!) approaching popular music texts in their own work. As Chapter 3 also demonstrated, approaches associated with the newly emerging disciplines of cultural analytics and digital humanities are attempting to gain exactly such an understanding by putting the technologies in question to use in the service of practice-led academic enquiry. This allows for the possibility of the techniques in question, including exploratory data visualisations, clustering, and natural language processing, allied to the kinds of data collection techniques detailed in Chapter 4 that facilitate such activity, to be examined as both modular and distinct research tools, and also in their use in combination as "not transmitters but rather producers of sociality" (van Dijck, 2013: 57).

The aim here has been to examine through practice what is a large part of the problem with attempting to understand complex data-related systems. This problem can be located in the combination of their relative novelty and their technical/commercial opacity (see: Ananny, 2015). Most of the powerful entities operating within the digital space of the commercial music industries, for instance, are less than twenty years old (Wikström, 2013) and the technologies they are using, particularly those associated with ideas of data-derived knowledge creation, are even younger. As such the digital space is not fixed in structure and remains highly contested commercially, as I showed in chapter 1, which renders the agencies concerned (in terms of both the companies involved and the technologies they use (see: Webster et al, 2016) as perpetually fluid. This results in a volatile and developing online space that means the internal systems behind many of the services people use on a daily basis are not only changing at great speed (see: Amatriain, 2013), but are also unavailable for scrutiny since the

particulars of their operation are valuable trade secrets. These difficulties have been discussed in terms of what Boyd and Crawford (2012) call a divide between those who have access to not only data but also the knowledge and resources required to process it, and those who do not.

Taken as a whole, this makes studying the role, benefits and potential consequences of digital and data technologies extremely difficult, and particularly for humanities researchers who may not necessarily possess the technical skills or other resources (see: Sandvig and Hargittai, 2015; Savage and Burrows, 2007) that such an enquiry requires. The work of De Certeau (1984), discussed in Chapter 2, has been shown as instructive in terms of providing an overall framework for the practical engagement I have devised. It is a model that can help us understand instead what can be seen as the ongoing negotiation and dialogue between strategic aims of governance (e.g. record labels, streaming services, interfaces, etc.) and the tactical activities of listeners engaged in their individual forays into Felski's (1999: 5) "more exotic and esoteric worlds" through the reception of music and other cultural content. Harkive, as discussed in Chapters 3 and 4, is conceived of as both a place/space where respondents are simultaneously "the informed, the informant, and the information" (Michael and Lupton, 2015) and is thus able to usefully replicate and explore many of the (technical) conditions under which contemporary modes of music reception can and do occur. Where many digital interfaces and platforms are defined by the integrated "capture, analysis and output" of data (Rieder, 2016), Harkive – as a modular method – attempts to reflexively examine each of these processes in turn. Chapter 4 discussed the rationale behind Harkive's data capturing and analytical processes, this chapter has explored ideas and consequences of a specific forms analysis of the information those process facilitate[12].

The relationships between the marketplace, technologies and cultural practices of music, how these are changing, and the issues that are raised by the ongoing

[12] My post-doctoral work will focus on developing Harkive into exactly the integrated process Rieder describes, and I am particularly interested in exploring the possibility for real-time analysis of data. At present, my analyses are based around off-line, batch processing of data. The principles behind both, however, are largely similar. Each takes data and produces insight. The difference being that only with real-time data analysis can online interfaces change dynamically and thus alter or influence the type of data they are able to collect.

dialogue between them, is the wider focus of my longer term intellectual project and will inform much of activity I hope to undertake in post-doctoral work. Within the limited scope of this research project, however, the aim is to explore the means by which I can build and improve upon our understanding of those relationships and changes through the Harkive project in its present form. This chapter demonstrates, then, through providing answers to my central research question through recourse to computational techniques and machine learning, how I have made a contribution to contemporary debates by developing a new and innovate means by which cultural data can be collected, sorted and analysed. In doing so I make a contribution towards ameliorating the disconnection between the present skills sets of many cultural studies and popular music studies researchers (myself included) and the technologies and systems that I have speculated should now become an important (if not central) focus of our collective work. By deploying computational techniques often involving (to greater or lesser extents) 'black box' algorithmic analysis – complex mathematical calculations that produce results about large and complex cultural data sets – I have demonstrated that analyses of this kind can be effectively and reflexively achieved by popular music scholars.

CHAPTER 6

Streaming music services, playlists and discovery

6.1 – Overview

In this chapter, and towards answering my central research question of what an analysis of the data generated by The Harkive Project can reveal about the music reception practices of respondents, I focus on examining one particular element of music reception through an analysis of Harkive respondents' use of music streaming services. There has been a growing body of work recently that explores the use of streaming services (see, for example, Hagen, 2016; Prey, 2015; Webster et al., 2016) and here I demonstrate how the methods of data collection and analysis I have developed help contribute to those debates by providing new insights. Through the analysis in this chapter I show a number of things related to the use of streaming services: I show that despite the efforts of streaming services to engender a sense of brand loyalty amongst consumers, respondents here appear to reject aligning themselves with particular streaming brands and instead view streaming as primarily a new, additional mode of music reception that augments existing listening practices. In other words, although streaming service interfaces can be viewed as strategic places (De Certeau, 1984; Mejia, 2012) that users inhabit as spaces through their various individual and collective tactical listening activities, I show that users are better understood at this time as visitors rather than residents (White and Le Cornu, 2011) of the strategic places of streaming.

Respondents are also shown to switch between streaming brands throughout the course of their everyday listening and appear much more inclined to align themselves with media brands (e.g. newspapers, magazines, radio stations) or else musical genres and artists, that together can be understood to pre-date digital technologies. In this sense streaming services instead function

primarily as what Markham (1998) refers to as "tools" and are as yet not embraced in terms of Markham's conception of a more engaged mode of "ways of being". There is also very little evidence from the Harkive corpus that streaming services have been successful to date in what Morris and Powers (2015:5) describe as their attempts to "acculturate [users] to and find reciprocity within a service's musical ethos" as instead users incorporate streaming into existing practices and attitudes. Morris and Powers also argue that streaming services are "pushing not only..specific qualities [of their individual services], but also the benefits of streaming music more generally" (2015:7). Based on the findings of this chapter it cannot be argued that the rapid growth in streaming as mode of music reception is entirely attributable to that project.

The analysis here is based on a combination of a distant reading (Moretti, 2013) afforded by a computational analysis, and a close reading of texts foregrounded by that process that the previous chapter demonstrated as a useful means of analysis. As the previous chapter demonstrated, my modular method is able to reveal broad trends within the large dataset of 7582 individual Harkive stories, and through this I will isolate small clusters of respondents and their stories in order to perform a close reading of their texts to test and explore claims made by researchers examining the use of and attitudes towards streaming music services as a particular facet of contemporary practices of music reception. The chapter proceeds in three inter-related parts. Building on the work of Hagen, Puschman & Burgress, Lakoff & Johnson, and Markham (2016; 2008; 2003, 1998; 2014), I turn my attention first to music streaming as an activity broadly conceived by examining the metaphors used by Harkive respondents in describing their experiences. I then look at two particular aspects that are key to the manner in which streaming services are framed by the companies operating them by examining questions around playlists and ownership through the work of Morris & Powers and Sinclair & Tinson (2015; 2017), before looking at ideas and questions around discovery and recommendation with particular reference to the recent work of Nowak (2016).

This examination allows me to show that despite their growing use, streaming services do not in and of themselves lead to expressions of greater/better levels of consumption in terms of discovering and recommendation of new music for Harkive respondents. It is instead pre-existing routes to discovery and recommendation, along with ideas of ownership that are played out within digital environments. I also demonstrate that an analysis of respondents' stories shows very little evidence of users' explicitly voicing concerns over activities related to data collection, privacy or algorithmic curation in their stories. This apparent lack of concern from respondents is in opposition to the concerns theorized by academics, including those who look at popular music (Morris, 2015; Prey, 2015; Webster et al., 2016), and it would appear that in the context of music reception as reported by Harkive respondents who are regular users of streaming services, the growing role and potential consequences of digital monitoring is not something that overtly concerns them at this point in time. The presence, however, of quantitative data that suggests otherwise, gathered by the survey element of my data collection in response to specific questions regarding data privacy and related concerns, suggests that such concerns may be worth exploring further.

Overall this chapter demonstrates that the methods of data collection and analysis developed by this project are able to provide useful routes to insight, particularly when operating at scale, but that there is considerably more work to be done if we are to fully understand the changing nature of the environments and cultures around the reception of music in light of recently emerging digital technologies, something I will explore in more detail in chapter 7 and 8.

6.2 – Additional notes on method

As shown in Fig 21 below, the Harkive dataset over the lifetime of the project thus far demonstrates some correlation with figures reported by the RIAA

and BPI discussed in Chapters 2 and 3. For example, the recent growth in commercial revenues for vinyl records, along with the downturn in revenues for downloads, is reflected here. In particular, the dataset appears to correlate with the recent growth in use of one particular service, Spotify, and shows evidence of respondent discourse around activities related to playlists.

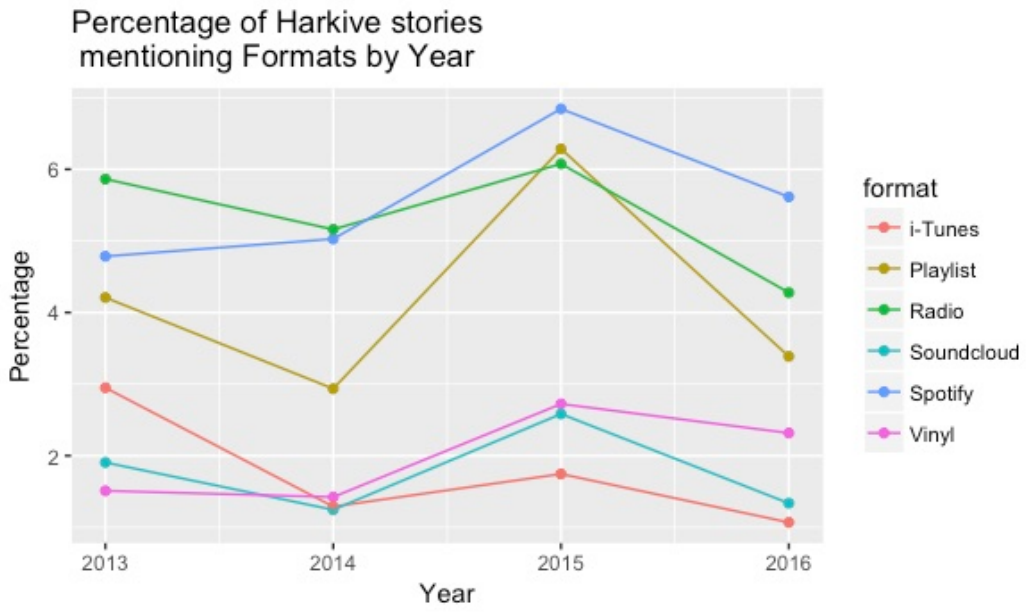

Fig21: Percentage of Harkive stories containing mentions of specific formats and technologies over the period 2013-2016.

This type of exploratory computational analysis developed through approaches outlined in Chapters 4 and 5 suggests that the contents of the Harkive dataset may provide useful insight into the manner in which people describe their use of Spotify in particular, and of streaming and playlists more generally. Allied to this data from the story-gathering element of data collection, Chapter 4 also demonstrated that respondents were invited to complete the Harkive music listening survey. This generated quantitative data that has been used to provide additional context to the analysis in this chapter. Based on a combination of these two datasets, clusters of respondents that may help reveal insights specific to streaming use have been isolated for further, close analysis based on a process that segments the corpus and survey datasets into a focussed sample in the following manner:

- 122 (6.34%) of the 1922 respondents to the story-gathering element of the project also completed the Harkive music listening survey. Between them they provided 1750 (23.1%) of the total 7582 stories[1].

- Of the 122 respondents to the story-gathering element who also completed the survey, 58 (47.54%) identified as regular users of streaming services, indicating that they used streaming services *Often*, *Very Often*, or *Daily*. Between these 58[2] respondents, 1268 stories were provided, which equates to 16.72% of the total corpus.

- The corpus segment derived from those identifying as regular streaming users was then automatically searched for mentions of words and services foregrounded by the topic modelling analysis undertaken in chapter 5 which I have interpreted as being associated with streaming activity[3]. In so doing the 1268 stories from 122 respondents was further reduced to 203 (2.69% of total corpus), from a total of 47 respondents (2.44% of total).

- This segment of the corpus and survey datasets is on the one hand a significant reduction of the entire corpus of Harkive data, but can be understood to contain data specific to the aims of this chapter in the following ways: contextual information derived from survey responses; where respondents indicated they are regular users of

[1] 23% of the total story dataset originating from just 6% of respondents can be explained by the fact that those using platforms such as Twitter are able to post more than one story across the course of the day. By using the survey data to subset the corpus based on respondents identifying as regular streaming user, I am able to extract all of their entries, and not just those that mention words associated with streaming.

[2] Of the **58** respondents, **22** (37.93%) identified as female, **35** (60.34%) as male, and **1** (1.72%) indicated they would rather not say. In terms of age, **46** respondents (79.31%) were aged between 31 and 50, 12% (n=**7**) were under 30, and the remaining 8.62% (n=**5**) over 50. A total of **46** (79.31%) respondents came from the UK, **3** (5.17%) from the USA, **2** (3.44%) from Ireland, and the remaining 7 respondents (12.06%) were made up of one person from each of the following countries: Bulgaria; Canada, China; India; New Zealand, Poland; Spain[2].

[3] List of words used for search took into account automated text processing, stems words to roots to avoid repetition, as discussed in previous chapters. The words are presented here in the stemmed form with their 'true' value in brackets: **spotif** (Spotify); **deezer** (Deezer); pandora (Pandora); **applemu** (AppleMusic); **soundcloud** (Soundcloud); **data** (Data); **algor** (Algorithm); **privacy** (Private, Privacy, etc); **automat** (Automatic, Automated, etc); **curat** (Curated, Curation, etc); **recomm** (Recommendation, Recommended, etc); **discov** (Discovery, Discover, etc); **playlist** (Playling, Playlisted, etc); **stream** (Stream, Streaming, etc)

streaming services; and which contain the use of specific words or phrases commonly associated with streaming in their stories.

The process of segmentation and reduction described above thus informs the analysis that follows and serves as a demonstration that the larger Harkive corpus and the methods of data collection and analysis I have developed have the potential to be usefully deployed in numerous different analyses related to respondents' reception of music. The process above could, for example, be used to isolate discourses around vinyl records, or radio listening, and honed further by segmentation based on survey responses, the presence of certain words, and also the results of the unsupervised machine learning analysis performed in chapter 5. For the purposes of this chapter, however, I will examine respondents' stories according to the three lines of enquiry outlined in the introduction, proceeding from an examination of metaphorical descriptions of streaming activity, through to specific aspects of streaming related to playlists and ownership, and then discovery and recommendation.

6.3 – The metaphors of streaming

In her recent exploration of streaming listeners Hagen (2016) builds on the work of Lakoff and Johnson (2008) and Markham (1998) to argue that "shared, implicit frameworks of meaning allow experiences to become more widely available (explicable, even familiar) to others"[4]. This allows me to speculate that an exploration of the manner in which Harkive respondents use metaphors in their descriptions of their experiences with streaming services can usefully lead us to a greater understanding of streaming experiences more generally. Further to that, if "metaphorical understanding is..partially culturally determined and dependent upon past individual experiences" then exploring, as Hagen goes on to suggest, the metaphorical descriptions of activity related to streaming through the wider lens of other cultural practices of music reception (i.e. ones that also involve radio, physical

[4] Hagen's article appeared online on the First Monday website and so no page numbers are attributed to her quotes used here. The full article is available here: http://firstmonday.org/ojs/index.php/fm/article/view/6005/5277

formats, and so on) may be similarly useful. This is particularly so if we consider Sterne's notion of mediality (2012:9), which he uses to describe the manner in which the form and content of technologies often refer to one another in complex ways. We may therefore expect to see "native" descriptions of streaming activity, i.e. those which Morris and Powers suggest are "birthed in an era of digitalisation" (2015:2), where metaphors are directly related to new streaming technologies, and at the same time descriptions that recall earlier, or other modes of music reception, including what Morris and Power call "pre-digital norms of distinction.. (e.g. genres or taste cultures)" (2015:2). To illustrate this point we can see an example of the latter in the following excerpts:

> Dug out (if you can "dig out" on Spotify) Feist's The Reminder for the first time in ages. It remains a perfect summer record (#7445)
>
> Kicking off #Harkive with a 2nd spin of Torres via Spotify on my phone.. (#1033)

The respondents here call upon exploratory metaphors closely associated with vinyl records, that of 'crate-digging' or the rotation of vinyl records, to describe the manner in which they have chosen or are playing dematerialised recordings from the Spotify catalogue. Simultaneously, however, the respondent in one story questions the notion of whether such an activity is possible through a digital interface, where the tactility of physically searching through a collection is necessarily abstracted. Despite the transposition of activity over media forms, however, we can observe that the recording remains central – 'ideal' in one example – and is the focus of the familiar experience that Hagan describes. Although the form the recording takes is relatively new, it is primarily a familiarity with and affection towards musical content in recorded form and to an extent existing practices of listening that predate streaming, which remain intact and provide anchoring functions for music reception within streaming services. These observations were largely borne out across the rest of the sample. Although 68.96% of respondents (n =40) reported that the experience of listening via physical media (i.e. vinyl/CD) differs from listening via digital formats (streams, mp3s),

respondents were much more equally split on whether they felt physical formats were better than digital formats: 25.86% (n = 15) disagreed with that statement; the same amount neither agreed nor disagreed; and the remaining 48.27% (n = 28) agreed to varying extents. In addition the majority of respondents (62.06%, n = 36) agreed to varying extents that whether the format of listening was digital or physical had no bearing on their enjoyment, and a further 20.68% (n = 12) neither agreed nor disagreed with that statement. Only 17.24% (n = 10) indicated that differences between physical and digital modes of listening impacted on their enjoyment, with the majority of these (n=9) agreeing only somewhat with that statement. The survey indicates here, then, that the mode of listening is a less important factor than the act of listening itself, and that respondents are (for the most part) transposing "pre-digital norms of distinction" (Morris & Powers, 2015:2) and existing listening practices across to digital and streaming modes.

