**Television by the Numbers: The Challenges of Audience Measurement in the Age of Big Data**

**Introduction**

On August 24th 2013, leading global market research firm A.C. Nielsen celebrated its 90th anniversary. To mark the occasion, the company launched an interactive web-based timeline highlighting the key achievements over their 90-year reign. According to a press release that accompanied the launch (Nielsen, 2013), the website was designed to showcase the numerous innovations introduced by Nielsen over the past nine decades. Scrolling through the timeline, however, a different picture emerges; one that reveals a relatively consistent and conservative approach to audience measurement. This alternative narrative is consonant with scholarly accounts of the television ratings industry (Balnaves and O’Regan, 2011; Bermejo, 2009; Buzzard, 2012; Meehan, 1990). Indeed, having established itself as the dominant force in market research in the 1940s, and with relatively limited competition since that time (Buzzard, 2012), Nielsen has had little incentive to innovate and has therefore been able to maintain its dominant position in the market by pursuing the same core strategy of sampling audiences. The logic that underpins this approach dictates that a small sample, if truly representative of the television viewing audience, can accurately apply to the entire market. As one ratings executive put it: ‘If you have a bowl of soup in which all the ingredients are thoroughly mixed, you do not have to eat the entire bowl to know how the soup tastes’ (Seiler in Buzzard, 2011: 47)

Today, however, small group sampling is not nearly as effective. Audiences are increasingly fragmented, watching at different times and on different screens. To return to the analogy above: it’s hard to know how the soup tastes if the ingredients are in different bowls. Moreover, likening the audience to a bowl of soup presupposes that the recipe is always the same, whereas in reality the demographic make-up of the audience is constantly changing. Thus, as a result of widespread changes in the profile of the audience as well as the development of new viewing practices, the long-established system of sampling audiences as they watch content in their living rooms is clearly no longer as effective as it once was. To be fair, the Nielsen timeline does highlight some examples of how the company has responded to these changes, such as the inclusion of time-shifted viewing figures and the introduction of a more comprehensive methodology that allows for the monitoring of content across multiple screens. Yet the timeline stops at a rather critical juncture in the company’s history. Just two months after the website’s launch, Nielsen announced the introduction of their latest and arguably most innovative measurement service: a new metric that utilises enormous quantities of data mined from Twitter, aptly named the Nielsen Twitter Television Ratings (NTTRs). In contrast to the company’s legacy system of monitoring audiences on the basis of small group sampling, this new approach to audience research uses vast, if not quite complete, datasets and thus represents a major shift in ratings methodologies; the implications of which are the subject of this article.

In order to properly examine the significance of these recent developments, it is necessary to begin by defining some key concepts. The NTTRs are an example of social media metrics; a relatively new approach to market research and audience measurement that utilises large quantities of data generated by social media activity. Social media metrics, in turn, can be considered part of the emerging science and industry of Big Data. Though the precise definition of Big Data is the subject of ongoing scholarly debate (boyd and Crawford, 2012; Kitchin, 2014; Kitchin and McArdle, 2016; Manovich, 2012), in her pioneering study of the its role within the media industries, Martha L. Stone offers a broad but useful working definition of Big Data, describing it as ‘an umbrella term for a variety of strategies and tactics that involve massive data sets, and technologies that make sense out of these mindboggling reams of data’ (2014: 2). Whilst many definitions of Big Data have focused on volume as its defining feature, Rob Kitchin has produced a more nuanced definition in which he notes that ‘Big Data is characterized by being generated continuously, seeking to be exhaustive and fine-grained in scope, and flexible and scalable in its production’ (2014: 2). In many respects, Kitchin’s definition recalls one of the earliest attempts to describe this phenomena, namely Doug Laney’s (2001) “Three Vs” of Big Data which identified volume, variety and velocity as its key features.[[1]](#endnote-1) Of course, data has always played an important role in the television industry, particularly in relation to ratings and market research, but the volume, variety and velocity of Big Data clearly marks it as distinct from earlier forms of “small(er) data”. For instance, data generated through Twitter is enormous in size (volume), is produced in real-time (velocity), and is comprised of a range of different data types such as text, URLs, geographical location, date published, and retweet count (variety). As such, the developments discussed throughout this article, such as the NTTRs, constitute a marked departure from earlier audience measurement practices and therefore have significant implications for how the television industry operates.

