Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Music and big data: a new frontier

David M Greenberg[1,2] and Peter J Rentfrow[3]

There is an unprecedented opportunity for psychologists and behavioral scientists to merge prior theory and research with big data to develop profound insights into the way people use and are affected by music. There are now **streaming services** that store data from millions of people on their day-to-day musical listening habits; **song-level** data that tags sonic and emotion attributes for millions of songs; **wearable devices** (e.g. watches and earbuds) that capture physiological metrics including heartrate and galvanic skin response; **mobile technologies** that track a person's moment-to-moment activity, location, mood, and sociability; and **survey instruments** and **digital footprints** that capture personality and other biopsychosocial metrics in just under a minute. We propose that merging these technologies can create a new age in music psychology that exponentially expands the present knowledge and scope of the field. The new data will advance general areas of music psychology, but will also provide an important opportunity to establish new knowledge about health and well-being that can have a direct impact on the public. By scientifically mapping how music changes behavior and health in the short-term and long-term, Big Music Data can lead to future health initiatives including the development of new evidence-based treatment modalities to be utilized by medical physicians and mental health practitioners. Importantly, industry and streaming services can use these new insights to optimize their technologies and develop music-based health and wellness platforms aimed at improving the well-being of its users, ultimately impacting the way music is used by millions of people globally.

Addresses
[1] Department of Clinical Psychology, City College of New York, City University of New York, USA
[2] Department of Psychiatry, University of Cambridge, UK
[3] Department of Psychology, University of Cambridge, UK

Corresponding author: Greenberg, David M (dmg39@cam.ac.uk)

Current Opinion in Behavioral Sciences 2017, **18**:50–56

This review comes from a themed issue on **Big data in the behavioural sciences**

Edited by **Michal Kosinski** and **Tara Behrend**

http://dx.doi.org/10.1016/j.cobeha.2017.07.007

Introduction

Music is pre-historic and its impact on health has been documented for thousands of years [1,2•,3]. Music is not simply entertainment — scientific research from the past decade has shown that it played an integral role in human evolution, and is closely tied to communication, social bonding, and human development [4–8]. Today, music remains a central part of human experience across cultures and the lifespan [9]. People listen to it in a multitude of contexts for a total of 11–44% of their waking lives [9–13]. With the advancement of big data and technology, we are entering a unique period with an opportunity to gain new insights into the uses and effects of music that was not previously possible. This new knowledge has the potential to scientifically inform how music can benefit medical and clinical settings, and industry. In this paper, we will review prior scientific research on music, explore different areas of Big Music Data, and outline a conceptual approach that we feel will be most fruitful this endeavor.

Prior research

Prior theory and research in music psychology has shown the impact of music on neurological, personological, social, and cultural levels: Research in music cognition has outlined how we process musical information and form cognitive and affective representations of it [14]; research in neuroscience has identified the reward networks in the brain that are activated when listening to music and the hormones and chemicals that are excreted in response to it [15–17]; research on affect and music has explored the underlying mechanisms for how people perceive and experience emotions from music and the process of how the emotional intensions of the musician are expressed through music and then perceived and felt by the listener [18,19]; research in personality and social psychology has shown how individual differences in musical preferences is linked to personality, values, and cognitive styles [20,21,22•,23]; research on the uses and gratifications of music has shown that music plays distinct functions in a variety of contexts including concentration during work and motivation during fitness [24,12]; medical research has shown how music can impact physical health and rehabilitation such as increasing recovery rates after surgery [25•]; and research in music therapy has shown that music-based treatment interventions can be successfully used to address mental and emotional health issues in those with autism, depression, post-traumatic stress disorder, and dementia [26,27•,28,29•]. Findings in the areas have provided a thorough theoretical and scientific base which can be immediately applied to big data research.