The observations above are particularly interesting when we consider Morris and Powers' argument that "[streaming] services sell branded musical experiences, inviting consumers to see themselves and their attitudes, habits and sentiments about music reflected by the service they choose to adopt." (2015:2) Given the attempts of streaming services to differentiate from each other on experiential grounds through interfacial features, as discussed in chapter 2, we may expect to see some alignment from respondents with particular services based on a perceived similarity within interfaces with respondents' existing, pre-digital practices. However, within the 203 Harkive stories from regular streaming users, there are instead numerous examples of streaming services being used primarily in terms of what Markham (1998) describes as "tools", where streaming services instead facilitate listening based on respondents' alignment with other music-related brands and artists, or else activities that reside outside of or else predate the streaming space. According to Markham, the "tool" function is very much the first level along a sliding scale of more meaningful and engaged use of Internet technologies, one that can then proceed to the incrementally more engaged states of "place" (which can be understood in similar terms to De Certeau's space) or the final

stage of “ways of being”, where ideas of tool and place fall away and the user ultimately enfolds the technology concerned into their everyday routine. There is little or no evidence here, however, that a particular service has to date successfully achieved their aim, as it has been proposed by Morris and Powers, of developing a close relationship with users’ preferred or habitual modes of listening practices. Instead, and as mentioned above, several respondents are shown to align themselves closely with existing media outlets, rather than streaming services, in their descriptions of streaming use:

> Listening to the Pitchfork "Best New Tracks" playlist on Spotify (#888)
>
> ..prior to this NME-embedded excellent Waxahatchee Soundcloud stream, heard Waves by Glass Gang via Gold Flake Paint blog (#4206)
>
> The bonus material from the CODY reissue (via guardianmusic stream) is sounding great. Looking forward to the boxed set. (#4695)

In the above examples it is media outlets (Pitchfork, NME, a music blog) that for respondents provide the focus and impetus for listening, with the mode of listening (streaming) or a particular service (Soundcloud, Spotify) reduced to the role of a tool through which that listening is facilitated. Meanwhile, the following respondent is led by both the content of media outlets and their own tastes through a period of listening during the day that takes in a series of streaming services, all used primarily as “tools” and only occasionally as “places”:

> Pitchfork streaming. First heard these guys [London Grammar] featured on a Disclosure track and instantly loved them but couldn't find any of their material. Heading over to their Soundcloud now to listen to more…After having a quick look at Throats' Last.fm page, I learn that some of the members have a new band called Exhaustion, so I head over to their Bandcamp to check it out. It's very raw. Too raw it turns out. I go to Consequence of Sound to stream Nicolas Jaar's remix of Daft Punk's Random Access Memories as I need something a little less intrusive so I can concentrate on work. The only annoying thing about streaming an album from Soundcloud like this is I can't pause it easily if someone interrupts to talk to me. Find myself

wishing it was on Spotify (#6040)

Ultimately here, the respondent mentions what they perceive to be the greater affordances of Spotify over that of Soundcloud, but not before mentioning their use of an array of different streaming services (Bandcamp, Last.fm), dextrously switching between them depending on the availability of catalogue and the affordances and limitations of each of the services concerned. Once again, the streaming services are "tools" used to facilitate listening, and the habits and preferences (the "ways of being") primarily exist outside of the streaming platforms as "places". In a similar way the first respondent below views Spotify listening as a proxy for their physical collection of music and listening modes closely associated with that collection, in a manner that recalls Markham's later conception of prosthesis (Markham, 2003:5), where Internet technologies are used to access "information which would otherwise be out of practical or physical range of motion"(2003:3)

> I mostly use Spotify for streaming full albums in order – either stuff I want to listen to before deciding whether to buy, or as a convenient way of listening, at work, to records that I already own but don't want to bring with me. Rarely use playlists.
> (# 4654)

> Work listening is [a] combination: Soundcloud, Mixcloud, Spotify, YouTube, Bandcamp & iPlayer. Streaming wins; rarely physical media. (#7540)

The examples above demonstrate adoption of streaming as a mode of listening that is not yet characterised by a sense of belonging or loyalty towards a particular service. Although 72.41% (n =42) of survey respondents indicate that digital music technologies (streaming services, MP3 stores, etc) enhance their enjoyment of music, only 25.86% (n =15) indicate that they like to try digital services when they are first launched, and only 20.68% (n = 12) feel it is important that they are using the most up-to-date services and devices. It is the existing and familiar modes of reception that seem to hold sway, with streaming services being used to facilitate those, rather than act as catalysts to new or different modes. Finally, the three examples below suggest

the polar opposite of positive alignment with any given streaming service. The first refers to a 'well known streaming website' put seemingly pointedly does not name it, the second feels the need to qualify, or else somehow excuse their use of Spotify, while a third suggests they have recently adopted Spotify's premium service but have done so with a degree of reluctance.

> Time for a bit of #BestDebutAlbums preparation with The Texas Campfire Tapes by Michelle Shocked via a well known streaming website (#1026)
>
> Tunes in the car – had to be Let's Eat Grandma this morning (streaming from Spotify, I know...) (#7687)
>
> Newish premium Spotify user (I caved, but it will never be complete solution!) (#880)

This apparent reluctance or inability of respondents to emotionally align with streaming services can be partially explained by the relative infancy of streaming as a technology and practice. A fondness for music, listening and related activities of music reception, all of which can be understood as predating streaming technologies, and certainly for the majority of listeners in this sample given the predominance of those in the 30-50 age range (see footnote 2), would perhaps explain the primacy of past experience over any alignment with new streaming brands. Indeed, the examples above show regular users of streaming services in terms of what White and Le Cordu (2011) would describe as visitors rather than residents of the services concerned and indeed Hagan makes this observation in her work looking at streaming users (2016). We can thus see streaming users as largely service agnostic, regularly jumping between different services throughout the day and, moreover, streaming more broadly defined as primarily a mode of listening that is used alongside and within the pre-existing contexts of other modes of listening despite the centrality and dominance suggested by commercial figures discussed in chapter 1 (see: BPI, 2017).

The evidence here speaks to the task that Morris & Powers suggest remains for those operating in the streaming space, where "services must interweave

the newness [of streaming]..with norms, demographics and predilections that predate streaming" (2015:9), which is a similar point to that made by Spotify's Matthew Ogle in chapter 2. However, there is little evidence that these norms and predilections are being carried over by Harkive respondents into discreet practices related to particular services. Instead it is existing practices that are transposed to streaming as a whole via a process that recalls Sterne's mediality. Given the relative novelty and complexity of the streaming space, which itself further complicates understandings of listener practice (Nowak, 2014), we can view the dexterous moves of respondents between streaming services in terms of their attempting to exert degrees of control over their engagement with music more widely conceived. Rather than expecting to see alignment, then, with a particular service, it is the dexterous jumps from one to another that are perhaps better understood in terms of Markham's "ways of being", in the sense that a degree of instability and restlessness are the conditions that are becoming commonplace, everyday and "domesticated" (Baym, 2010), rather than a particular streaming brand. We may wish to consider this further in terms of ideas of ownership, which the next section focuses upon.

6.4 – Playlists and ownership

Morris and Powers (2015) demonstrate that the word stream or streaming has historically implied "freedom and bounty as well as limits and constraints" (2015:2), and argue that the metaphor of streaming as it relates to popular music "obfuscates how these new services inhibit the circulatory flow of music, altering longstanding notions of property and rights associated with the music commodity" (2015:2). Related to this, and within the context of Belk's (2014) idea of a post-ownership economy, Sinclair and Tinson (2017) have examined streaming from a marketing studies perspective and attempt to address a gap in the related literature in terms of psychological ownership, which they demonstrate as being previously linked to increased brand loyalty and a willingness to purchase. This can be considered in terms of the perceived aims of streaming services in terms of customer retention

and acquisition suggested by Morris and Powers. Sinclair and Tinson argue that streaming users "organise their music consumption, manage and project their identity and establish a sense of control in their everyday routine" (2017:2), all of which are also in evidence within the Harkive data and which can be further understood through Markham's notion of digital technologies relating to "ways of being". The manner in which the respondents above switched between services can be seen as ways of organising consumption in the face of numerous new channels, and evidence of presentation of the self can be demonstrated in the very act of contributing to Harkive, particularly in terms of the vast majority who did so via a public forums such as Twitter. There is also very strong evidence even within the small, segmented data set of stories, of mood management, a key facet of Markham's idea of "ways of being", particularly where streaming is used in work and other everyday situations, as in the following indicative examples:

> Too hot and unsettled at work, need music on to concentrate. Soundcloud faves on shuffle, guarantees songs I love and throws up some surprises. Emperor Zero – Heart of Iron, class tune, Hookworms, haven't heard this for ages, so good, Autobahn covering Iceage, can't wait for the album. (#3506)

> After a blast of Esquivel I've now gone classical to Janacek piano (still on Spotify), suits maudlin mood of paperwork perfectly (#6555)

> Thankfully I managed to have a proper lunch hour today and dragged myself to the gym for a long overdue bit of exercise. I took my iPhone with me and listened to a playlist of new music on shuffle as I plodded along on the cross-trainer, listening to a playlist of new albums via Spotify on my phone (#1033)

> I woke up here in London and began my commute on the bus with the usual NPR downloads from the US yesterday on my phone. There was a very boring interview with Liev Schreiber and a review of the American version of The Bridge so I switched to Spotify and put my starred songs on random play while I read an interview with Malcolm McLaren's widow on my Kindle. I began with 'S Club Party'. This moved into 'Hung Up' by Madonna which I thought would mix nicely with 'Love Is A Bourgeois Concept' on the new Pet Shop Boys album. Next came on the Designer Drugs remix of 'Untouched' by The Veronicas as I passed Russell Square. I skipped over Fleetwood Mac and The Woodentops but let 'I

> Don't Care Anymore' by Doris Duke play because it is so depressing ("I'm all alone on a lumpy bed") that I thought it could only make my day seem amazing by comparison. Skipped The Cramps and Philip Glass and settled on 'I'll Remember' by Madonna as I came across Waterloo Bridge and got off the bus at the National Theatre before walking along the South Bank (#3988)

The responses above demonstrate how the use of streaming services in work, commuting and exercise situations are used in similar ways to those identified by De Nora's (2000) work around music in everyday life, and also in terms of Walkmans and later iPods which Bull (2006, 2000) demonstrated were able to "fine [tune] the relationship between mood, volition, music and the environment in ways that previous generations of mobile sound technologies were unable to do" (2006:136). What is different about streaming services in these situations, and as discussed in chapter 2 in relation to Mejia's work around mICTs (2012), is the additional layer of complexity in terms of content and apparent choice that comes with the connectivity of these devices. In the examples above, music (and other content) is variously: drawn from numerous sources and services on demand; enabled by the possibility of creating playlists and favourites for later use; by switching between a variety of songs with ease; or else of following recommendations from numerous, similarly connected sources. Sinclair and Tinson observe that it is precisely this additional functionality that sets streaming apart from the iPod and its antecedents and "leads to both greater control and integration of music into the everyday routine" (2017:9), a further step in the direction of Kassabian's idea of "ubiquitous music" (2013), an idea to which I will also return and expand upon in Chapter 7. Meanwhile, Sinclair and Tinson go on to argue that this additional element of control "over mood and to an extent the spaces [of listening]..can facilitate strong feelings of psychological ownership for music streaming applications that were perhaps deemed to be diminished in the digital age of music consumption." (2017:6) However, Morris and Power make the point also that control is slowly being relinquished by users as musical content is no longer owned, but rather accessed via connected devices, and indeed that the cloud-

based location of the material puts collections broadly conceived in a risky position due to the volatile nature of licensing issues, another idea I will expand on in the next chapter in related to Sterne's contention that a consequence of digital formats is that "collections may no longer outlive their owners" (2009:58). Alongside the affordances in terms of control highlighted by Sinclair and Tinson, then, we should consider the limitations and constraints of streaming services, particularly given the rapid moves from one service to another that we saw earlier.

If control is a contested and volatile area, and instability is a driver for our "ways of being", then there may be other routes towards users building familiar relationships with streaming that can help us understand the widespread use of services (even if service loyalty is not yet present). Turning once again to the work of Hagen, who found that, "practices of gathering music into streaming playlists regarded as curatorial activity are clearly associated with principles of collection" (Hagen, 2015), we can perhaps begin to see other ways where elements of the new streaming experience find footing in familiar practices. Indeed there is evidence from the survey responses that regular streaming users consider playlists to be part of their wider music collections. 81.03% (n =47)[5] of respondents consider online playlists to be part of their wider music collections, which supports Hagan's claim, but within that there is also evidence of a divide in terms of emotional attachment to digital formats that sets it apart from the positive connotations with which we associate physical collections (see Straw, 1997b). 39.65% of those respondents (n =23) report feeling an emotional connection to digital collections, whilst for 48.27% (n = 28) there was no emotional connection[6], despite considering digital formats as part of their wider collections. However, from respondents on either side of this divide, we see that there are more similarities in their descriptions of activity related to streaming and playlists than there are differences, and in fact a commonality across both is the use of streaming in work and other everyday contexts that we saw above.

[5] Q71: My online playlists (e.g. in streaming services) are part of my wider music collection
[6] Q72: I feel emotionally connected to my online/digital music collection in the same way as I do with my physical (CD, vinyl) collection

Morris & Powers argue that "digital music services no longer sell discrete musical objects, nor do they focus exclusively on content offerings. Instead, services sell branded musical experiences, inviting consumers to see themselves and their attitudes, habits and sentiments about music reflected by the service they choose to adopt" (2015:2). Much of the rhetoric surrounding this is encapsulated in the ways in which services present themselves to consumers. We can recall the Spotify press release regarding the launch of the Discover Weekly Service in July 2015 discussed in chapter 2 ("It's like having your best friend make you a personalized mixtape"[7]), and consider this in terms of Morris and Powers' observation that "streaming..is not just a technical form of transmission, but a key metaphor for the flow of information in the digital age"(2015:2). One key differential feature deployed by services is that of automated recommendation and curation, which through processes of digital monitoring can be understood as a way in which services can reflect users tastes, and thus increase the possibility for feelings of psychological ownership. In the final section of this chapter I will turn my attention to ideas of discovery and recommendation as expressed by respondents.

6.5 – Discovery and recommendation

Given the observations of Morris and Powers and Sinclair and Tinson above about notions of belonging, control and ownership, and considering that streaming services are metaphorically framing the negotiation of their huge catalogues in terms of personalised curation as a form of "chivalrous" (Straw, 1997a) guidance, an examination of discovery and recommendation as it occurs and is framed within streaming may help us further understand streaming as a mode of music reception. Datta et al (2016) argue that "consumer adoption of streaming leads to substantial increases in quantity and variety of consumption, and in discovery of new music. Relative to music ownership where experimentation is expensive, repeat listening increases for

[7] Quote taken from the original 2015 press release circulated online by Spotify when they launched Discover Weekly. Full text available here: https://press.spotify.com/li/2015/07/20/introducing-discover-weekly-your-ultimate-personalised-playlist/

consumers' best new discoveries" (2016:1). Although not perfectly mapped to Datta et al's work, the survey data here[8] reveals that attitudes towards variety and volume of listening would support such a claim. 67.24% (n =39) indicate that they listen to more music because of digital technologies, and 91.37% (n=53) report that this variety increases with their use of digital technologies. A key adjunct to this, however, is that many in fact report that much of their discovery of new music is achieved by listeners themselves, with 63.79% reporting that they find it easy to discover new music without assistance.[9] This highlights what Kjus (2016) describes as "key discrepancies [that] arise between the promise and the reality of streamed-music discovery, both for artists seeking new fans (and funds) and for audiences expecting streaming to supersede existing forms of musical exploration" (2016:2). Coates (2016), meanwhile, has suggested that the present efficacy of automated recommendation is skewed towards well-known artists in a manner that replicates existing models of radio play, which may shed light on why respondents here appear to prefer to perform much of their exploratory work themselves.

Indeed, the above speaks to an argument made by Nowak (2016), that although the act of discovering new music is widely acknowledged be an essential component of popular music culture, something discussed in Chapter 2 in my attempt to update the work of Straw in light of streaming services and automated curation, questions of how people discover and what a discovery is remained poorly theorized. He offers a potentially useful route towards a better understanding of discovery by challenging the often "taken-for-granted association[s] either with the social positioning of consumers or with the increasing array of material agencies at their disposal" (2016:137), which can be understood as comprising one of the central tenets of algorithmic curation when the results of data collection about taste are deployed as curatorial mechanisms through service interfaces. In other

[8] Q57: I listen to more music now than in the past because of digital technologies) & Q59: Having access to more music, and more information, means I listen to a wider variety of music than I did previously.
[9] Q83 I tend to stick with music that I already know & Q84 I find it easy to discover new music that I like without assistance.

words, Nowak argues that it is thus either a case of structure or agency, or the ongoing dialogue between the two, that defines our thinking about discovery and suggests this is replicated in the commercial models of streaming services. We can, of course, consider this also in terms of much older questions regarding music reception that date back to Adorno's critique. We might expect then, if we combine this with Hagan's suggestion that producers' metaphors are often repeated by users, to find examples within the Harkive corpus of either socially derived recommendation, discovery through agency afforded by technology, or combinations of the two. This is indeed the case.

The following examples show how respondents utilise digital technologies within the interfaces of streaming services to find useful recommendations and routes to discovery. Within these descriptions we can see evidence that such practices display elements of the habitual, suggesting they are becoming incorporated and embedded into everyday practice. Spotify's Discovery Weekly function has a 'good hit rate' and is 'on fire' in two examples, and 'awaits' in another, whilst the observation about the efficacy of Soundcloud's interface for discovering new music is suggestive of regular use by the respondent.

> Pretty good hit rate this week from Discover – Inner City Blues by Sly Dunbar is a nice slice of dubby poppy reggae. (#6825)
>
> It's worth adding that once you're on Soundcloud, from whatever source, it's easy to spend a while on there discovering new music (#1253)
>
> Settling in to the work day. Spotify Discover playlist awaits. Kicking off with 'Door of the Cosmos' by Sun Ra (#6828)
>
> Discover Weekly on fire this week. World Gone Deaf by Bill Baird. Catchy, Beck-y indie. (#6818)

We can however see more examples of different digital interfaces being used in combination rather than in loyal isolation, where the affordances of one feeds directly into the use of another in a manner that recalls my earlier point regarding the ordinariness of instability as a "way of being". BBC Radio's

online interface 'Playlister' is a good example here, discussed by a respondent in an extract below in terms of providing a bridge between different technologies of discovery that allows the respondent to hear new music on radio broadcasts and then save them to their own Spotify playlist for later use. Likewise, the Spotify-native apps of media outlets such as Pitchfork inform discovery that can be further explored within Spotify[10]. In other examples, discovery occurs through listening to podcasts and online mixes. Once again, this speaks to the observation made above regarding the primacy of "tools"-based use of streaming services, where users are visitors rather than residents of streaming places, and where Markham's "ways of being" are characterised primarily by the dexterous and repeated moves from one place to another. It is through this that users, to recall De Certeau, act tactically and carve out a space that is tolerable. In much the same way that streaming services often facilitate rather than guide listening, the same can be observed regarding discovery. The tactical moments of discovery below – in other words the space being carved out – are occurring within digital interfaces not closely associated with streaming, but rather the online places of more traditional forms of cultural intermediation.

> So many discoveries to be found via JDTwitch RinseFM mixes. This, from Graham Philip D'Ancey, is ace: [LINK] (#146)
>
> The list is heavily influenced by what I've heard on BBC 6 Music recently, particularly on Mary Anne Hobbs and Gideon Coe's shows. I use the BBC Playlister a lot. If I hear a good track I can save it to a playlist on the BBC site and then export it to Spotify. I have about 70 songs from this year so far. If I then really like a track I will download from amazon mp3 or iTunes. This playlist is on my iPod so is the stuff that bore enough repeated plays on Spotify for me to buy the tracks, or the relevant albums (#5147)
>
> Now listening to Liverpool band Strange Collective and their new EP Super Touchy. As recommended by theriderpodcast (#6823)
>
> The 'New stuff' list is usually populated from within Spotify these

[10] Third-party apps from companies such as The Guardian and Blue Note Records presented ways of engaging with Spotify catalogue via interactive visual displays. These provided additional content and context (album reviews, sleeve notes, visual timelines, etc.) within the Spotify interface that could be viewed as the music played, or else as a means by which to navigate through catalogue. Spotify eventually phased-out native apps in February 2015, although maintained a relationship with MusixMatch in order to provide users with lyrics to songs. – see: http://www.musicbusinessworldwide.com/spotify-introduces-lyrics-but-phases-out-third-party-apps-on-desktop/

> days, via Spotify's New Releases page and a couple of apps – any Decent Music and Pitchfork. (#1062)

There are also numerous examples of technological affordances meeting social milieus as routes towards discovery, and particularly through Twitter. Often these will relate to media organisations using Twitter as a means of communicating with audiences, such as in the 'website links to articles' or the Tweet from a record company examples below, but predominantly these relate to specific recommendations that come from others within a respondents' personal network. This speaks again to the work of Hagen and Markham, who suggest that digital environments are places that people enter and make themselves available to others within the conditions of ubiquitous Internet connectivity that can provide ways of engaging with music that have almost become expected as part of our everyday life. In other words, it is through the constant moves from one interface to another, which are mundane and everyday and closely linked to ways of being, that respondents act tactically and create their own or collective spaces. Once again, however, it is worth noting that this tactical activity is not linked solely to the use of music streaming services, but rather includes social media, email and online conversations, which help facilitate music listening through the streaming services that themselves remain in the role of "tools". This seems at odds with Hagan's observation that streaming services are mediators rather than intermediaries, that shape rather than facilitate, as the evidence here would suggest the opposite.

> I find a lot of music via Twitter these days. Websites tweet links to articles, friends post recommendations, all on one site (#1251)

> #harkive first up: cover of David Bowie's Young Americans by The Cure via YouTube, recommended by [Twitter name] and [Twitter name] (#4216)

> I discover a lot of new music on Twitter, simply by following people who are enthusiastic and knowledgable about music. One such is the writer [name], who has just tweeted a link to Swamp Dogg's 'My Hang-Ups Ain't Hung Up No More', an extraordinary 1974 southern soul track about going to a shrink! I immediately

> head for Discogs and added it to my wants list… I click on another tweet, this time one of Domino records inviting me to watch the new video by Matthew E White for his song 'Vision'. Oh, it's gorgeous. I must get his album. I loved his first. (#3823)
>
> Sometimes recommendations creep in and, when they do, it's via Twitter. For example, [name] mentioned that Jungle's debut is available so I added that to the list. That went to the 'Good stuff 2014' list too. I also listened to Royksopp & Robyns 'Do It Again' and Owen Pallett 'In Conflict'. Work was interrupted by lunch (watched a saved video on Vimeo – no music) and a bus down to the Southbank Centre – talky podcasts while walking, Mr Fine Wine's Downtown Soulville while reading articles. Meeting done and back to Owen Pallett and then some S. Carey before a tweet from [name] mentioned Gotan Project, which made me think of St Germain (possibly also influenced by an email asking me to proofread a short French translation). I queued up a few tracks from Tourist, which was played everywhere when I lived there many years ago. (#1062)
>
> ..finally got the headphones on playing some music via PC while working: mix of YouTube and SoundCloud tunes recommended via Twitter (#4214)

A final observation to be made regarding the intersection of technological affordances and sociality is the role of the Harkive Project itself in terms of recommendation and discovery, which was a surprising and unexpected factor. Many people contributed to the project in online forums such as Twitter, and several reported enjoying reading the stories of others. This led to conversations and discoveries that were a direct result of the project taking place, and as such Harkive itself became another online, strategic place within which respondents acted tactically.

> If I learn nothing else from #Harkive I have at least discovered Clout's Substitute thanks to [name of another Harkive respondent] (#4061)
>
> [@Reply to another Harkive Respondent] Missed the news about Moebius myself. Sad days for experimental music, but an opportunity for the world to discover it! (#6537)
>
> This is the third year I've participated in the harkive project. I like the feeling of being part of a larger music community, all of us

> connected across data points across the whole world. I used to feel this way about napster when I found myself chatting with users scattered across the globe, discovering and sharing music (#907)

There are numerous examples above that demonstrate Nowak's observation that discourses of discovery are often rooted in ideas around technological discovery and social milieus, but it is questionable whether the responses of Harkive respondents here are perpetuating conceptual metaphors of discovery posited by companies operating in the streaming space. Indeed, there are also examples or outright rejections of the language and efficacy of streaming service recommendation which variously suggest that items suggested are either not aligned with listener tastes or preferred modes of listening, are variable in their efficacy, or else are the result of serendipity rather than any listener-perceived technical prowess of automated recommendation. The survey responses provide useful additional context here. Only 13.79% (n=8) of respondents state that online/automatic recommendations are often better than those they receive from friends, and only 29.31% (n=17) are surprised by how accurately online music service recommendations reflect their tastes.[11]

> Discover Weekly is now playing "I Believe in Miracles" for the second time this week. It pops up most weeks. It's discovered. (#6817)

> On to wider Discover tab in Spotify 'Charlene' by Psychic Mirrors. Not for me. 80s style plodding R&B with annoying vocals. (#6824)

> Enjoying Spotify's Discover, though I'm not a playlist kinda gal. Does feel like I will discover new bands (#1248)

> Zammuto session for KEXP…Courtesy of YouTube mailer which is usually 90% pish but recommends one good video (#4671)

> A YouTube recommendation I agree with! Earlier Gershwin has led to Bronski Beat's Ain't Necessarily So #serendipity (#468)

Rejecting dominant notions of technological and social conceptions of

[11] Q81: The online/automatic recommendations are often better than those I receive from friends & Q85 I am surprised by how accurately online music service recommendations reflect my taste

discovery, Nowak (2016) instead argues that it is the affective responses to music – whether discovered through social connections or technological agency – that are important and "entangles much more than the question of the origins of such discoveries" (2016:142). A discovery has to be memorable, then, argues Nowak, or it would not be reported as such, and theorizes that it is epiphanies and rediscoveries that can allow us to differentiate from music that we simply have access to via our social milieu or which comes to us via technological affordances (or combinations of the two). It is not so much the point at which discovery occurs that is important, then, but "the interaction that leaves an affective mark on individuals" (2016:143) which he argues is the site of and indeed makes up the actual discovery. This is an interesting idea to pursue, and indeed there are examples from regular streaming service users of exactly these epiphanies and rediscoveries. 50% (n = 29) of survey respondents indicate that music is important to them when they wish to reminisce about something, or someone, and 79.31% (n=46) report often recalling or remembering a song and seeking it out to play it[12]. In the examples below it is respondents' memories and previous discoveries that inform music listening via streaming services

> "Pale Blue Eyes"– The Velvet Underground. I vividly remember on the night I heard that Lou Reed had died, I laid in my room with no light but a candle and played this album aloud. I discovered the Velvets soon after I moved to New York for the first time, when I was eighteen, and their music has been a constant ever since. Sure, some of it is nasty and dark and dirty, but there's an undercurrent of gentleness there, perhaps all the more rewarding because it runs so contrary to his provocative image (#4554)

> This morning, as I was finishing my breakfast, I got the urge to listen to Alex Chilton by The Replacements. I've only recently begun listening to them in a serious way, and I don't possess much of their work, so it was off to YouTube. In the sidebar, I of course found Can't Hardly Wait, which so far is my favourite song of theirs. The building of anticipation that happens during the line "I'll be home when I'm sleeping" is such a wonderfully

[12] Q48: Importance of music when I reminisce (remember someone/somewhere) & Q33: I listen when I remember a song and seek it out to play it

> pure romantic moment. After that, I wanted to hear Teenage Fanclub's 'Songs From Northern Britain', so I opened up iTunes and listened to the entire thing. I discovered that album last September, and it's shepherded me through a lot over the past few months. I've even occasionally thought that, if I ever manage to hoodwink some poor innocent man into becoming my husband, I'd play 'Planets' at our wedding reception, for us to dance to. (What can I say; I'm at an age where people I know are starting to get married.) (#4541)

> When I got home I remembered I'd seen on Twitter earlier in the day one of my favourite DJs, Kutmah, share a link to a mix he's recorded to promote the Low End Theory Festival in LA this weekend. Low End Theory is a clubnight held every Wednesday in Los Angeles and has played a pivotal role in the development of the 'beats' scene, where people like Flying Lotus and Daedelus emerged. I first came across Kutmah, a Brit who moved to LA aged 12, a few years ago when he was arrested and threatened with deportation from the States, so there was a campaign on social media to support him and through that I found his mixes. So I didn't want to sleep on this mix, which he'd recorded for Wire magazine, and promptly put it on my iPad through my monitor speakers while I cooked my post-gym dinner and tried not to sweat to death on the hottest day of the year (#7255)

To conclude this discussion of ideas of discovery I will consider something that is almost entirely absent from the database of stories from regular streaming users. As has been discussed in earlier chapters, and particularly in the case of Spotify, the mechanics of automated recommendation within interfaces are driven by activities of data collection, digital monitoring, and algorithmic processing. As the opening chapters demonstrated, the potential influence of activity of this kind, where services operate what Hartmann et al call data-derived business models (2014), is the subject of considerable attention and concern from academics in numerous fields. It is interesting to note, then, that apart from the occasional complaint about the efficacy of automated curation services, the database of stories does not contain any evidence of concern regarding issues of privacy or data collection techniques. Future research may wish to consider why this is the case. The absence, however, of respondent discourse around these issues is particularly interesting when we consider that the survey responses do provide some

evidence of concern[13]. Of the respondents to the survey who indicated they were regular users of streaming services, 94.82% (n =55) indicated that they were aware that data can be collected by companies/organisations about them based on their online activity (e.g web browsing, music listening, social media posts), and 67.24% (n =39) indicated that they agreed data collected about them was a necessary part of the exchange involved when using certain online services, but only 51.72% (n =30) indicated that the data they generated about themselves when browsing, purchasing or listening online was given over voluntarily. Indeed, 53.44% (n= 31) felt that data collected by companies/organisations about their online activity was an invasion of their privacy, and only 17.24% (n =10) exhibited trust that companies and organisations collecting data about them would keep it secure. This awareness and degree of concern over digital monitoring takes on an interesting additional dimension when we consider that 81.03% (n =47) would like to know more about how the data collected about them is used, and 67.24% (n =39) would like to have access to the data collected about them. Interestingly, and in consideration of Boyd and Crawford's conception of data "haves" and "have nots" discussed in earlier chapters, 53.44% (n = 31) felt that they possessed sufficient technical knowledge to be able to make sense of the data collected about them were they to be given access to it. These final responses suggest that projects involving elements of citizen science, where 'ordinary' users may engage with data, particularly that which is generated about their own activities, may be fruitful routes towards further work in this area. This is an idea I will return to in Chapter 7 in a discussion about potential avenues for future research. In the meantime, however, it would be useful to place the observations above within the wider context of

[13] Q61: I am aware that data can be collected by companies/organisations about my online activity (e.g web browsing, music listening, social media posts), Q67: The data collected about me is a necessary part of the exchange involved when I use certain online services, Q74: The data I generate about myself (e.g. when browsing, purchasing or listening online) is given over voluntarily, Q75: I trust that companies and organisations that collect data about me will keep it secure Q62: Data collection by companies/organisations about my online activity (e.g. web browsing, social media activity, music listening) is an invasion of my privacy, Q66: I would like to know more about how the data that is collected about me is used, Q65: I would like to have access to the data collected about me by companies and organisations, Q69: Were I to be given access to it, I feel I have sufficient technical knowledge to make sense of the data collected about me

how streaming is becoming an important means of engaging with music in everyday life.

6.6 – Discussion

In this chapter I have demonstrated how the two data sets collected by Harkive are able to provide useful and unique insight into the practices of people who identify as regular users of music streaming services. Using first a series of automated analysis techniques that enabled a distant reading of text and survey responses, I was then able to use elements of that data to reveal broad trends in the datasets that facilitated segmentation based on survey responses and text mining, ultimately isolating a particular group of respondents from the larger dataset who could be empirically understood as likely subjects for an exploration of streaming listening as an emerging practice.

The chapter went on to demonstrate that, just as the business models and practices of companies operating in the streaming space are still in a state of flux and uncertainly, streaming as both a mode of listening and a cultural practice is similarly unstable. In part I suggested that this is due to the relative infancy of streaming, which is still in the process of being incorporated into wider, existing practices and through this, via Markham, posited that it is paradoxically the very same instability that forms the basis of streaming services' present incorporation as "tools" and occasionally "places" into respondents' wider "ways of being" as they relate to everyday music reception. Despite the understandable desire of companies operating in the streaming space to "acculturate [users] to and find reciprocity within a [given] service's musical ethos" (Morris and Powers, 2015:5), the findings of this chapter suggest that Harkive respondents are not yet prepared to make such a leap or commitment. Streaming as a mode of listening is instead spread widely across a number of different services and thus any attempts to map listening in this way on to 'older' forms of listening, where consumers may own one record player, or may tune to a particular radio station, begin to

break down. Streaming as a practice, then, is something listeners are still auditioning and experimenting with and display a level of uncertainty towards that is in contrast to recently published figures suggesting that streaming has become the dominant mode of music reception. The revenue figures derived from streaming may indeed suggest that it is the dominant mode of consumption, and indeed the broad trends revealed by my data would offer some support to this, but within the detail of individual respondents' there is also considerable evidence that streaming is merely but one of a number of ways in which people listen.

The work carried out in this chapter differs in scale and application from existing and emerging work that looks at music streaming. Rather than carefully assembling groups of subjects based on demographic or other identifiers, this study instead draws from a larger body of respondents and has sought to use exploratory computational methods that are closely linked to commercial practices, combined with a close engagement with texts more commonly associated with humanities work. It is not my position that this way of drawing together subjects is better than other methods, but I do suggest that my mode of operation more closely resembles the practices of companies and organisations operating in the commercial online space at this present time, with Harkive operating as simultaneously an online, strategic place that is able to capture evidence of practices related to tactical movements of space creation. Such a route in the particular instance of this chapter leads to observations about streaming as a practice, on the one hand, but also – and perhaps more importantly – to a further process of reflection and discussion regarding the manner in which data collection and computational analysis leads to a form of knowledge production that is emerging, increasingly influential but, as yet, not adequately understood.

CHAPTER 7

Digital technologies & the cultural practices of music reception

I have previously argued that the activities Harkive stories describe can together be viewed as the tactical movements (De Certeau, 1984; Mejia, 2012) of respondents as they carve out their own spaces through their uses of music and technology in the strategic places of contemporary everyday life. As a researcher, operating within the strategic place of the academic process, and within the confines of a findings chapter within a thesis aiming to answer the specific research question of what an analysis of the data generated by The Harkive Project can reveal about the music reception practices of respondents, here I follow a similar rationale to both De Certeau's walkers and to the Harkive respondents and – by 'walking' through a close engagement with the many narratives contained – attempt to carve out a space of my own in which to think through the many issues of debate and method I have engaged in through my research. I first explore the issue of how the reductive and productive processes of data collection and computational analysis are able to deal with the complexity of everyday life, whilst simultaneously showing that the use of technologies by Harkive respondents that facilitate precisely such analysis are inspiring new cultural practices related to engagement with music. From this I argue that new means and models of understanding the activities and outcomes of music reception should be explored in order for people to be able to make sense of both this process, and the conditions by which their practices are increasingly monitored and guided.

7.1 – Walking with Harkive respondents

Broadly, the issues I have engaged with have related to the use of data collection and computational analytical techniques as a means of understanding the

complex realities of music reception in everyday life. In this chapter, and specifically through examining the various ways that Harkive respondents engage with those technologies, I explore how we are collectively coming to terms with the tensions that exist between reductive/productive processes of computational systems and the sheer complexity and abundance of detail that comprise our everyday lives[1]. Through this I speculate that the challenge facing popular music scholars today is to think about how we may arrive at an understanding of the tensions between digital abstractions, data-derived knowledge, and the ways in which we all live, think and be in the world. This chapter demonstrates that within the detail of Harkive stories, there are suggestions towards how we may do that.

A key finding of this chapter is that a consideration of the type of data-derived abstractions produced by and through the use of digital and Internet technologies is never far from respondents' minds. In this chapter respondents are shown to exhibit a degree of familiarity and communication with new digital technologies and their abstracted, digital selves, and that this in turn is helping to form intriguing new cultural practices. This allows me to speculate that acknowledging these cultural practices offers a route towards what popular music scholars may wish to focus on in the coming years.

I begin by highlighting the inherent difficulty of studying music reception and everyday life, and specifically the difficulties of attempting to arrive at that understanding through the use of data-derived computational analytical techniques. These techniques are, on the one hand, able to capture a greater degree of detail than has previously been possible about peoples' music reception activities, but on the other they are unable to account for the complexity of that engagement. By showing how the use of various technologies of music reception can help reveal elements of that complexity, I demonstrate that the uniqueness of the cultural practices Harkive respondents engage in

[1] As Feyerabend observed, "the world we inhabit is abundant beyond our wildest imagination. Only a tiny fraction of this abundance affects our minds. This is a blessing, not a drawback. A super-conscious organism would not be super-wise, it would be paralysed." (2001:6). It was a consideration of this quote alongside the conceptualisation of respondents (and researchers) as De Certeau's walkers, that provided the springboard for the approach taken in this chapter.

during their everyday lives can be understood in terms of De Certeau's walkers, with respondents engaging in a form of negotiation with the structural spaces of contemporary music reception. This enables me to suggest ways in which this negotiation is a productive route towards understanding contemporary conditions, and primarily because of the ways in which new technologies are altering everyday routines and routes to meaning.

I show that it is a combination of always-on connectivity, engagement with digital interfaces, and an awareness of practices related to data-collection and analysis, that are providing the impetus for those alterations to routine and meaning-making for Harkive respondents. These new elements are closely inter-related, which complicates our understanding of them. The route I suggest is one that considers the respondents' activities with and within the new digital environments of listening as being based on an engagement with a process, rather than with particular digital objects and their consequences. In other words, respondents are shown to be developing their own epistemic responses to digital, data and Internet technologies. Through this I suggest that the intriguing challenge for popular music scholars is to develop new means and methods that may enable us to look at the possibilities for practices such as data collection and computational analyses to become further enfolded into the everyday routines of listeners, which in turn could help inspire new cultural practices and routes to meaning. Ultimately this chapter suggests that by combining the computational work I have developed with an consideration of these new cultural practices, there is the potential for popular music scholars to help develop, in a manner suggested by Piper (2016) in chapter 3, new reciprocal relationships between culture and computation that may be beneficial. Specifically this means developing new means and methods that can enable people engaging with music through digital technologies to develop further and more reflexive epistemic responses to the changes in routine and meaning that these technologies are helping to produce. Towards that end, this final chapter explores the following, inter-related issues of debate: what are Harkive respondents' everyday relationships to the new technologies associated

with music reception, and can an understanding of this inform the future work of popular music scholars?

In order to achieve this, however, we will 'walk' with the Harkive respondents through their everyday engagement with the digital, data and Internet technologies of music reception.

7.2 – Listening technologies and the everyday

As a starting point we can consider how an ordinary (every)day on the "wildly, unimaginably abundant earth" (Feyerabend, 2001, p. 6) begins for Harkive respondents. It is from here that we can 'walk' with them as their relationships to music, technology, and the world around them, variously unfold. It is here, also, where the problem of how we can understand contemporary music reception practices is foregrounded and can be considered.

> 6.30am The sun is shining in Rome and I'm out for my morning run. I could never go for a run without music playing in my ears. My iPod seems to know me very well - the first song that comes on shuffle is 'Scorpio Rising' by Death in Vegas. Top tune. Reminds me of the old days...A band I used to tour with played it every single night minutes before getting on stage. It was their backstage ritual. (#7501)

> As per usual the day starts groggily, with my alarm blaring incessantly for the umpteenth time that morning. It's 8am. After a frantic 30 minutes getting ready for work I dash out the door to walk to work, shove my over-ear headphones on (horses for courses) and fire up the ol' iPhone. Given the sunny weather it's definitely not the sort of day for serious music, nor songs with a sad tinge. In fact I'm decidedly indecisive this particular morning so rather than pour over the seemingly infinite number of songs on the ol' phone, it was time to live dangerously and put it on shuffle. Disclosure - When A Fire Starts To Burn. The beat sets the pace of my journey to work. (#2953)

> 9:40am - I listen to an album called 'Zentropy' by Frankie Cosmos on my walk to work using my iPhone, Spotify and Apple headphones. This album is super short so it ends before I arrive - to fill the last couple of minutes I listen to 'Paper Bag' by Fiona Apple twice through. I haven't really listened to Fiona Apple properly but this

> song makes me feel like I should be Ally McBeal or something. (#2919)
>
> 10am - I am walking around the supermarket, doing some shopping. The radio is playing, but I am not really aware of what I am listening to as I concentrate on my shopping. Then, a song I half-recognise comes on. I am sure that I have heard it before, and it makes me feel nostalgic for something, but as I pause in the aisle I cannot think of where I had heard it, or who the song was by, so I shrug it off and continue shopping. (#655)

The extracts above are taken from four different Harkive stories collected between 2013 and 2016, from respondents in different countries. They provide a brief snapshot of how an ordinary day begins that would be familiar to many people, and particularly those for whom music listening and the technologies associated with that activity are part of their everyday experience. The extracts also point towards the difficulties of studying such a thing. On the one hand we can observe the experience of waking and preparing for the day that is common to many people, but on the other we can see that the detail of the everyday differs in subtle but important ways from person to person. Some respondents start their day with exercise while others begin theirs 'groggily'. Some head to a place of work for the common start times of the 9-to-5 day, another heads to the shops. Some choose the music they listen to, others hand off this task to mobile devices and digital interfaces. The devices we see being used are common to many people, but they are used here in different combinations and scenarios. Music variously 'fills the time' or 'sets the pace', and for others it instead occurs around them, easily shrugged off. It prompts one respondent to feel briefly, and to their apparent amusement, like a TV character. For one person the song they hear leads directly to a specific memory, for another a memory remains elusive. And so on. How do we make sense of this complexity?