Limitations of prior research

Though there have been significant advances in our scientific understanding of music, the field has been

hindered by methodological limitations. First, with the exception of a few studies [9,30], the majority of research in music psychology has reported relatively small samples sizes. This is in part due to limited access to undergraduate samples and subject pools, less funding opportunities, and lower research exposure than is available for research in more traditional fields. Though in recent years online recruitment platforms have increased the scope of recruitment strategies, the sample sizes reported in music psychology journals remain smaller when compared to related fields like personality and social psychology. Larger samples are important because they increase statistical power, and allow researchers to control for confounding variables, observe small effects, analyze within and across sample replication, and examine affects across age groups and geography. Considering that small sample sizes contributed to lack of replication in social psychology, larger samples could prevent such a crisis in music psychology.

Second, the representative musical stimuli used in studies are limited. Often, experimental and correlational studies use stimuli that are brief (15–30 s in length) sometimes it is computer generated or manipulated, and in some cases has never been heard before by participants [31,32]. Though these approaches limit confounding variables, they also lack ecological validity. Third, regardless of the stimuli administered, given the time restrains of online and laboratory experiments, the music that is used does not capture the ways in which people naturally interact with music in their daily life, and the breadth of music they are exposed and listen to — it only captures a snapshot of the way someone listens to, responds to, and engages with music. As will be shown in this paper, big data has the possibility to advance beyond these limitations.

Big Music Data

Recent technological advances, including the Internet, streaming services, online social media, and audio file formats, has generated the collection of large amounts of data relevant for psychological research on music. This includes big data on the human-level and song-level. To date, there have been four approaches to Big Music Data discussed below.

Mass Internet surveys

One approach to Big Music Data is administering music surveys and experiments to masses of individuals. This predominantly descriptive and correlational approach provides a powerful platform for mapping a range of psychological and music-related phenomena. Typically, the data generated by online surveys are self-reports of demographic and psychological characteristics combined with information about music use that includes musical preferences, and affective reactions or perceptions of musical stimuli. The primary advantage of this approach is that it provides access to large and diverse samples of people around the world.

There have been several examples of successful platforms where large amounts of music psychology data have been collected: (a) In the myPersonality project [33•], over 20,000 Facebook users provided their affective responses to musical stimuli and completed measures of personality and other psycho-demographic measures. From this data, researchers have been able to examine the structure of musical preference and its correlates with personality and cognitive styles [21,22•,31,34]; (b) As part of the Internet-based music preferences project at Out of Service (www.outofservice.com/music-personality-test), a quarter of a million participants have provided self-reports of their personalities, demographics, and music genre-preferences. From this dataset, researchers have been able to examine musical preferences across the lifespan, showing that normative trends in musical preferences correspond to Erikson's psychosocial stages of development [9]; (c) As part of the BBC 'Lab UK' project, nearly 150,000 participants completed self-report and behavioral tasks of musical ability. These data were used by researchers to explore the structure, correlates, and geographic distribution of musical sophistication [30]. Further, the data were combined with a separate BBC 'Lab UK' dataset on personality and well-being to show that personality traits predict musical ability in both musicians and non-musicians [35]; (d) Most recently, Greenberg created the Musical Universe project (www.musicaluniverse.org), which is one of the most extensive datasets in terms of the quantity and breadth of musical and psychological variables. Over 100,000 people have completed measures on musical preferences, personality, well-being, and demographics (including musicianship, musical consumption, music training, geographic location and clinical diagnoses). In terms of psychological information, large subsamples have completed measures on mood, cognitive style, emotion regulation, values, the dark triad of personality, and mind-reading. In terms musical information, large subsamples have completed measures on musical engagement style, music perception, musical and creative arts performance attributes, and experimental listening tasks that detected changes empathy in response to music listening.

Online social media

Online social media (OSM) is one area where musical behavior can be observed 'in the wild' via digital footprints. OSM are forums where people come together for the purpose of interacting with each other and sharing information — an interaction which invariably includes the expression of musical information. Typically, the data available from OSM include behavioral records of the music people like that can be gained from Facebook likes and Twitter. One of the biggest advantages of these data is that they are behavioral and therefore overcome some