Felski's (1999) definition of everyday life is a useful next step, because it serves to highlight the specific difficulty of understanding the relationships between people, music, and digital technologies in Harkive stories. For Felski the everyday is "the essential, taken-for-granted continuum of mundane activities that frames our forays into more esoteric or exotic worlds" (1999:1). The

reception of music – which Negus (1997: 8) defines as "how people receive, interpret and use music as a cultural form while engaging in specific social activities" – is the activity here that plays out alongside both the taken-for-granted and the forays into the esoteric and exotic. Much of everyday life in Harkive stories is ordinary at a macro level, in the sense that the activities we see above from respondents detail familiar ways in which people begin their days with music (waking to music on the radio, for instance, is a common theme of Harkive stories). However, the music reception activities above are also extraordinary in that they are unique at the micro level of the individual. To borrow a natural metaphor associated with the recent technological developments around data technologies (see: Puschmann and Burgess, 2014), those who seek to understand audiences for music by recourse to digital monitoring and data analysis technologies, are faced with both the tantalising promise and the intriguing problem of abundance at the granular level. Interestingly for the purposes of this chapter, we can see from the above that is often as much the choices and uses of technology that mark out this granularity, this uniqueness of cultural practice, as it is any of the other potentially measureable characteristics of the individuals' activities, such as their age, gender, or their choices of particular types of music. One scenario where this is also evident is during the commute to work:

> I get the taste for apathetic punk rock so I listen on my commute. The Internet on my phone drops out on the train so I switch from the EP on Spotify to the album stored on my phone. I've always preferred the production on the EP, can't help but feel a bit disappointed at having to interrupt it (#5804)

> Listening to https://ultimae.bandcamp.com/album/9980 on ferry (yes I commute to work via a boat! Its a one-hour trip but it is indoors). Devices I use are IBasso DX90 as transport (with Flacs on a 128GB micro SD card - I guess about 200 albums), Chord Hugo as DAC / amp, Shure SE846 earphones. This is heavy (I walk 3 miles a day with this lot in my bag!) and costly but the sound is so great it makes it worth it! I spend a lot of time using this gear so the cost is worthwhile to me (#5814)

> The commuting hour. I drive to work and on a Tuesday [and] my toddler is also in the car. This morning we listened to a five track Beat Mark sampler on cassette (yes, my car still has a radio cassette

> player) that the band gave me after a gig on Sunday night. We also listen to a bit of Tim Burgess 'Oh No I Love You More' on my iPhone via the car stereo and together we sing a couple of nursery rhymes (#5335)
>
> Much as I love (and often rely) on last.fm and Spotify to recommend new artists to me, I still like CDs curated by people who know a scene inside out. For my morning commute to work today, it's the 'The Blues Magazine' issue 7 CD. I make a mental note of the tracks I like so I can look up the artists online later today. (#2311)

From the above examples of commutes via different modes of transport, either alone or in the company of others, we can see that uses of different technologies are as individual as the choice of music that accompanies them. What is interesting in terms of next steps is that the technological elements of these stories highlight unique elements of the listeners' choices that simultaneously replicate (in the metaphorical sense described by Hagen (2016) in chapter 6) and challenge the progressive narratives that have emerged around digital technologies of listening. For one respondent the 'curated' CD is superior to the recommendations of online services, for another a legacy in-car technology is used equally to play analogue and digital formats (and is then ignored in favour of the ancient practice of singing). For another respondent the portability and affordability of digital devices is complicated by weighty, expensive additional equipment, even when they know they must carry it with them on a 3-mile walk. Ideas of recommendation, convenience, portability and affordability – in other words the progressive selling points of digital technologies of music reception – are variously accepted, rejected or otherwise negotiated by individuals in their unique scenarios. We can thus see not only that the activity of engaging with music through the use of technology is something that is both "synonymous with the habitual, the ordinary and the mundane" and "strangely elusive, that which resists our understanding and escapes our grasp" (Felski, 1999:2), but also that listeners here are De Certeau's walkers, carving out their own space through the tactical uses of listening technologies. As such we can see that although patterns of technology use are possibly even more difficult to gain a handle on than other means of differentiation (such as songs listened to, or the demographics of the listeners), they may – paradoxically – be of potentially more use in terms of

generating ways of understanding contemporary conditions. The question becomes, then, whether an exploration of the ways in which new cultural practices are emerging alongside digital, data and Internet technologies can arrive help us arrive at a different kind of understanding.

Bijsterveld and van Dijck (2009:16) define cultural practices related to listening as "the ways in which people are used to doing things and commonly attribute meanings to these routines". This is a useful next step, because considering the relationship between routine and meaning is an important way we may approach issues around understanding how music reception now occurs. The subtle changes to routine (understood at both macro and micro levels) that accompany the emergence of practices associated with digital and mobile technologies – for instance, where the choice of commuting music is handed over to a recommendation algorithm – would suggest that not only are ideas of routine being altered, but so also is the manner in which meanings are attributed to those routines. As such, by examining the routine cultural practices that are emerging alongside these new technologies, we may edge closer to a better understanding of contemporary methods of meaning-making. Before we run the risk, however, of over-privileging new technologies, we should consider that Bijsterveld and van Dijck make the point also that now, as in the past, new technologies "inspire" new cultural practices[2] and that the promotional rhetoric of manufacturers "usually advertise[s] new use options which they aim to embed in both well-established and newly imagined cultural practices." (2009:16). In Chapter 6 we saw the manner in which metaphors associated with older listening practices and the rhetoric used by companies associated with newer modes of listening spill over into the discourses around practices linked to streaming services. We can observe a similar process occurring in other modes of contemporary listening, but in ways that are linked to intriguing configurations of uses of new and old technologies in combination that are specific to individual negotiations between routine, use and affordance:

[2] They are referring here specifically to the relationship between cultural practices, audio technologies and memory, but the efficacy of new technologies to 'inspire' (rather than, say, 'lead') does, I think, hold for listening practices more generally conceived.

> We just moved into this awful open space office, and the CEO is in town, and people are all around and won't shut up. What to do? In grade 11 biology we learned about Pavlov. I thought it was interesting, and decided to make a studying playlist so that I would have a Pavlovian response to the music, and be productive. I don't know if it actually works or I just think it works, and therefore it does. So, that's what I'm listening to. The playlist hasn't changed since 2005, and it starts with Tears for Fears' "Everybody Wants to Rule the World" and ends with Wilco's "Jesus, Etc" (#1513)

> My music today is a playlist I built last night of over 300 tracks, selected from my favorite artists in my library. Every track is a track I love. It's like the ultimate personal radio. Worth investing the time to pick them, because it makes me smile constantly. Right now it's 'Wish You Were Here' by Pink Floyd, one of my favorite songs ever. But I've said that 20 times today. (#1102)

> After lunch worked on my computer dealing with e-mail etc. and listened to my music, held on a server (over 800 CDs worth) and played on a Logitech Radio in a random way.. In about an hour of music heard some Blues, some Latin and various old early 60's pop tracks (I've taken to buying collections such as 'The Monument Records Story' with many well known and obscure US pop songs) plus one or two tracks of African music - I have catholic tastes! I'll be listening to more as I work on the PC again this evening - but probably played through a tablet, connected to music centre and playing from 'Google Play Music', which accesses my collection as well - and a very good system it is - makes interesting playlist suggestions too (#6183)

> I find [vinyl records] at car boot sales and charity shops in my little corner of the UK... This year I began making a radio show style cloudcast where I showcase some of these finds, crackles and all. I'm spending Harkive Day at home with a pile of records to listen to, with a view to picking out tracks for the next episode. I love rooting through boxes of old records looking for the next addition to my car-boot collection; it's a great way of discovering new old music and I've gained loads of favourite artists this way. The fact that I can share this with others via Mixcloud makes it even more of a fun hobby (#5126)

In the examples above, digital and internet-enabled technologies are used in everyday, routine contexts in ways that are intriguingly linked to pre-digital practices. Digital playlists are for some static and unchanging, despite the ease with which they can be updated, and function for one respondent as the 'ultimate

personal radio'. For another, Internet connectivity linked to listening is primarily a means by which they can make their physical record collection portable so that it functions as a personal streaming service. For another the idea of the digital playlist is directly linked to 'rooting through' boxes of vinyl before ultimately contributing to their online identity in the form of a shared playlist, broadcast as online audio, and thus acts as bridge between two worlds, between the real and the virtual. We can thus see that contemporary routes to meaning through routine are informed as much by older technologies as they are newly emerging ones, and can make the observation that this is a process that has a much longer lineage and would have likely played out also with the arrival of the record, of radio technologies, or the CD. We may ask, then, what is it specifically about the contemporary conditions of music reception that is new. By examining one particular element of these new cultural practices we can give this question more consideration.

To look at this more closely we can consider in particular the idea of the playlist. A new and distinct cultural practice on the one hand, in the sense that the components of a playlist can now be drawn from comparatively larger pools of songs – i.e. the millions available via streaming services, or the thousands of files contained on computers, MP3 devices, or personal servers – which can in turn facilitate the creation of playlists that can be relatively easily assembled, reconfigured, and quickly and potentially more widely shared. On the other hand, however, these are activities that have their roots and familiarity – the kernels of their routineness, and thus their meaning – in older cultural practices, such as the creation of cassette mix-tapes, or in the curatorial knowledge/display elements of record collecting (see: Straw, 1997b). Contemporary activities with playlists can therefore be understood as both new and familiar, common and unique. We may wish to consider also that, as Bijsterveld and van Dijck point out, the manner in which "users appropriate the audio technologies in their everyday lives is never self-evident" (2009:16). In other words, what results is a form of dialogue between users and manufacturers, which we can understand in terms of De Certeau's model of strategic activities and tactical movement, as evidenced in the following examples:

> My primary method of music listening at the moment is on my laptop. I have two auto-updating playlists, both sorted in alphabetical order by filename, in Rhythmbox. The first of these is the top thousand tracks that haven't been listened to since I got this computer. The second of these, which I play when my wife is around, is the top 600 tracks which haven't been listened to and which aren't by the Beach Boys, Monkees, Jan & Dean, Harry Nilsson, Captain Beefheart, Frank Zappa, or anyone else she's got sick of hearing me listen to (#4426)

> The drive home from work is always a battle through two congested Midlands motorways and a busy high street in central Birmingham so I really need to be alert. This means a soundtrack in the car that's upbeat and rockier than today's previous sounds. I connect my iPhone via Bluetooth to the in-car Ford Audio (app/facility?) and hit shuffle. Track list: 'Ain't Nothing But A House Party' by The Showstoppers, 'Harry, Turn The Music Up' by Otis Taylor, 'Hells Bells' by AC/DC, 'Hands On Your Stomach' by Otis Taylor, 'Hard To Handle' by The Black Crowes, 'Somebody To Love' by Jefferson Airplane, 'Back in Black' by AC/DC. It seems odd that two artists appear twice from a shuffled 16GB music library, so I guess there are some other factors at work here (#2311)

> I need something peaceful and mellow, and my current band crush is The Staves. I cue up 'In The Long Run' on the iPhone and listen to it twice on repeat. Then I decide to embed it on the blog and end up listening to a live version on Soundcloud, too. I love the finger-picked arpeggio that underpins the melody and make a mental note to look up the fingering online. I make a further mental note to buy a guitar (#1728)

Above we see interactions that manifest for respondents as part of an on-going dialogue and negotiation between cultural practices that are informed in part by the previous, strategic technologies of the record, the radio and musical instruments, but also by tactical cultural practices newly emerging which are inspired by digital and Internet technologies. What is perhaps more interesting and potentially useful to us in the manner in which these newer technologies of listening are developing, are the new conditions and consequences of their use.

7.3 – Connectivity, interfaces, and abstractions

As was discussed in the opening chapter, because data collection (and subsequent analysis) is, as Housley et al show, a "functional prerequisite of network intergration" (2014:2) these newer technologies have, according to both proponents and critics, the potential to close the gap highlighted by Felski between the ordinary and the elusive. To its proponents, datafication (Mayer-Schönberger and Cukier, 2013) suggests that many (and progressively more) elusive elements of the mundane can now be captured, monitored, and perhaps even predicted. This is one of the key consequences of our use of devices now closely associated with listening. But here we return to the question regarding the extent to which these technologies can account for that which is evidenced above, i.e. the on-going, abundant detail contained within the strategic/tactical negotiation between individual cultural practices and technologies of reception that occur in mundane, everyday situations. Once again, as Prey (2015) observes, the qualitative is never fully absorbed by the quantitative, and as such De Certeau's strategies of place and tactics of space remain in tension. However, the gap that exists between the two is where cultural practices of listening and digital and Internet technologies collide, and as such it is a useful and interesting location for further enquiry.

> Playing right into the Spotify marketing dudes hands and in a fit of rebellion against the Spotify Discover Weekly playlist (zzzzzzzzzzzzzzzz) that apparently with its swanky algorithm can predict what I want to listen to (you can imagine how well that's gone down), I made my own. So basically I'm using and sharing their product just as they'd like - REALLY rebelling there then ;-) ooops. I started off with the De Lux track on Lauren Laverne's show this morning on 6Music, got side-tracked with Talking Heads (the De Lux singer sounds like David Byrne to me) then sort of meandered round a bunch of 2015 releases before getting side-tracked again. What? It's summer, you're meant to meander ;-) (#6268)

The example above demonstrates an instance where public perceptions of commercial monitoring come up against individual cultural practice, and it neatly encapsulates the tensions that exist in the negotiation between the two. Thinking through and updating Bull's (2009) idea of "auditory nostalgia" allows us to start to make sense of this negotiation and tension, and also enables us to consider the conditions and issues that are new and unique to the contemporary

landscape. Auditory nostalgia is an idea Bull developed through an examination of iPod culture and the manner in which communication technologies are used in the management of daily experiences. Bull describes listeners as "living in an interiorised and pleasurable world of their own making, away from the historical contingency of the world, and into the certainty of their own past" (2009:84). Since Bull wrote, however, the novelty and role of the iPod – and thus the new cultural practices it helped inspire – have been overtaken by connected devices and digital interfaces that perform the similar tasks. The important addition is of functions that provide and forge connections between internal and external worlds. Playlists in streaming services, for example, can be more easily assembled and shared than those created with an iPod. There is thus an emerging, new ground created by this recent change in the means through which people engage in the reception of music. It is a shift that sees the personalised, internalised narrative experiences suggested by Bull's work cross over into more collective, connected experiences of the present. This occurs through the close integration of always-on, mobile communication technologies into the existing technologies and cultural practices of listening. We can see exactly this in the now-ubiquitous Smartphone and through the digital, connected interfaces such devices enable: a streaming service is both a personal music library and a technology that links "internalised and pleasureable" worlds to the connected world. The intriguing issue raised by this is how, where, when, and with whom, these new collective, connected experiences manifest themselves, and whether the nature of our reception experience (our routines, and thus our routes to meaning) alters, inspiring new cultural practices as a consequence.

> I currently have an iPhone, one of the reasons being because of its integration of iTunes Match....Previously keeping music on my phone (or other handheld media device) involved having to make a choice of what particular albums I particularly wanted to have access to and fitting them on an 8gb, 16gb, or 32gb (depending on what year it was!) storage card. Even with the 32gb card I was always frustrated at how little I could keep to hand, so £21.99 a year to take that frustration away from me is easily money well spent….for my commute I usually set my phone to play music in a random shuffle manner, from song to song. This has an added bonus that actually quite a sizable portion of my music collection I've never actually heard before - so it's as good a way of making new discoveries as the

> traditional way of listening to the radio! In order to log what I've heard, I use the iTunes star rating system on each track as it plays _ basically one star indicates I intend to delete the track (or the whole album, or even the whole artist) from the library, two stars puts the album or the artist at risk of deletion, three is my basic default of yes, I like this, four is an indication of 'So, this is interesting'. I'll go back and listen to the whole album and more from this artist at some point in the future, and five stars is a log that the track is a definite favourite with me. (#2897)

Here the listener is using the mobile device not only as means of accompaniment through an everyday commuting experience, exactly as Bull's iPod would have, but also as a means through which the respondent can manage both the unknown elements of an abundant music collection, and their future activity with it. They are not only heading towards the "certainty of their own past" (Bull, ibid), but also somewhere else. In this instance, through the logging of ratings, knowingly written to digital monitoring systems in the cloud, the respondent attempts to guide activity as yet only imagined, because the results of the user's categorisation activity will appear as a form of playlist at some future point when using the iTunes interface. This is seemingly the intended and desired outcome of their present use of the available technologies. Through listening in this manner they are not only drawing upon, or creating memories, but also facilitating the possibility of their creation at some future point. The respondent is, in other words, seemingly comfortable with the act of communicating with a manifestation of his digital self – his music preferences rendered in abstract, numeric form – that will eventually inform the activity he engages with once 'back' in the real world at some future, imagined point. Through the tactical use of available strategic technologies, the respondent is turning the processes of datafication to his own ends and becomes part of what Webster et al (2016) would describe as the assemblage of his own music recommendation algorithm. He is taking personal ownership and attempts to exert a degree of agency and control over his 'personalisation'. He is – in De Certeau's terms – carving out a space that is tolerable through the act of communing with his own digital abstraction.

This activity offers a vital clue as to where we may go next, and can be further understood in terms of one of four inter-related functions Weber (2009) ascribes to the use of mobile devices as "sound souvenirs". Here technologies act inversely to the souvenirs gathered by tourists visiting foreign lands, and instead provide listeners with the possibility of configuring and shaping their individual and group cultural identities when they are removed from familiar surroundings. By taking familiar music into unfamiliar territory, listeners use sound technologies in their "rewind" function, and by making the routine of a commute (or similar) new and unique, they utilize the "forward" function. Both help create "future" sound souvenirs in that the familiarised unfamiliar, or the unique routine, can become a memory and help maintain "a sense of emotional and cultural' identity" (2009:80). We can see exactly this in the following examples:

> I'm listening to 'African Lullabies' which I downloaded from iTunes a couple of days ago. It's a compilation of songs from African artists such as Floxy Bee. I have no particular claim to the cultural or geographical history of Africa, but I have visited Kenya and South Africa. This album reminded me on my boring SouthEastern train commute that people make music and express their feelings in many different ways, but the feelings themselves are common to all of us. The joy and peace of a quiet song which you sing to children is universal.(#1077)

> Listened to these songs on Spotify for first time in years! Some songs that I uncovered/rediscovered Here Without You - 3 Doors Down Black Balloon - Goo Goo Dolls Climbing The Walls - Backstreet Boys Empire In My Mind - The Wallflowers I was surprised to find myself singing along to these songs word for word; not a single lyric missed. They are not even in my top 20 favorite songs. Yet my brain remembers. This gotta be the power of music (#2881)

> I used to think local music was just angst ridden spotty boys with weak voices playing morose songs about lost love, or hippies with harps. Then I met [Twitter name] and my mind was blown. Since that meeting I have made friends with so many local musicians who ARE AMAZING! Then podcasts from @atlumschema and @RightChordMusic opened my ears up to more, more, more…Literally, [Twitter name] saved my musical life. <3 (#5148)

Sound souvenirs are thus a useful way of conceptualising the use of connected devices, certainly, with mobile and connected music listening acting in various ways on Weber's "emotional and cultural identity", in both "rewind" and "forward" modes, in much that same way that the Walkman or the iPod may have been used. But I suggest that is the very connectivity of these newer devices and modes of listening that provide the intriguing developments and the genesis of new cultural practices, particularly in the sense of users exerting degrees of control and agency in terms of technologies associated with processes of datafication and digital abstraction. Indeed many of the examples above pivot almost entirely around the connectivity that facilitates such processes. This is a factor not considered in Weber's study, which stops at the Walkman, or in Bull's work that focuses on the comparatively unconnected iPod.

We can consider further implications of this by revisiting the work of Mejia, discussed in chapter 2, which demonstrates that connected mobile devices impact not only the aesthetic experience of our surroundings – the places and things we encounter in the world – but also on our structural understanding of them (Mejia, 2012). The manner in which mobile devices are used in the home, for instance, illustrates how this arises:

> Listened to Radio 2 in the car for the school drop-off and after-school pick up at my daughter's request (9 yrs old). - now premium Spotify at home and in the office, & my family sideload cached files when unable to stream it. Now using it to listen to as much new stuff as stuff I had forgotten about...and discovered some great artists, too, like Laura Welsh whom I've listened to for 2 hours today, repeat playing her tracks through Spotify (#3284)
>
> After breakfast, comes the usual diet of streaming from Deezer. It acts as perfect background noise for if I'm multitasking with other work going on in my life. Chasing Yesterday by Noel Gallagher's High Flying Birds seems to be on my top must-play albums. I listened to the free stream on iTunes and I fell in love at the transformation of one of the finest Manchester musicians of the 90's. Just awesome. (#6343)
>
> Home again, now dinner and new music discovery time. BBC 6 music radio, listening online from the opposite side of the globe, wired wireless. Spotify, "recommended for you!"(#3748)

> I have moved my iPod speakers into the garden today to listen to Bombay Bicycle Club and basque in the sun. Everything about their albums brings back memories of summers gone by so they are my perfect band to listen to on a sunny day. I started my morning listening to 6Music. Lauren Laverne's biorythms never fail to bring my normally wandering attention to new music and always manage to have me too bloody close to tears with their much loved personal stories. I don't understand how music manages to capture us all so much. How does it create that impact?! I just accept it and let it take me over. (#4540)

As we see in the examples above, the terrain of mobility becomes geographically much smaller as the move from room to room (and to the garden) replicates ideas of an everyday journey, such as that taken from the home to the office. As internet and digital monitoring-enabled devices are now increasingly domesticated in the home (including not only the domestic use of mobile devices, but also in more static technologies, such as Smart TVs) then practices of sound souvenir use begin to follow a reverse trajectory than the ones previously conceived of by Weber and Bull. In other words, rather than taking a familiar 'domesticated' cultural practice with us as we move, we are instead developing the habit of re-producing practices linked to mobility when we are in situations not normally linked to mobility, such as the home. This in turn changes the nature of the home (and similar spaces) as places – exactly as Mejia describes – because the ubiquity of mobile, connected devices (which Mejia calls mICTS) and their centrality to cultural practices such as engaging with music, social media use, the consumption of information, and so on, not only enable us – voluntarily or otherwise – to remain connected (and thus out in the open) while being physically situated behind closed doors, but also because we engage primarily through digital interfaces and with what Mejia calls avatars. In other words, our aesthetic experience of the structural world around us occurs through engagement with digital abstractions of real-world people, places and items. One such item, of course, is the recorded music object, now rendered in digital, abstracted form through interfaces such as Spotify.

It is here we can consider observations made by Sterne (2009) regarding digital technologies and music reception. He points out that in making recordings more

portable and easily stored, in other words one of the advertised progressive benefits of digital technologies, we have also made them paradoxically more and not less ephemeral. In describing the manner in which his friends are getting rid of their CDs and records in favour of 'fragile' hard-drives, he points out that collections may no longer outlive their owners. This is a process that is "less a simple kind of forgetting, like forgetting where one left one's car keys; it is more properly a *forgetting of forgetting*[3]" (2009:59). He argues that this has been a fundamental condition of recording throughout its history, and I would add that this is particularly so in the case of streaming, the latest incarnation of the recorded object[4]. Within the streaming interface all one truly engages with in terms of objects (as distinct from sounds) are the visual representations of metadata associated with recordings, and not the physical representations of the recordings themselves. The examples below point toward how much of Harkive respondents' music reception activities involve interfaces that provide not recordings, but rather the representations of recordings, which Mejia would call avatars.

> Currently listening to Stromae, after a recommendation from a friend this morning. For ~15 years I gave the same middle-of-the-road answers whenever asked about my favourite artists. But for a few months now I've been discovering all sorts of music, and going to gigs also. And all thanks to two free apps on my phone: Napster, for access to the music, and Songkick, for early notification of concert tickets. Both have truly transformed my life by bringing music back into it. Given the power of music that's not an overstatement. (#3091)

> Listening to Little Dragon on Vimeo after hearing it on the Tour de France highlights show yesterday. I looked it up on Shazam to find out what the name of the tune was - I knew it was Little Dragon. She's cool. Really like the tambourine rhythm, almost like a samba... http://vimeo.com/31246558 @LittleDragon Listening on headphones at work. One ear on, one ear off. Busy day. (2819)[5]

[3] Italics in original text

[4] Sterne argues that recording technologies as a whole promote disposal since the commercial imperative that produces them is one of continual novelty. This position is somewhat problematic because, I would argue, the central positioning of the commercial imperative is questionable. However, we can acknowledge that the process of forgetting posited by Sterne does not slow down but rather increases with streaming.

[5] It would be interesting to see, considering Sterne's points, how many of the links to songs and videos, or indeed the mentions of services such as Shazam, Songkick, Napster, that people have described and included in their Harkive stories, remain active and/or understandable over time, and for how long. In the case where links are removed (or the licensing that facilitated the link has altered, or expired), all that would remain is the record of the listener's description – and only then if the 'fragile' hard-disk holding the Harkive dataset survives.

Here the progressive developments in terms of playback technologies become closely linked the ways in which we may remember our cultural practices associated with music reception. I suggest that this has implications for ideas related to the relationship between routine and meaning previously identified as key elements of music in everyday life. What is experienced in the present, or what remains in the future, are less likely to be – to take one example – a record collection in the traditional sense of CDs or vinyl records gathering dust on a shelf, or even in terms of files on hard-drives, but rather instead the digital trace of a collection. This, by extension, implies the digital trace of the listener and their cultural practices, and here we can consider Crawford's observation that there are often "far-flung assumptions about what the human subject does and what our data traces reveal..[where] the assumptions that go into the making of the trace body have become so attenuated and, in some cases, ridiculous, that it's critically important that we question these knowledge claims at every level"[6]. It follows, then, that it may be the case that the ways in which we as popular music scholars understand cultural practices linked to music become lost (forgotten), or potentially redundant. To select but one example, using Benjamin's (1931) thoughts about his book collection in order to understand vinyl record collecting, for instance, presupposes an understanding of both what a book collection *and* a vinyl collection are – and what they mean – as objects in the physical world now. This is a position that changes when individuals are instead "content curators" (Hagen, 2015) of Sterne's "fragile" abstractions, and not collectors of objects.

This suggests new directions and aims for our work as popular music scholars. Alongside developing an understanding of the materiality of digital music technologies, as Magaugga (2011) suggests, we may also (and perhaps more importantly) consider devising ways to understand how collections and other cultural practices related to music become manifest in terms of digital, abstracted space. In other words, I suggest there needs to be a way to understand, to somehow make concrete in our conceptions, the ephemeral and

[6] Crawford was in conversation with Hito Steyrl, a transcript of which was published in The New Inquiry: http://thenewinquiry.com/features/data-streams/

abstract realities of the zeros and ones that now operationally (see: Liu (2016) in chapter 3) represent both music and the cultural practices of consumers associated with it. This requires an understanding of not only what constitutes their materiality, but also of how they are, as Prey (2015) and Lefebvre (1991) help us to consider, conceived, perceived and lived in the contested areas that occur when everyday, 'lived' space meets the 'abstract' space of digital and connected technologies. What adds further difficulty to this task, is not only that the data which may provide a route towards an understanding of such materiality is largely an abstraction, but also that the digital traces of individuals and their collections, if they can be understood at all, may only be understood in the aggregate sense of trends and correlations. The question becomes, then, whether the individual (data point) experiencing music via technology in a given time and place can be understood when considered in isolation once he or she is rendered as an abstracted digital trace.

To consider this in more detail, the concept of ubiquitous subjectivity developed by Kassabian (2002) is useful in terms of plotting our next steps. The following examples, which detail music listening experiences for Harkive respondents that pivot around numerous online services carrying abstracted, digital content that is accessed through mobile interfaces and devises that in turn facilitate the creation of digital traces of cultural practices (and thus routine and meaning), illustrates how for Kassabian the development of recording technologies over time has "disarticulated" the listening space. There is no longer "a place" in which to listen since everywhere is such a place; listening "blends into" environments without calling attention to itself.

> Radio 1 is always my introduction to music every day whilst I get ready for work. Today was no different. For my commute I listened to the new Slow Club album 'Complete Surrender' on Spotify. Tuesday is 6 Music day on the office radio so I get in and switch that on which will be playing in the background all day. After lunch I popped on the Alvvays album, which was streaming in advance of release on NPR.org. After that enjoyable Lo-fi indie interlude it was back to 6 Music on the office radio again. 3:30pm coffee break time, checked out the new Lykke Li video for 'Gunshot' on YouTube. This was swiftly followed by the La Roux album which was streaming

> directly from the artist's official site. Back to Spotify to listen to playlist 'Top of The Poptastic 2014' compiled by pop blogger @Poptastic. Mainly to avoid Steve Lamaq on 6zzzz. En route to play football after work I put on Cher Lloyd's new album Sorry I'm Late to fire me up for an hour of running around kicking people. Not exactly an obvious choice but it worked (#5792)
>
> Mr Fine Wine's 'Downtown Soulville' show while reading on the tube. It's a radio show that I get as a podcast every week. While having some lunch I watched a 60 Minutes interview with Eminem that had showed up on Reddit. The new Ratatat album, Magnifique, via Spotify. The new album by the Bird and the Bee, Recreational Love, also via Spotify. Wilco, Star Wars. Also on Spotify. Someone tweeted about how their niece (Holly Macve) has been signed to Bella Union and will be touring with John Grant. Listened to one of her demos on Soundcloud. Listening now hijacked by Soundcloud. Listened to Major Lazer's cover of Lost by Frank Ocean then a DnB playlist from UKF. (#1064)

For Kassabian the unpacking of catalogue into recontextualised forms (for example, the multi-genre, multi-temporal playlist) makes it "easier to use the music as an environmental technology, conditioning and conditioned by a new kind of subjectivity" (2002:134). The listener, in this sense, is not an individual, but rather a part of "an always moving and ever-present web". Not a listener understood by categorical variables hitherto used in conceptions of audiences in popular music, such as genre, class, or gender, but rather one understood in terms of categorical variables imagined only by the aggregation of abstracted data points, where all listeners are connected by the ubiquitous music that 'cables all of us together'. It is here also we can recall Cheney-Lippold's (2017, 2011) "measurable types" and the potential for Manovich's "programmable user" (2001). Kassabian, quotes Clough, who proposes "a new ontological perspective and an unconscious other than the one organised by an oedipal narrative" (2002:20), which speaks to both emerging calls around post-humanities scholarship (see for example Chun et al , 2013) and also, in the words of Prey's update of Williams' (2011) work, the idea that "there are no individuals, just ways of seeing people as individuals"[7] – thus Kassabian's demonstrates her theory of subjectivity, which is arrived at via the proliferation of a ubiquitous

[7] This comes from an as-yet unpublished paper of Prey's I heard him deliver at AoIR conference September 2016 in Berlin.

music which "forms the network backbone of a new, ubiquitous subjectivity". This can be understood as a state of being where we "prefer to be connected, need to listen to our connections, cannot breathe without them. [Where] we already live in a network we insist on thinking of as a dystopian future". And here, of course, another paradox emerges, at least according to Kassabian, in that we seem to somehow welcome such a future/present because "isolated consciousness – silence – is unpleasureable in the extreme" (2002:141)

> Music for me is a means of escape and improvement. I misplaced my iPod for a few weeks and it felt very odd leaving the house without it, since I tend to walk most places and felt strangely exposed without it for those journeys. I listen to a relatively wide selection of things and would consider myself informed although not really knowledgeable about music (#2715)

> Woke up at 7, first thing I did when fully awake was to launch music on my computer. Whole music library, randomly playing. - Started work at 8.30, ended up on a music streaming website. I chose a user playlist: classic and/or mainstream rock - mostly tracks I already know, as it turns out. But if I come across some unknown & nice tracks, I will dig into them. I tried another one (EDM), but stopped listening to it because I couldn't concentrate. This may sound weird, but I need music to concentrate. It keeps me in a little private space with no distraction. - Going out after work, and I'll probably take my mp3 player with me. Actually, I listen to music most of the day. I hate silence. I listen to almost everything, depending on my mood. I listen to music on my computer, for I have no other device to do it - and I am a computer geek, my computer's on all day. (#3610)

It is through a consideration of all of what has been discussed throughout this chapter where the route towards interesting new questions reside, and where Housley et al's "digital remastering" (2014: 5) of older questions suggest themselves. Given the rapid rise of connected forms of music consumption in recent years, and through this close engagement with Harkive narratives in this chapter that has allowed me to provide further answers to my central research question, I offer the suggestion that there is a need to arrive at an understanding of modes of music reception and the cultural practices associated with them that can consider the additional dimensions added by connectivity and datafication in relation to routine and meaning. In other words, popular music scholars may wish to consider how digital technologies and connectivity fit into what

Hesmondhalgh (2002) says are the crucial rhythms and routines involved with understanding music listening and everyday life now that communications technologies are, as Liu (ibid) described in Chapter 3, operationally and thus culturally "ordinary". What could also be considered when attempting to address this is the extent to which the use of connected/mobile devices can be viewed in relation to the "domesticated" (Baym, 2010) nature of such forms of connectivity, in other words, when listening via always-on devices becomes the norm rather than a special event. We may wish to consider also the data collection that go hand-in-hand with this, and which have "nestled into the comfort zone of many people" (Van Dijck, 2014:1), or the mICTs that impact on not only our aesthetic experience of our surroundings but also our structural understanding of them (Mejia, 2012).

7.4 – Environments, musicking and techoustemology

Through the layers of complexity and further issues engaged with throughout this chapter, I am suggesting that an attempt should be made to create a means by which popular music scholars can understand music reception – and its associated cultural practices in everyday life, and thus routine and meaning-making – in a manner that is able to take into account exactly that complexity and those issues. Such an attempt, I suggest, would be directed by the desire and intention that an understanding of the relationships between music and the individual as they occur when mediated through digital, data and Internet technologies, in different times, places and spaces, is able to grasp the epistemic and ontological shifts that are currently occurring. This shift is variously altering ideas of place, space, agency, routine, meaning, subject and object, and is creating digitally imagined variations of them via enabling representations/manifestations of each that increasingly play a role in everyday life.

Here I speculate that one such route toward that new understanding is to take Nowak's (2016) model of "environments", which too seeks to understand music

listening in the everyday, as a starting point, and to then consider Porcello's (2005) concept of techoustemology alongside Small's idea of "musicking" (1998). This, I suggest, is a means by which we may understand the various issues and activities around data collection, music reception, always-on connected devices, and 'ubiquitous subjectivity' discussed through my analysis of Harkive data in this chapter. These ideas need to be explored in turn in relation to the observations I have made based on that data.

Nowak argues that music listening can be understood as the relationship between a subject (the listener), and an object (a sound, noise, etc), that is "performed" through technological artifacts within specifics of space and time. According to this model, we have a way of conceptualising listening as being comprised of four inter-related components. We can indeed see numerous examples of this in the Harkive stories, such as this one, which encapsulates the idea of a listening environment in Nowak's terms:

> It's the hottest day of the year thus far in Dublin so this suburban white boy is bumping 'Islah' by Kevin Gates at a sensible volume from my iPhone via the Sony Bluetooth stereo in my car on the drive to work. The sounds of Dublin's only (and sadly soon to be defunct) alternative music radio station 105.2 TXFM competing with the clanks and crashes from the construction site opposite the office. At time of writing, it's the very apt 'Blister In The Sun' by The Violent Femmes courtesy of Cathal Funge on TXFM Breakfast. Part of my job involves editing audio for video so I often have to spend time trawling through stock music sites for music beds. Today I'm hunting for something inspirational and upbeat for a corporate video for a garden design school so am auditioning watermarked previews of suitable tracks on AudioJungle.net. Enough musical wallpaper, time for '15 Songs Indebted to Suicide's Alan Vega' compiled by Stuart Berman for Pitchfork. (#7116)

Via Nowak, if we consider that music is always mediated, the meaning of listening activities are always subject to change, and it is thus "the environment of activity [which] mediates the relationship..and creates a 'system of interactions' between the subject, the music content, the technology and the environment" (2016:21). Through this Nowak builds an approach that "considers listening activities as uncertain in their outcomes, with individuals

engaging in an endeavor to grasp upon the "adequate music" for a given situation" (2016:21). Further to this, the proliferation of digital technologies related to music listening, which contribute to its growing omnipresence in everyday life, in turn "augments the contestation of its meaning" (2016:21). Every given situation, in other words, is thus an opportunity or challenge. Digital technologies, then, have the effect of further multiplying the possibilities for "activities and materialities", which we have seen throughout this chapter. As we see in the responses to Harkive, reception of music does not only occur within the confines of digital interfaces, but jumps between numerous digital interfaces operated by different companies, and further spills over into cultural practices linked to CDs, radio, vinyl records, songs in the memory, etc, that make a complete picture of listening activity difficult to grasp. Activity that is performed through digital interfaces also creates digital traces of the real-world activity, but these do not capture the entirety of cultural practice, routine or meaning. This means that whilst, for instance, Spotify may well be able to gather detailed information about the activity of listening to music during a commute, they are unable to gather information regarding the music you have listened to on your kitchen radio, or your turntable, or the tune you whistled from memory. As such, activities which lead to concerns of the programmable user are only ever based on an incomplete inference – the inherent reductions and abstractions highlighted by Berry – but the results, as Cheney-Lippold shows, manifest as consequences in the real world.

Nowak and Bennett (2014), meanwhile, argue that often concepts of music as an object are sometimes too rigid and suggest that combining Small's concept of 'musicking' could change the "conventional interpretation of music as a relatively static cultural 'object', to recast it as a 'process' individuals interact with, and are acted upon" (2014:429). I have consistently used Negus' definition of music reception throughout this thesis as a way of understanding cultural practices associated with music that include not just 'objects' but also memories, discussions, and other immaterial elements, and it is from this point that I suggest that a similar approach could be applied to the components of connectivity, mobility and abstractions discussed throughout this chapter, and

particularly those related to data collection and analysis. Whilst coining 'dataing' in an attempt to update Small would be perhaps too cumbersome, the idea that data-derived business practice is something individuals interact with and are acted upon during the ongoing negotiation between strategic place and tactical space is nevertheless a potentially useful one, but it also requires additional thought. It is here where we may return to Bijsterveld and van Dijck, with whom we began this chapter. They were writing in the introduction to their edited collection, "Sound Souvenirs"(2009), a book with which they aimed to make a contribution to Porcello's (2005) idea of "techoustemology". This is a means by which, according Porcello, we may "[provide] a window into how people deploy technology to engineer (whether by making, listening to, or circulating) their musical and sonic lives" (2005:270). This is an interesting concept and approach that can perhaps begin to tie together and present a way forward through the exploration of the inter-related ideas in this chapter.

Porcello describes techoustemology as "the implication of forms of technological mediation on individuals' knowledge and interpretations of, sensations in, and consequent actions upon their acoustic environments as grounded in the specific times and places of the production and reception of sound" (2005:270). My attempt to 'walk' with the Harkive respondents through their music reception practices in this chapter leads me to speculate that a consideration of the implications of digital, data and Internet technologies, and to the results of data collection/analysis in particular, may be a useful starting point for subsequent work. In other words, to the complexity of ideas of technological mediation (for instance, the recording and production of sound) we may wish to add to the mix a consideration of practices linked to data collection and analysis, the deployment of dynamic interfaces, and so on, as elements which - as I have shown - also have implications for how audiences experience the world around them, and specifically through their acts of music reception. Such a view enables us to consider ideas of digital traces and the abstraction of social action in terms of a technological mediation of music reception that, following Porcello, have implications of their own. Related to this, it is interesting that Porcello makes the further point that even when technological mediation is not present "its absence

has a ghostlike implication[s] for our expectations, our templates, of how we expect sounds to sound" (2015:271) and that it demands "an accountability not only to the specifics of how technology is deployed but to serious consideration of the resultant *sounds* of sociotechnical action, and to how individuals and groups conceptualise, rationalize, and discursively render their own techoustemological ideologies and practices" (2005:271). Likewise, the centrality of digital, data and Internet technologies to contemporary practices of music reception suggests that these elements too may have consequences even in their absence: the listener attuned to the daily use of the digital music interface; to the audio quality of streaming services; or to their habitual use of an automated recommendation service as a means towards discovery, may in various ways – and similar to the manner discussed in chapter 6 – transfer across, repeat, or in some way be renegotiated in the non-digital spaces of the present and future. Considering the data, digital and Internet technologies of music reception in these ways is promising in terms of a route forward, because it allows us to consider ways that we may begin to account for their role, function, and potential benefits and consequences, as they become further enfolded into everyday practice. As promising as this suggested route forward may appear, however, the task remains a hugely complex one.

Cultural practices associated with music and the everyday and the relationships we are developing to new digital technologies, evidenced by my analysis of the Harkive data in this chapter, are – if we follow Porcello and Small – too complex, fluid and idiosyncratic to be understood by rigid systems. We could understand rigid systems in terms of both Nowak's 'environments' model and also the 'Taste Profile' of a Spotify user, and here we can begin to think of Feyerabend's (2001) work and the implications of what he describes as the imposition of epistemological dichotomies. Whether patterns emerge across individuals, whether the individual practices are too idiosyncratic to be unified, argues Porcello, is the question that has "characterized – perhaps even 'plagued' – media and cultural studies" for decades. As he observes, "one person's agency seems unfailingly to be another's determinism" (2005:272), which is yet another dichotomy. The new question that emerges from my analysis becomes, then,

twofold: to what extent can either academic inquiry or commercial activity, each of which seek to explain and understand cultural practice as it relates to music and the everyday, ever adequately reach the point where they can explain the object under their gaze; and secondly, what are the consequences of those attempts when each in their own way, through their inherent positions of relative authority, (re)produce the status quo in terms of knowledge and being. Porcello describes this in terms of "they do, we theorize"; we could easily adapt this to Spotify (and others) as 'they do, we analyse'. Porcello argues that by reverting to the locality of ethnography we can "give voice" to the "do-ers who possess eloquent, theorized voices of their own" (2005:272). The 'do-ers' in this case have been the respondents to Harkive providing windows into their individual, unique worlds, and who have been shown in this chapter to be developing ways of curating their cultural practices, routines and routes to meaning, with efficacies of their own.

7.5: Discussion

This brings me to my concluding suggestion and towards the introduction of some ideas for next steps my research could take based on the observations of this chapter, which themselves are the culmination on my engagement with the various issues of debate and method that have unfolded across this thesis. In short, I suggest that popular music scholars make efforts to enable everyday listeners to attempt to understand data technologies, digital monitoring and related systems as adjuncts to Porcello's concept of techoustemology. The focus in my findings chapters has largely been on formats, services and devices, and the manner in which these intersect with cultural practices in the everyday, but the suggestion for future work is to look at the possibilities for data collection, visualization and computational analyses to become similarly enfolded into the everyday routines and cultural practices of listeners. What would happen, for instance, if these processes and their role in cultural knowledge production and ontology were to became as similarly 'domesticated' as digital technologies of music reception, and if the tools and analytical methods that they facilitate were to made available to "amateurs" (Gomart and Hennion, 1999) in a way that

enabled them to seek out and alight upon their own 'adequate' relationships with them now that the sociotechnical systems, the abstracted and reduced representations of music, of time and space, and of the self, form part of their everyday experience?

It could be argued, of course, that many of us already do this to an extent. Each time we engage with the content of websites, the interfaces of streaming services, and so on, we are involved in a relationship with them. Indeed, the example Harkive stories in this chapter demonstrate how this is already occurring. The new questions that my research has led me towards, however, relate to the extent that this is/could be done more productively and reflexively, and whether there is space, or the possibility for the creation of space, for the users – the everyday users – to develop epistemic responses to their new technological and societal conditions, and whether this could 'inspire' new cultural practices.

Here I envisage something similar to the scholars employing digital technologies in their particular pursuits of knowledge, such Liu (2016), Kitchin (2014), Piper (2016), Manovich (2016) and others, who variously suggest that such engagement should occur critically and reflexively. We can think here also of the workers (human and non-human) embedded in the design and delivery of digital interfaces and data monitoring systems that are (consciously or otherwise) driven by the commercial and rational imperative for Feyerabend's "prodigious power of performance" (2001:6). Here also there is a reflexivity of sorts as each (to varying extents) consider the ethical and social boundaries within which they operate because they work under the gaze of the regulators and publics struggling to keep pace with it all. What if the tools and skills that I have shown are now available to both scholars and commercial interests were to become available to all? What kind of new cultural practices might this help produce?

Based on the findings of this chapter, arrived at through 'walking' with the Harkive respondents, I suggest that new means and models of understanding the activities and outcomes of the reception of music could be explored in order for

listeners (and everyone in the everyday) to be able to make sense of both the processes and the conditions under which their practices are monitored and guided. Spotify's previously discussed 'taste profiles' are a digital abstraction of us and our tastes that are rendered in digital space and in a form that we presently lack the conceptual capabilities to fully grasp. And this is but one of many abstracted selves created by our activities when engaging with music through connected devices. On Discogs, Twitter, Facebook, and so on, we continually create (other) versions of ourselves (or have them created for us). How might we come to terms with that as popular music studies but also in the management of our everyday lives?

One way is through the development of the work I have begun. I have attempted, in rudimentary terms, to create a representation of everyday listening in abstracted form through the experimental space of Harkive: a simultaneous space/place where respondents are the informed, the informant and the information (Michael and Lupton, 2015), and through which various different forms of understanding have been produced via abstracted computational processes. The understandings produced in some ways show a degree of efficacy, in other ways they are lacking – they cannot come close to accounting for the sheer abundance of the everyday. This is the central problematic I have wrestled with. The systems, processes and analyses I have conceived of are limited in size and scope, particularly when compared to the sociotechnical systems deployed in commercial and others operations. They are limited in scope also to systems used in other areas of academia, particularly more advanced work in areas of cultural analytics and digital humanities. Nevertheless the work has, at least, begun. In the concluding chapter of this thesis I will undertake a reflection on the successes and failures of this project as a piece of practice-based research.

CONCLUSION

The meeting: next steps

On the afternoon of Friday 24th March 2017, with just a few months of my 3-year PhD project remaining, I sat down in a meeting room to discuss my research. The meeting room was located on the campus of a different university to my home institution, and within a department that does not focus on media, culture or popular music. The meeting had been set up by a friend of mine working at that University, and within that department, who had thought it might be interesting for me to meet one of his colleagues.

The colleague in question is a professor who works with complex data systems, mathematical models, and algorithms. The work his team of researchers performs is concerned with assessing the design and functionality of large-scale engineering projects. They develop computer-simulated models to assess how things such as engines, bridges, buildings, and other complex mechanical operations may perform under certain theoretical conditions: extreme heat; freak snowstorms; intense acceleration; rapid deceleration; unexpected impacts. The results of their work inform the design, manufacture and performance of multi-billion dollar commercial and governmental projects around the world, and sometimes even in outer space. We were in the room, however, to discuss popular music, digital technologies, data and – specifically – my work. The professor, it transpired, was just like you and I and millions of others: he is a huge fan of music.

I found myself in the position that many researchers dread. I had to succinctly explain the aim, purpose and outcomes of my work. I began by explaining that I was interested in the changing nature of the popular music landscape, and in particular how digital and Internet technologies and data-derived business models were playing a role in that change. In order to explore this, I explained

how I had developed The Harkive Project, and that this in turn had collected thousands of stories from people about their experiences with music. I then described how these stories were rich in detail and revealed the complexity of what went on in everyday situations when people engaged with music. I then described how I had arrived at the decision of using computational analytical techniques in order to help me make sense of the data I had gathered, because I conceived of this as a means that could simultaneously help me deal with the complexity within the stories, whilst also learning more about how similar technologies operate within the field of popular music. I described how – and perhaps more interestingly – that an unexpected outcome of taking that path was that I had been able to build an understanding of how systems of computational knowledge creation related to research processes in the humanities. I talked specifically about popular music culture, and how – for instance – machine-derived recommender systems might be changing the nature of that in important ways. I talked also of how the idea of having access to millions of songs via connected, hand-held devices was something that we – the people who have music woven into our everyday lives, our identities, our plans, activities and memories – were still getting used to. I explained also that several scholars from within my discipline, the humanities, were attempting to marry the tools and frameworks for critical engagement that we do have at our disposal with a greater practical understanding of complex social-technical data systems. It was at this emerging, intriguing intersection where I saw my work.

***

I opened my thesis with a personal account of how digital, Internet and data technologies played a role in a specific experience I had with music. My use of a connected, mobile device that provided me with on-demand access to a large catalogue of music had, through practices associated with data-collection and analysis, foregrounded a particular song to me that had in turn produced a strong emotional reaction. This experience prompted me to think about the questions that I was beginning to develop in the early stages of this research process, about the role of digital, Internet and data technologies in the ways we

experience music. These questions were informed in part by my personal and professional experiences with music over a period of many years, and eventually those led me towards more academic pursuits and to the creation of The Harkive Project that forms the central piece of the work I have presented here.

In this work I have been concerned with a central research question: what can an analysis of the data generated by The Harkive Project reveal about the music reception practices of respondents? In answering this question I have been engaged with three inter-related issues of debate and method: how cultural practices associated with popular music were changing in light of digital technologies; how these changes may be explained or understand through an analysis of the music reception stories gathered by The Harkive Project; and what the methods I had developed in order to perform that analysis may suggest in terms of possible new directions and approaches for popular music studies research. In undertaking that work, and in answering my central research question, I have built and reflected upon an innovative and experimental methodology that has been able to engage with those inter-related issues.

From its promising beginnings as an MA project in 2013, and over the last three years of my doctoral research, I have developed Harkive into an online space that encourages people to reflect on their engagement with music whilst it simultaneously acts as a place capable of replicating many of the commercial practices related to data collection and analysis that I have shown throughout this thesis as now playing central roles in the means by which we engage with music. By conceiving of Harkive in this manner I have not only been able to critically engage with the role of digital, Internet and data technologies in contemporary cultural practices associated with music reception, but have simultaneously explored the use of computational techniques in popular music studies research. Through this I have argued that a greater practical understanding and critical engagement with these technologies is possible, both for music consumers and popular music scholars, and I have demonstrated that my work represents a significant step towards that.

***

As my meeting within the unfamiliar department at the unfamiliar University drew to a close, I agreed to share with the professor and his team some of the data I had collected. This data comprises not only of the original texts supplied by respondents, but also the additional variables now associated with them that were produced through my process of computational analysis.

What may come from such an exchange? Where does this (at present) unofficial collaboration take us? These are the interesting questions that make meetings such as this worthwhile. My hope and expectation is that these researchers, with their greater understanding of computational processes and analytical techniques, may find interesting avenues within the data that would neither occur nor be available to me. They may make suggestions for additional data collection, or different types of analysis, or indeed different types of questions. This may lead to a more official, structured collaboration at a future point.

Equally, of course, there is also the chance that it may not lead anywhere at all. Indeed at two other points within the process of working through this project I have embarked upon similar attempted collaborations. One was with a commercial organisation that was attempting to launch a new music-related online product and who were interested in the results of my research, another was with an ad-hoc group of professional data scientists located in various countries who had expressed an interest in analysing the data I had collected. Neither of these collaborations came to fruition. What drives me in these attempts to facilitate collaborations of this kind, however, is the huge potential I see in The Harkive Project, and the fact that my ambition for it is not (at present) matched by my technical capabilities. My entire thesis can be understood in part as an attempt to reconcile these things.

Although I am extremely proud of the achievements I have made in terms of acquiring new technical skills – I have, from a standing start, learned the rudiments of data science and have deployed them in a real-world research

context – I am nevertheless frustrated by my inability to have developed the sufficient skills to have taken this project further than I have managed, certainly in terms of the technical perspective that I have placed at the centre of my enquiry. The constraints of a 3-year period in which to complete my work is of course a mitigating factor here, but nevertheless a significant regret is that I did not enter a period of self-directed learning around data analysis techniques much earlier in the process. I do wonder how much further I may have taken this research had I have begun that process sooner. That being said, I have developed numerous skills that I did not possess at the beginning of this process and it is precisely these skills that will form the foundation for what I foresee as a lifetime of further learning in this area. This will involve continued immersion in and use of computational analytical techniques alongside the continual pursuit of collaborations, both official and unofficial, in order to facilitate work that would always take as its starting point the critical perspectives that are so central to the humanities. I have, if nothing else, made progress in terms of forging an approach that has attempted to consider how these things may work productively in unison.

That the work undertaken in this doctoral project evolved from a previous project developed during my Masters studies has been instrumental in the direction I have taken. This can be understood in simple terms by the observation that the data used in my analysis preceded the formulation of my research question. This is, in terms of a research thesis, perhaps the opposite of what one would normally expect to happen in terms of project design. Questions are usually formulated first, based on either an engagement with literature in the field that seeks to highlight important theoretical gaps, or else through changes in the cultural, economic, technological or social landscape of a field that help suggest new, updated, or reformulated questions. Each scenario, or more usually a combination of the two, leads the researcher to a project that can be seen to produce the desired outcome: the furtherance of collective knowledge within a given field. Research design necessarily follows from such a position, leading to methods of data collection and analysis that are, the researcher hopes, sufficiently robust to be able to answer specific, carefully formulated questions.

My work has proceeded in a different route to that described above. At the point at which this work began, there already existed a large dataset that contained within it rich textual descriptions of the experiences of over a thousand people, gathered across four years. In other words, data collection was not informed by a carefully formulated research question. Instead, a consideration of the data collected and the conditions within which it was generated (i.e. the everyday lives of those contributing) was used as a springboard for my engagement with the numerous issues of debate and method that have informed the creation of a methodology that in turn led to the formulation of a research question. This is a hugely important point and one on which my entire project and research activity has pivoted. This thesis has developed through a process that encapsulated the initial challenges of the project: how could I devise a useful way of making sense of my data; and how could I do so in such a way that generated the requisite contribution to knowledge in the field of popular music studies. Because of that, the research question of this project was amongst the final pieces of the jigsaw to emerge, not the first.

The process, then, was not one where data would be collected in such a manner that it could be analysed through a pre-determined theoretical framework or methodology. Instead it was a process that took data as its starting point and auditioned a set of experimental analytical techniques in order to lead the way through issues of debate and method. Along the way, those issues became as much about popular music as they did about the process through which data and related practices have epistemological and ontological consequences. Those debates became inter-twinned in my imagination and, fortunately for my purposes, both could be explored through the thousands of stories Harkive had gathered, and the means through which they were examined. This has enabled me – I argue – to make a contribution to knowledge that on the one hand remains 'within' the field of popular music studies, but that also demonstrates potential efficacy if applied the others.

At various points throughout the completion of work, however, I have wrestled with the complexity and difficulty of my undertaking. I suspect, although will perhaps never know, that a different approach may have been an easier route. I suspect, although equally will never know, that my chosen approach has produced in many ways a better outcome. That I am left with not only more questions, but also with ideas, new analytical skills, considerable impetus, and – perhaps most important of all - a voracious appetite for pursuing my work further, goes some way to vindicate my chosen path.

I have regularly asked myself to what extent I am still (or ever was) operating within my home discipline of popular music studies. I have questioned whether taking a more nomadic approach that carries ideas from popular music studies to and from the fields of data science, digital humanities, the commercial field of popular music, and elsewhere, has been the best method. What has consistently kept me going through these feelings of confusion, these crises of confidence in my technical and intellectual abilities, was precisely the strangeness: if I was feeling alone, then perhaps I was doing something right, and perhaps I was on the path to something new and sufficiently unique. In the final analysis I realise that all of the above feelings were correct to varying extents, and the outcome has been a move towards a new form of popular music studies upon which I can make a significant imprint. Recalling the early days of my MA studies, back in 2011, some advice from tutors regarding the arcane world of academia still resonates: You must, I was told, carve out and vigorously defend an area that is your own. I feel confident in defending the manner and extent to which I have done that.

What is further interesting about the somewhat back-to-front process I have been through, where questions were the end product rather than the means, is that upon reflection it more closely resembles the present day conditions through which data-derived knowledge is created. Although this was the result of a combination of expediency, of the limitations of my initial design of The Harkive Project, of my personal and professional background, ultimately it has been advantageous and instrumental to the longer-term development of my

future as a scholar. As my engagement with literature from various academics has enabled me to observe, the recent emergence of technologies that facilitate the collection and analysis of data at huge scale is altering previously held ideas about what research is and how it can and should be done.

The wider context within which new forms of research can be undertaken is also that which provides the argument for its necessity. Rather than collecting datasets that are characterised by a limited number of variables, for specific, predefined purposes, that are contributed to by publics in full or some knowledge of their participation, we are currently in the midst of a paradigm shift that instead is characterised by the collection of huge amounts of data, about many aspects of everyday life, all of the time, and which is often collected in ways that does not require or seek the explicit permission of publics. In other cases the data collected is itself a by-product of other activities, or is a creation based on analysis of extant data. In each of these scenarios, then, the data often comes first and the 'use cases' then follow. From these large and ever-growing datasets, it is exploratory analysis of trends, clusters and patterns that lead to questions, and not the other way around. Related to this, the goal of commercial and governmental data collection and analysis can shift from the explanatory and descriptive, and towards the realm of the predictive. I have demonstrated numerous examples of exactly such activity in the field of popular music, where experiential product features such as algorithmically generated curatorial tools, are deployed at scale in the pursuit of a competitive advantage. Similar models now preside over many elements of everyday life, sometimes concerning more important elements of our existence than the ways we engage with music, and can suggest consequences that may not always be benign. This represents a significant epistemological and ontological shift. It is precisely this shift, I have argued, that we need to attempt to understand and I have shown that my work makes a small but significant step in that direction.

BIBLIOGRAPHY

Adorno, T.W. and Simpson, G. (1942) On popular music. *Zeitschrift für Sozialforschung*, 9(1), pp.17-48.

Allington, D., Brouillette, S. and Golumbia, D. (2016) Neoliberal tools (and archives): A political history of digital humanities. *LA Review of Books*, pp.1-05.

Amatriain, X. (2013) Mining large streams of user data for personalized recommendations. *ACM SIGKDD Explorations Newsletter*, 14(2), pp.37-48.

Ananny, M. (2016) Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science, Technology, & Human Values*, 41(1), pp.93-117.

Anderson, C. (2007) *The long tail: How endless choice is creating unlimited demand*. Random House.

Andrejevic, M. (2014) Big data, big questions| the big data divide. *International Journal of Communication*, 8, p.17.

Andriessen, W. (1999) 'THE WINNER'; compact cassette. A commercial and technical look back at the greatest success story in the history of AUDIO up to now. *Journal of magnetism and magnetic materials*, 193(1-3), pp.11-16.

Auerbach, E. and Said, E.W. (2013) *Mimesis: The representation of reality in Western literature*. Princeton University Press.

Banks, M. (2007) *The politics of cultural work*. Springer.

Barnes, S.B. (2006) A privacy paradox: Social networking in the United States. *First Monday*, 11(9).

Bartmanski, D. and Woodward, I. (2015) Vinyl: The analogue record in the digital age. Bloomsbury Publishing.

Baym, N.K. (2010) *Personal connections in the digital age*. Polity.

Belk, R. (2014) You are what you can access: Sharing and collaborative consumption online. *Journal of Business Research*, 67(8), pp.1595-1600.

Benjamin, W. (1931) Unpacking my library: A talk about book collecting. *Illuminations,* ed. Hannah Ardendt, trans. Harry Zohn, New York: Harcourt:Brace.

Benjamin, W. (2008) *The work of art in the age of mechanical reproduction*. Penguin UK.

Berry, D.M. (2011) The computational turn: Thinking about the digital humanities. *Culture Machine*, 12.

Bhattacharjee, S., Gopal, R.D., Lertwachara, K., Marsden, J.R. and Telang, R. (2007) The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science*, 53(9), pp.1359-1374.

Bijsterveld, K. and van Dijck, J. eds. (2009) *Sound souvenirs: audio technologies, memory and cultural practices (Vol. 2).* Amsterdam University Press.

Blei, D.M. (2012) Topic modeling and digital humanities. *Journal of Digital Humanities.* 2, 8–11.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), pp.993-1022.

Bodle, R. (2011) Regimes of sharing: Open APIs, interoperability, and Facebook. *Information, Communication & Society*, 14(3), pp.320-337.
Bogdanovic, D. and Longhurst, B.J. (2014) *Popular music and society*. Wiley.
Borges, J.L. (1998) On the exactitude of science. *Collected Fictions. Translated by Andrew Hurley*. New York: Penguin, p.325.
Bourdieu, P. (1984) *Distinction: A social critique of the judgement of taste*. Harvard university press.
Bowers, S. (2015) *Apple's cash mountain grows to $178bn.* [ONLINE] Available at: https://www.theguardian.com/business/2015/feb/02/apple-cash-mountain-grows. [Accessed 10 February 2018].
Boyd, D. and Crawford, K. (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society,* 15(5), pp.662-679.
Boyd, D. (2014) *It's complicated: The social lives of networked teens*. Yale University Press.
Bragg, B. (1992) *Mother of the Bride*. Sony/ATV Music Publishing LLC
Brown, J.D. (2011) Likert items and scales of measurement. *Statistics*, 15(1), pp.10-14.
Bucher, T. (2016) The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), pp.30-44.
Bull, M. (2000) *Sounding out the city: Personal stereos and the management of everyday life*. Berg.
Bull, M. (2006) Investigating the culture of mobile listening: From Walkman to iPod. *Consuming music together* (pp. 131-149). Springer, Dordrecht.
Bull, M. (2009) The auditory nostalgia of iPod culture. *Sound Souvenirs: Audio Technologies. Memory and Cultural Practices*, pp.83-93. Amsterdam University Press
Burkart, P. and McCourt, T. (2006) *Digital music wars: Ownership and control of the celestial jukebox*. Rowman & Littlefield.
Burkart, P. (2014) Music in the Cloud and the Digital Sublime. *Popular Music and Society*, 37(4), pp.393-407.
Caplan, L. (2016) Method without methodology: Data and the digital humanities. *e-flux journal*.
Chanan, M. (1995) *Repeated takes: A short history of recording and its effects on music.* Verso.
Chen, Y., Kreulen, J., Campbell, M. and Abrams, C. (2011) Analytics ecosystem transformation: A force for business model innovation. *SRII Global Conference (SRII), 2011 Annual,* pp. 11-20.
Cheney-Lippold, J. (2011) A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28(6), pp.164-181.
Cheney-Lippold, J. (2017) *We are data: Algorithms and the making of our digital selves*. NYU Press.
Chu, Z., Gianvecchio, S., Wang, H. and Jajodia, S. (2012) Detecting automation of twitter accounts: Are you a human, bot, or cyborg?. *IEEE Transactions on Dependable and Secure Computing*, 9(6), pp.811-824.
Chun, W.H.K., Grusin, R., Jagoda, P. and Raley, R. (2016) The dark side of the digital humanities. *Debates in the Digital Humanities 2016*, pp.493-509.

Coates, N. (2016) The age of discovery. *Popular Communication*, 14:3,192?193, DOI: 10.1080/15405702.2016.1196363
Cohen, L. (2003) *A Consumers' Republic.* New York: Alfred a Knopf Inc.
Collins, H., Evans, R. and Gorman, M. (2007) Trading zones and interactional expertise. *Studies in History and Philosophy of Science* Part A, 38(4), pp.657-666.
Danietz, J. (2016) *The Fundamental Limits of Machine Learning - Facts So Romantic* - [ONLINE] Available at: http://nautil.us/blog/the-fundamental-limits-of-machine-learning?utm_content=buffer9ffc0&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer. [Accessed 10 February 2018].
Dant, T. (2003) *Critical social theory: Culture, society and critique*. Sage.
Dasu, T. and Johnson, T. (2003) *Exploratory data mining and data cleaning*. John Wiley & Sons.
Datta, H., Knox, G. and Bronnenberg, B.J. (2016) Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Working Paper.*
De Certeau, M. (1984) The practice of everyday life. Berkeley.
DeNora, T. (2000) *Music in everyday life*. Cambridge University Press.
Dredge, S. (2016) *Matt Ogle talks Discover Weekly and Spotify evolution.* [ONLINE] Available at: http://musically.com/2016/03/21/matt-ogle-discover-weekly-spotify/. [Accessed 10 February 2018].
Edwards, L., Klein, B., Lee, D., Moss, G. and Philip, F. (2015) Discourse, justification and critique: towards a legitimate digital copyright regime?. *International Journal of Cultural Policy,* 21(1), pp.60-77.
Felski, R. (1999) The invention of everyday life. *New formations*, (39), pp.13-31.
Feyerabend, P. (2001) *Conquest of abundance: a tale of abstraction versus the richness of being*. University of Chicago Press.
Finnegan, R. (2007) *The hidden musicians: Music-making in an English town.* Wesleyan University Press.
Forde, E. (2014) *Apple reportedly in crisis over iTunes and its future music plans* [ONLINE] Available at: http://musically.com/2014/04/10/apple-reportedly-in-crisis-over-itunes-and-its-future-music-plans/. [Accessed 10 February 2018].
Fox, M. (2004) E-commerce business models for the music industry. *Popular Music and Society*, 27(2), pp.201-220.
Frith, S. (1987) The industrialization of popular music. *Popular music and communication,* pp.53-77.
Frith, S. (2002). Music and everyday life. *Critical Quarterly*, 44(1), pp.35-48.
Gardiner M.E. (2000) *Critiques of Everyday Life*. NY: Routledge.
Gerlitz, C. and Helmond, A. (2013) The like economy: Social buttons and the data-intensive web. *New Media & Society*, 15(8), pp.1348-1365.
Gerlitz, C. (2016) What counts? Reflections on the multivalence of social media data. *Digital Culture & Society*, 2(2), pp.19-38.
Gilbert, N. and Troitzsch, K. (2005) *Simulation for the social scientist*. McGraw-Hill Education (UK).
Gitelman, L. ed. (2013) *Raw data is an oxymoron.* MIT Press.
Gomart, E. and Hennion, A. (1999) A sociology of attachment: music amateurs, drug users. *The Sociological Review*, 47(S1), pp.220-247.

Greene, P.D. and Porcello, T. eds., (2005) *Wired for sound: Engineering and technologies in sonic cultures.* Wesleyan University Press.
Hagen, A.N. (2015) The playlist experience: Personal playlists in music streaming services. *Popular Music and Society*, 38(5), pp.625-645.
Hagen, A.N. (2016)The metaphors we stream by: Making sense of music streaming. *First Monday*.
Haider, M. (2015) *Getting Started with Data Science: Making Sense of Data with Analytics*. IBM Press.
Hall, G. (2013) Toward a postdigital humanities: Cultural analytics and the computational turn to data-driven scholarship. *American Literature*, 85(4), pp.781-809.
Hall, S. (2001) Encoding/decoding. *Culture, Media, Language*, ed. Hall et al., London: Hutchinson, 128-140.
Hartmann, P.M., Zaki, M., Feldmann, N. and Neely, A. (2014) Big data for big business? A taxonomy of data-driven business models used by start-up firms. *A Taxonomy of Data-Driven Business Models Used by Start-Up Firms* (March 27, 2014).
Hebdige, D. (1979) *Subculture: The Meaning of Style*, New York, Methuen
Hendy, D. (2000) Pop music radio in the public service: BBC Radio 1 and new music in the 1990s. *Media, Culture & Society*, 22(6), pp.743-761.
Hendy, D. (2013) *Noise: a human history of sound and listening.* Profile Books.
Hesmondhalgh, D. (1998) The British dance music industry: a case study of independent cultural production. *British Journal of Sociology*, pp.234-251.
Hesmondhalgh, D. (2002) Popular music audiences and everyday life. In David Hesmondhalgh and Keith Negus (eds) *Popular Music Studies,* London: Arnold, 117–130.
Hey, T., Tansley, S. and Tolle, K.M. (2009) *The fourth paradigm: data-intensive scientific discovery (Vol. 1).* Redmond, WA: Microsoft research.
Heye, A. and Lamont, A. (2010) Mobile listening situations in everyday life: The use of MP3 players while travelling. *Musicae Scientiae*, 14(1), pp.95-120.
Hine, T. (1986) *Populuxe*. Alfred a Knopf Inc.
Hogan, M. (2015) *Bandcamp CEO Explains Payment Policy Changes* [ONLINE] Available at: https://www.billboard.com/articles/business/6738386/bandcamp-ceo-payment-policy. [Accessed 10 February 2018].
Housley, W., Procter, R., Edwards, A., Burnap, P., Williams, M., Sloan, L., Rana, O., Morgan, J., Voss, A. and Greenhill, A. (2014) Big and broad social data and the sociological imagination: A collaborative response. *Big Data & Society*, 1(2), p.2053951714545135.
IFPI (2015) *IFPI publishes Recording Industry in Numbers - an essential guide to the global music market.* [ONLINE] Available at: http://www.ifpi.org/news/IFPI-publishes-Recording-Industry-in-Numbers-2015. [Accessed 10 February 2018].
IFPI (2016) *IFPI Global Music Report 2016.* IFPI [ONLINE]Available at: http://www.ifpi.org/downloads/GMR2016.pdf [Accessed 10 February 2018].
Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3), pp.299-314.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) *An introduction to statistical learning*. New York: Springer.
Jenke, T. (2017) *Apple is reportedly set to buy Shazam at a bargain basement price*. [ONLINE] Available at: https://www.theindustryobserver.com.au/apple-buy-shazam-bargain-price/. [Accessed 10 February 2018].
Jockers, M. (2015). Revealing sentiment and plot arcs with the syuzhet package. [ONLINE] Available at: http://www.matthewjockers.net/2015/02/02/syuzhet/ [Accessed 10 February 2018]
Jones, S. (2000) Music and the Internet. *Popular Music*, 19(2), pp.217-230.
Kassabian, A. (2002) Ubiquitous listening. In David Hesmondhalgh and Keith Negus (eds) *Popular Music Studies,* London: Arnold, 131-142.
Kassabian, A. (2013) *Ubiquitous listening: Affect, attention, and distributed subjectivity*. University of California Press.
Khawaja, J, (2016) *The Rise and Fall—and Resurrection?—of Beatport* [ONLINE] Available at: https://thump.vice.com/en_au/article/bmadnw/the-rise-and-falland-resurrectionof-beatport. [Accessed 10 February 2018].
Kitchin, R. (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1).
Kitchin, R. (2014) *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
Kjus, Y. (2016) Musical exploration via streaming services: The Norwegian experience. *Popular Communication*, 14(3), pp.127-136.
Kramer, A.D., Guillory, J.E. and Hancock, J.T. (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24), pp.8788-8790.
Krause, A.E., North, A.C. and Hewitt, L.Y. (2015) Music-listening in everyday life: Devices and choice. *Psychology of Music*, 43(2), pp.155-170.
Kusek, D., Leonhard, G., Lindsay, S.G. (2005) *The future of music: Manifesto for the digital music revolution*. Berklee Press Boston, MA.
Laing, D. (1991) A Voice without a Face: Popular Music and the Phonograph in the 1890s. *Popular Music*, 10(1), pp.1-9.
Lakoff, G. and Johnson, M. (2008) *Metaphors we live by*. Univ. of Chicago press.
Laney, D. (2001) *3D data management: controlling data volume, velocity and variety*. Gartner. Stamford, CT, USA.
Latour, B. (2005) *Reassembling the social: An introduction to actor-network-theory*. Oxford university press.
Lears, J. (1995) *Fables of abundance: A cultural history of advertising in America.* Basic Books.
Lefebvre, H. (1991) *The production of space*. Oxford Blackwell.
Lim, K.W. and Buntine, W. (2014) Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (pp. 1319-1328). ACM.
Lin, J. and Ryaboy, D. (2013) Scaling big data mining infrastructure: the twitter experience. *ACM SIGKDD Explorations Newsletter*, 14(2), pp.6-19.

Liu, A., (2016) *Drafts for Against the Cultural Singularity - (book in progress).* [ONLINE] Available at: http://liu.english.ucsb.edu/drafts-for-against-the-cultural-singularity [Accessed 10 February 2018].
Liu, B. (2010) Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, 2, pp.627-666.
Lynch, M.P. (2016) *The Internet of us: Knowing more and understanding less in the age of big data*. WW Norton & Company.
Magaudda, P. (2011) When materiality 'bites back': Digital music consumption practices in the age of dematerialization. *Journal of Consumer Culture*, 11(1), pp.15-36.
Manovich, L. (2001) *The language of new media*. MIT press.
Manovich, L. (2007) Cultural analytics: Analysis and visualization of large cultural data sets. Manovich.net, 30.
Manovich, L. (2016) The science of culture? Social computing, digital humanities and cultural analytics. *The datafield society. social research in the age of Big Data*. Amsterdam: Univ. Press.
Markham, A.N. (1998) *Life online: Researching real experience in virtual space*. Rowman Altamira.
Markham, A.N. (2003) Metaphors reflecting and shaping the reality of the Internet: Tool, place, way of being. *Association of Internet Researchers Conference, Toronto*, Canada (pp. 16-19).
Mason, M. (2009) *The pirate's dilemma: How youth culture is reinventing capitalism*. Simon and Schuster.
Mayer-Schönberger, V., Cukier, K. (2013) *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.
McCourt, T. and Rothenbuhler, E.W. (2004) Burnishing the brand: Todd Storz and the total station sound. *Radio Journal: International Studies in Broadcast & Audio Media*, 2(1), pp.3-14.
McLeod, K. (2005) MP3s are killing home taping: The rise of Internet distribution and its challenge to the major label music monopoly. *Popular Music and Society*, 28(4), pp.521-531.
Mejia, R. (2012) Walking in the City'in an age of Mobile Technologies. *Race/Gender/Class/Media*, 3, pp.113-118.
Menn, J. (2003) *All the rave: the rise and fall of Shawn Fanning's Napster*. Crown Business.
Michael, M. and Lupton, D. (2016) Toward a manifesto for the 'public understanding of big data'. *Public Understanding of Science*, 25(1), pp.104-116.
Milner, G. (2010) *Perfecting sound forever: the story of recorded music.* Granta, London.
Moretti, F. (2013) *Distant reading*. Verso Books.
Morris, J.W. and Powers, D. (2015) Control, curation and musical experience in streaming music services. *Creative Industries Journal*, 8(2), pp.106-122.
Morris, J.W. (2015) Curation by code: Infomediaries and the data mining of taste. *European Journal of Cultural Studies*, 18(4-5), pp.446-463.
Murphy, G. (2014) *Cowboys and Indies: The Epic History of the Record Industry*. Macmillan.
Negus, K. (1997) *Popular music in theory: An introduction*. Wesleyan University Press.

Nissenbaum, H. (2009) *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
Nowak, R. and Bennett, A. (2014) Analysing everyday sound environments: The space, time and corporality of musical listening. *Cultural Sociology*, 8(4), pp.426-442.
Nowak, R. (2014) Investigating the interactions between individuals and music technologies within contemporary modes of music consumption. *First Monday*, 19(10).
Nowak, R. (2016) Music Listening Activities in the Digital Age: An Act of Cultural Participation through Adequate Music. *Leonardo Music Journal*, pp.20-23.
Nowak, R. (2016) When is a discovery? The affective dimensions of discovery in music consumption. *Popular Communication*, 14(3), pp.137-145.
O'Neil, C. (2017) *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
Ohmann, R.M. (199) *Selling culture: Magazines, markets, and class at the turn of the century*. Verso.
Palmer, T. (1977) *All you need is love: The story of popular music*. Penguin (Non-Classics).
Pasnick, A. (2015) *The magic that makes Spotify's Discover Weekly playlists so damn good.* [ONLINE] Available at: https://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good/. [Accessed 10 February 2018].
Piper, A. (2016) There will be numbers. *CA: Journal of Cultural Analytics*, 1(1).
Prey, R. (2015) Henri Lefebvre and the Production of Music Streaming Spaces. *Sociologica*, 9(3)
Puschmann, C. and Burgess, J. (2014) Big data, big questions| Metaphors of big data. *International Journal of Communication*, 8, p.20.
Rieder, B. (2016) Big Data and the Paradox of Diversity. *Digital Culture & Society*, 2(2), pp.39-54.
Rose, B. & Eliscu, E. (1929) *Without A Song*. W.B. Music Corp.
Russell, D. (1997) *Popular music in England 1840-1914: a social history.* Manchester University Press.
Sandvig, C. and Hargittai, E. (2015) How to Think about Digital Research. *Digital research confidential: The secrets of studying behavior online.* MIT Press.
Sanjek, R. (1988) *American Popular Music and Its Business: From 1900 to 1984 (Vol. 3).* Oxford University Press on Demand.
Savage, M. and Burrows, R. (2007) The coming crisis of empirical sociology. *Sociology*, 41(5), pp.885-899.
Silver, J. (2013) *Digital Medieval*. Xstorical Publications
Sinclair, G. and Tinson, J. (2017) Psychological ownership and music streaming consumption. *Journal of Business Research*, 71, pp.1-9.
Sloboda, J. A. (1992) Empirical studies of emotional response to music. In M. R. Jones & S. Holleran, (eds), *Cognitive bases of musical communication*, pp. 33-46.
Small, C. (1998) *Musicking: the meanings of performing and listening*. Hanover, Wesleyan University Press.
Smith, B.H. (1979) Fixed Marks and Variable Constancies: A Parable of Literary Value. *Poetics Today*, 1(1/2), pp.7-22.

Smithies, J.D. (2014) Digital humanities, postfoundationalism, postindustrial culture. *DHQ*, Vol.8 n.1
Southall, B. (2009) *The Rise & Fall of EMI Records*. Omnibus Press & Schirmer Trade Books.
Sterne, J. (2006) The death and life of digital audio. *Interdisciplinary Science Reviews*, 31(4), pp.338-348.
Sterne, J. (2006) The mp3 as cultural artifact. *New media & society*, 8(5), pp.825-842.
Sterne, J. (2006) What's digital in digital music. *Digital media: Transformations in human communication*, pp.95-108.
Sterne, J. (2009) The preservation paradox in digital audio. *Sound Souvenirs*, pp.55-65.
Sterne, J. (2012) *MP3: The meaning of a format.* Duke University Press.
Straw, W. (1997a) 'Organized Disorder': The Changing Space of the Record Shop. *The Clubcultures Reader*, pp.57-65.
Straw, W. (1997b) Sizing up record collections. *Sexing the groove: Popular music and gender*, 3, p.16.
Swafford, A. (2015) Problems with the syuzhet package. Anglophile in Academia: Annie Swafford's Blog. [ONLINE] Available at: http: https://annieswafford.wordpress.com/2015/03/02/syuzhet/ [ACCESSED 10 February 2018]
Sweney, M. (2017) *Music industry goes to war with YouTube* [ONLINE] Available at: https://www.theguardian.com/business/2017/apr/15/music-industry-youtube-video-streaming-royalties. [Accessed 10 February 2018].
Thompson, D. (2014) *The Shazam Effect - The Atlantic.* [ONLINE] Available at: https://www.theatlantic.com/magazine/archive/2014/12/the-shazam-effect/382237/. [Accessed 10 February 2018].
Thrift, N. (2005) *Knowing capitalism*. Sage.
Tufekci, Z. (2015) Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal* 13 (2015b).
van Dijck, J. and Olsson, T. (2013) Social media platforms as producers. *Producing the internet: Critical perspectives of social media,* pp.45-62.
Van Dijck, J. (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), p.197.
Vanderbilt, T. (2016) *You May Also Like: Taste in an Age of Endless Choice*. Simon and Schuster.
Wall, T. (2013) *Studying popular music culture*. Sage.
Weber, H. (2009) Taking Your Favorite Sound Along: Portable Audio Technologies for Mobile Music Listening. *Sound souvenirs: Audio technologies, memory and cultural practices,* 2, p.69.
Weber, W. (1977) Mass culture and the reshaping of European musical taste, 1770-1870. *International review of the aesthetics and sociology of music*, pp.5-22.
Webster, J., Gibbins, N., Halford, S. and Hracs, B.J. (2016) Towards a theoretical approach for analysing music recommender systems as sociotechnical cultural intermediaries. *In Proceedings of the 8th ACM Conference on Web Science* (pp. 137-145). ACM.
White, D.S. and Le Cornu, A. (2011) Visitors and Residents: A new typology for online engagement. *First Monday*, 16(9).

Wickham, H. (2014) Tidy data. *Journal of Statistical Software*, 59(10), pp.1-23.
Wikström, P. (2013) *The music industry: Music in the cloud.* Polity.
Williams, R. (2011) Culture is ordinary (1958). *Cultural theory: An anthology*, pp.53-59.
Williams, R. (2014). *Keywords: A vocabulary of culture and society*. Oxford University Press.
Witt, S. (2015) *How music got free: A story of obsession and invention.* Penguin.
Young, S. and Collins, S. (2010) A View from the Trenches of Music 2.0. *Popular Music and Society*, 33(3), pp.339-355.
Zimmer, M. (2010) "But the data is already public": on the ethics of research in Facebook. *Ethics and information technology*, 12(4), pp.313-325.
Zimmer, M. (2015) The Twitter Archive at the Library of Congress: Challenges for information practice and information policy. *First Monday*, 20(7).

Appendices

Appendix A - Harkive Stories Database Schema

Full list of variables in the Harkive story data base and the API elements of different services used to popular them. Where API elements are not available, N/A will be shown.

source

NB: is populated by a field added at the API collection stage.

Source	Element
Twitter	Twitter
Tumblr	Tumblr
Instagram	Instagram
Facebook	Facebook
Email	Email
Form	Form

story

Source	Element
Twitter	Text
Tumblr	Body
Instagram	Caption_Text
Facebook	Message
Email	Gsx$Story
Form	Gsx$YourHarkiveStory

screen_name

Source	Element
Twitter	User_screen_name
Tumblr	Blog_name
Instagram	User_username
Facebook	From_name
Email	N/A
Form	Gsx$Name

real_name

Source	Element
Twitter	User_name
Tumblr	Blog_name
Instagram	User_full_name
Facebook	From_name
Email	Gsx$Fromemail
Form	Gsx$Name

user_id

Source	Element
Twitter	User_id
Tumblr	Id
Instagram	User_id
Facebook	From_id
Email	N/A
Form	N/A

story_url

Source	Element
Twitter	url
Tumblr	Post_url
Instagram	Link
Facebook	Link
Email	N/A
Form	N/A

story_id

Source	Element
Twitter	Id
Tumblr	Id
Instagram	Id
Facebook	Id
Email	Id
Form	Id

summary

Source	Element
Twitter	N/A
Tumblr	Summary
Instagram	N/A
Facebook	N/A
Email	Gsx$Subject
Form	N/A

tags

Source	Element
Twitter	N/A
Tumblr	Tags
Instagram	Tags

Facebook	N/A
Email	N/A
Form	N/A

date

Source	Element
Twitter	Created_at
Tumblr	Date
Instagram	Date
Facebook	Updated_time
Email	Date
Form	Date

user_profile_photo

Source	Element
Twitter	User_profile_image_url
Tumblr	Photo
Instagram	User_profile_picture
Facebook	Picture
Email	N/A
Form	N/A

user_location

Source	Element
Twitter	User_location
Tumblr	N/A
Instagram	Location
Facebook	N/A
Email	N/A
Form	Gsx$Town, Gsx$Country

latitude/longitude

Source	Element
Twitter	Geo_cordinates
Tumblr	N/A
Instagram	Location
Facebook	N/A
Email	N/A
Form	N/A

story_image

Source	Element

Twitter	N/A
Tumblr	Photo
Instagram	Images_standard_resolution_url
Facebook	Picture
Email	N/A
Form	N/A

Appendix B – Sample R Script

Below is an example R script I have written to perform a particular set of tasks:

- Load in survey data
- Convert responses to questions from text to numeric values
- Perform basic mathematics
- Produce new file of converted data
- Create visualisations of the data
- Produce image files of visualisation

NOTES:

Ordinarily this script would be saved as a file with the extension .R

This script needs the corresponding data set to run.

Any line in the script that begins with ## and coloured green is a comment, and not a command. R automatically skips over lines beginning with #, so they can remain in the script without issue. All other text represents actionable code.

You will note that each step of the process is broken into discrete parts, so the elements can be amended according to the needs of the researcher without stopping the rest of the script from functioning. For example, numeric values assigned to text-based responses can be changed, whilst all other code can remain the same. The script will still run.

R works through scripts line-by-line. If the script encounters an error (e.g. a wrongly coded command), a red warning message will be displayed in the console describing the issue. Until the issue in that offending line is resolved the script will not run any further, even if subsequent lines are coded correctly.

The script below runs from start to finish in around 2 seconds. Scripts that are designed to process large amounts of data, or more commands, can take longer, but R is extremely efficient and powerful and can handle often very complex sets of commands in seconds.

```
## - - HARKIVE 2016 MUSIC LISTENING SURVEY - - ##
##CONTENTS OF DATABASE
##IDENTIFIER
## - - R.Num
##DEMOGRAPHICS
## - -Q1: Gender
## - -Q2: Age
## - -Q5: Harkive Story - Yes/No
##FORMATS
# - - Q6: MP3/Digital Files
# - - Q7: Streaming (Spotify, YouTube, Soundcloud, etc)
# - - Q8: Radio (incl. Online, Analogue, Catch-up)
```

```r
# - - Q9: Podcasts (incl. Mixcloud, etc)
# - - Q10: Physical Formats (CD, Vinyl, Tape, etc)
# - - Q11: Live Music
# - - Q12: Performing/Making Music
# - - Q14: Your Favourite
# - - Q15: Your Least Favourite
# - - Q16: Convenience
# - - Q17: Cost
# - - Q18: Portability
# - - Q19: Sociability/Sharing
# - - Q20: Sound Quality
# - - Q21: Collectibility
# - - Q22: Exclusivity/Rarity

##Load in survey data and assign to a variable h16

h16 <- read.csv("h16_sample.csv")

## Responses Q6:Q12 now to be converted from text to numerics
## Never (0) to Daily (6)
## 'blank' responses to be graded as (0)

#Using dpylr package to create a function for this.
#Based on string matching, text observations are converted to
# corresponding numbers

library(dplyr)

factorise_likert <- function(x) {
  case_when(x %in% c("Daily") ~ 6,
            x %in% c("Very Often") ~ 5,
            x %in% c("Often") ~ 4,
            x %in% c("Neither Rarely nor Often") ~ 3,
            x %in% c("Rarely") ~ 2,
            x %in% c("Very Rarely") ~ 1,
            x %in% c("Never", "") ~ 0)
}

##create new numeric variables bases on text responses

n6 <- sapply(h16$q6, factorise_likert)
n7 <- sapply(h16$q7, factorise_likert)
n8 <- sapply(h16$q8, factorise_likert)
n9 <- sapply(h16$q9, factorise_likert)
n10 <- sapply(h16$q10, factorise_likert)
n11 <- sapply(h16$q11, factorise_likert)
n12 <- sapply(h16$q12, factorise_likert)

#bind these to dataframe
```

```
h16$n6 <- n6
h16$n7 <- n7
h16$n8 <- n8
h16$n9 <- n9
h16$n10 <- n10
h16$n11 <- n11
h16$n12 <- n12

##remove related text columns (q6:q12)

h16$q6 <- NULL
h16$q7 <- NULL
h16$q8 <- NULL
h16$q9 <- NULL
h16$q10 <- NULL
h16$q11 <- NULL
h16$q12 <- NULL

##use dyplr mutate to add columns together and get total
## across new values

h16 <- mutate(h16, n6_12sum = n6 + n7 + n8 + n9 + n10 + n11 +
n12)

##then calculate mean and SD of those

library(matrixStats)
group6_12 <- c('n6', 'n7', 'n8', 'n9','n10','n11','n12')
h16 <- h16 %>%
  mutate(mean6_12 = rowMeans(.[group6_12]), sd6_12 =
rowSds(as.matrix(.[group6_12])))

##add column for sum divided by sd

h16 <- mutate(h16, n6_12sumsd = n6_12sum/sd6_12)


#Plotting - create graph of the data above

library(ggplot2)
library(RColorBrewer)

h16 %>%
  filter(q1 != "Rather Not Say", q1 != "Other") %>%
  ggplot() +
  aes(x = n6_12sum, y = sd6_12, colour=q2, shape=q1) +
  geom_jitter() +
  ggtitle('Harkive Survey Q6:12 - Multi Format Engagement') +
```

```
  xlab('Regularity of Multi Format Engagement') +
  ylab('Standard Deviation') +
  scale_colour_discrete(name  ="Age") +
  scale_shape_discrete(name  ="Gender")

#Now we can apply the same process to Q14 and Q15.
##Note the function from above has been renamed
##to factorise_format, and the values changed

factorise_format <- function(x) {
  case_when(x %in% c("", "Other") ~ 'I' ,
            x %in% c("MP3/Digital Files") ~ 'A',
            x %in% c("Physical Formats") ~ 'B',
            x %in% c("Radio") ~ 'C',
            x %in% c("Live Music") ~ 'D',
            x %in% c("Performing/Making Music") ~ 'E',
            x %in% c("Podcasts") ~ 'F',
            x %in% c("Streaming") ~ 'G')
}

n14 <- sapply(h16$q14, factorise_format)
n15 <- sapply(h16$q15, factorise_format)
h16$n14 <- n14
h16$n15 <- n15
h16$q14 <- NULL
h16$q15 <- NULL


##Now Coding q16:22, with function now amended to
##factorise_importance

factorise_importance <- function(x) {
  case_when(x %in% c("Very Important") ~ 3,
            x %in% c("Important") ~ 2,
            x %in% c("Somewhat important") ~ 1,
            x %in% c("Neither Important nor Unimportant", "")
~ 0,
            x %in% c("Somewhat unimporant") ~ -1,
            x %in% c("Unimportant") ~ -2,
            x %in% c("Not important at all") ~ -3)
}

n16 <- sapply(h16$q16, factorise_importance)
n17 <- sapply(h16$q17, factorise_importance)
n18 <- sapply(h16$q18, factorise_importance)
n19 <- sapply(h16$q19, factorise_importance)
n20 <- sapply(h16$q20, factorise_importance)
n21 <- sapply(h16$q21, factorise_importance)
n22 <- sapply(h16$q22, factorise_importance)
```

```
h16$n16 <- n16
h16$n17 <- n17
h16$n18 <- n18
h16$n19 <- n19
h16$n20 <- n20
h16$n21 <- n21
h16$n22 <- n22

h16$q16 <- NULL
h16$q17 <- NULL
h16$q18 <- NULL
h16$q19 <- NULL
h16$q20 <- NULL
h16$q21 <- NULL
h16$q22 <- NULL

##creating total, mean, sd and total/sd columns for 16:22

h16 <- mutate(h16, n16_22sum = n16 + n17 + n18 + n19 + n20 +
n21 + n22)
group16_22 <- c('n16', 'n17', 'n18', 'n19','n20','n21','n22')
h16 <- h16 %>%
  mutate(mean16_22 = rowMeans(.[group16_22]), sd16_22 =
rowSds(as.matrix(.[group16_22])))
h16 <- mutate(h16, n16_22sumsd = n16_22sum/sd16_22)

##create new csv file so the above process does not
## have to be repeated in subsequent analysis

write.csv(h16, "h16_6to22_converted.csv")

#PLOT TOTAL OF 'Motivations' w/ Deviation across Motivations
##NOTE that filter applied to remove response to q1 (gender)
## that are not male or female. This is purely for
demonstration of filter function. At the end of the plotting
command a png file of the image is created and written to
working directory. Again, this is optional.

h16 %>%
  filter(q1 != "Rather Not Say", q1 != "Other") %>%
  ggplot() +
  aes(x = n17, y = n22, colour=q2, shape=q5) +
  geom_jitter() +
  ggtitle('Harkive Survey Q17 & Q22:\nFormat Choice
Motivations: \nCost v Exclusivity') +
  xlab('Cost') +
  ylab('Exclusivity') +
  scale_colour_discrete(name  ="Age") +
```

```
  scale_shape_discrete(name  ="Yes/No") +
  ggsave(file="n17-n22.png")

##At this point any interesting clusters can be visualised.
##Based on interesting clusters, stories can be pulled from
dbase of harkive stories
```

Appendix C – Harkive Music Listening Survey Questions

Full list of Harkive Music Listening Survey Questions, with brief explanations of questions and types of responses (where applicable)

Section 1: About You

- First Name
- Surname
- Gender (Q1)
- Age (Q2)
- Email
- Town (Q3)
- Country (Q4)

Section 2: Harkive Stories

- Have you submitted a story to Harkive, either in 2016 or previously? (Q5)

NB: Respondents answering No are taken to Section 3. Those answering Yes are asked the following:

- What method did you use?

NB: If respondents selected Twitter, Instagram, or Tumblr, they are asked to provide usernames.

Section 3: General Music Listening

Participants were asked to provide information about their everyday music listening by responding to Likert scale questions. The first of these pertained to regularity of their use of seven different formats/services/modes of listening, ranging from Never to Daily on a 7 point scale:

- Q6 - MP3/Digital Files
- Q7 - Streaming (Spotify, YouTube, Soundcloud, etc.)
- Q8 - Radio (incl. Online, Analogue, Catch-up)
- Q9 - Podcasts (incl. Mixcloud, etc.)
- Q10 - Physical Formats (CD, Vinyl, Tape, etc.)
- Q11 - Live Music
- Q12 - Performing/Making Music

Then which of the above methods were their:

- Q13 - Favourite method of listening
- Q14 - Least favourite method of listening

Motivations for selecting formats above were rated from Not Important to Very Important along a 7-point scale

- Q16 - Convenience
- Q17 - Cost
- Q18 - Portability
- Q19 - Sociability/Sharing
- Q20 - Sound Quality
- Q21 - Collectability
- Q22 - Exclusivity/Rarity

Level of engagement with music, the regularity of the following were rated from Never to Daily along a 7-point scale

- Q23 – Days when I listen to no music at all occur..
- Q24 – Days when I listen for less than 1 hour
- Q25 - 1-2 hours
- Q26 - 3-5 hours
- Q27 - More than 5 hours

Use of formats in the following situations, 7-point Likert scale ranging from Never to Daily

- Q28 - To Relax/wind down
- Q29 - Whilst doing something else (exercise, cooking, etc.)
- Q30 - Whilst working
- Q31 - Whilst driving/commuting
- Q32 - During social occasions (parties, socializing with friends)
- Q33 - I remember a song and will seek it out to play it
- Q34- I play some music as pure background/sonic wallpaper
- Q35 - When I am somewhere and cannot control the choice of music (a café, workplace, etc.)
- Q36 - When I'm alone
- Q37 - When I'm with others
- Q38 - To reminisce (remember someone/somewhere)
- Q39 - To listen to something new
- Q40 - To listen to something familiar
- Q41 - A song will pop into my head, or get 'stuck' in my mind

7-point Likert scale rating important of music in these situations, from Not Important At All to Very Important

- Q42 - To Relax/wind down
- Q43 - Whilst doing something else (exercise, cooking, etc.)
- Q44 - Whilst working
- Q45 - Whilst driving/commuting

- Q46 - During social occasions (parties, socializing with friends)
- Q47 - When I'm alone
- Q48 - To reminisce (remember someone/somewhere)

Section 4: Technology and Formats

7-point scale ranging from Strongly Disagree to Strongly Agree.

- Q48 - I almost always carry a portable music device (e.g. MP3 Player, SmartPhone)
- Q49 - My music listening habits have changed considerably in the last 10 years
- Q50 - The experience of listening to Physical formats (vinyl, CD) is different to Digital formats (streams, MP3s)
- Q51 - The experience of listening to Physical formats (vinyl, CD) is better than listening via Digital formats (streams, MP3s)
- Q52 - Digital Music Technologies (streaming services, MP3 stores, etc.) enhance my enjoyment of listening to music
- Q53 - Being able to easily find out additional information online (videos, band websites, biographies, reviews, etc.) is important to my enjoyment of music
- Q54 - I like to try new digital music services when they are first launched
- Q55 - It's important to me that I'm using the most up to date music services and devices
- Q56 - I listen to more music now than in the past because of digital technologies
- Q57 - Whether a format is physical or digital has no effect on my level of enjoyment when listening to music
- Q58 - Having access to more music, and more information, means I listen to a wider variety of music than I did previously
- Q59 - The currently available mix of digital and physical listening methods are perfectly suited to my needs

Section 5: Data, Privacy, Identity and Ownership

7-point Likert Scale ranging from Strongly Agree to Strongly Disagree

- Q60 - I am aware that data can be collected by companies/organisations about my online activity (e.g. web browsing, music listening, social media posts)
- Q62 - Data collection by companies/organisations about my online activity (e.g. web browsing, social media activity, music listening) is an invasion of my privacy
- Q63 - I always ensure that the privacy settings of my online services and social network accounts are set to the maximum possible settings

- Q64 - Sharing details of my music listening online (e.g. on social networks) is an important element of my enjoyment of music
- Q65 - I would like to have access to the data collected about me by companies and organisations
- Q66 - I would like to know more about how the data that is collected about me is used
- Q67 - The data collected about me is a necessary part of the exchange involved when I use certain online services
- Q68 - Data collection helps companies improve their services in ways that are beneficial to me
- Q69 - Were I to be given access to it, I feel I have sufficient technical knowledge to make sense of the data collected about me
- Q70 - I feel my 'digital self' is a close representation of my real-world self
- Q71 - My online playlists (e.g. in streaming services) are part of my wider music collection
- Q72 - I feel the same sense of ownership over my digital files and playlists as I do with my physical collection
- Q73 - I feel emotionally connected to my online/digital music collection in the same way as I do with my physical (CD, vinyl) collection
- Q74 - The data I generate about myself (e.g. when browsing, purchasing or listening online) is given over voluntarily
- Q75 - I trust that companies and organisations that collect data about me will keep it secure

Section 6: Recommendation and Discovery

7-point Likert scale ranging from Strongly Agree to Strongly Disagree

- Q76 - I often recommend music to my friends
- Q77 - I often have music recommended to me by friends online (via social media, email, etc.)
- Q78 - I often have music recommended to my by friends offline (face-to-face, social situations, etc.)
- Q79 - I seek out new music (through radio, print and online media)
- Q80 - I often listen to music recommended to me by online services
- Q81 - The online/automatic recommendations are often better than those I receive from friends
- Q82 - With so much music to choose from, I sometimes find it difficult to make a choice of what to listen to
- Q83 - I tend to stick with music that I already know
- Q84 - I find it easy to discover new music that I like without assistance
- Q85 - I am surprised by how accurately online music service recommendations reflect my taste
- Q86 - My online listening and browsing influences the type of music I am recommended by online services

- Q87 - Automatic recommendation systems are useful to my experience as a music listener
- Q88 - I sometimes engage in 'private' online listening so that my choices are not recorded
- Q89 - When I recommend a song to friends, I am helping to promote the band/artist
- Q90 - Having elements of my musical taste displayed online (e.g. in a shared playlist, or a link I post online) is important to me