Over the past decade or so, Big Data has been utilised by, and subsequently shaped, a number of different fields including finance, retail, healthcare, and policing amongst many others. Yet as the introduction of the NTTRs indicates, Big Data is also making significant inroads within the entertainment industries. This move has been driven by a number of factors, including the proliferation of viewer-related data generated directly through services such as Netflix and the BBC iPlayer or indirectly through platforms such as Twitter, Facebook or Telfie (formerly GetGlue). Thus, Big Data in the context of the television industry refers to a range of data types and sources, though it is the latter of these (“indirect data”) that will be the primary focus of this article, in part because of the limited availability of “direct data”. The volume, variety and velocity of data described here, coupled with the increasing fragmentation of the television audience has therefore forced market researchers and the ratings industry to develop more innovative and effective ways to accurately measure viewers and their increasingly diverse viewing habits.

In the context of this rapidly evolving television landscape, Allie Kosterich and Philip M. Napoli suggest that the industry has subsequently ‘become enamored with notions of “Big Data” and how they can be harnessed to generate strategic insights and enhanced revenues’ (2015: 9). For many, this new approach to audience research promises more stability, perhaps even predictability, for an industry typically characterised by risk and uncertainty. In his keynote address at the 2014 Edinburgh Television Festival, Channel 4’s chief executive David Abraham took the opportunity to underline his long-standing belief in Big Data: ‘Over the past few years I’ve been encouraging the TV industry to embrace the power of data’, explained Abraham, before warning attendees that, ‘a TV channel without a data strategy is like a submarine without sonar’ (2014). Abraham’s faith in big data is consistent with many of his industry peers. Shortly following her appointment as Chair of the BBC trust in 2015, Rona Fairhead observed that ‘when it comes to using data to understand its audiences the BBC is a long way behind the competition’ (in Reevell, 2015).

Whether these attitudes are driven by a fear of being left behind – as indicated by Fairhead’s comments – or whether there is a genuine belief that Big Data can have a positive creative and cultural impact – as was the nature of Abraham’s keynote address – the emerging consensus amongst network executives is that data has become an integral part of televisual culture; an essential tool for survival in the increasingly fragmented, crowded and competitive marketplace of digital TV.

Though Abraham and Fairhead’s comments specifically relate to the data strategies of public service broadcasters in the UK, the growing international trade of television coupled with the emergence of transnational services such as Netflix and Amazon Prime demonstrates that Big Data is clearly a worldwide phenomenon and is therefore very much on the minds of network executives across the globe. Thus, while many of the examples discussed below stem from the US, they are nevertheless indicative of how Big Data is being adapted and adopted around the world.

In contrast to the widespread optimism shared by those within the television industry, there is a noticeably more cautious and critical tone in much of the scholarly work on Big Data within the humanities. Whilst its proponents are quick to point out the economic and scientific[[2]](#endnote-2) efficacy of this new research paradigm, more sceptical observers have stressed the need to acknowledge and interrogate its limitations (see, for instance, Manovich, 2012). In their polemical “Six provocations for Big Data”, danah boyd and Kate Crawford (2011) describe how Big Data has been hailed by many – including those in market research – as a more rigorous, veracious and objective system of producing knowledge. However, their account highlights a number of problems with Big Data, calling the integrity of this emerging research paradigm into question. Big data, they explain:

tempts some researchers to believe that they can see everything at a 30,000 foot-view. It is the kind of data that encourages the practice of apophenia: seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions (boyd and Crawford, 2011: 2)

In light of these epistemological concerns, boyd and Crawford maintain that ‘it is crucial to begin asking questions about the analytic assumptions, methodological frameworks, and underlying biases embedded in the Big Data phenomenon’ (2011: 2); a sentiment shared by a number of other critics including Rob Kitchin (2014: 10). In contrast, then, to the television industry’s seemingly uncritical acceptance of this new research paradigm, this article assumes a more cautious and critical tone in examining the relationship between television and Big Data. Rather than focusing exclusively on the potential value or benefits of Big Data (creative, cultural, economic, or otherwise), this essay is concerned with also investigating the challenges and limitations that the industry faces as it continues to integrate Big Data within its existing business model – challenges and limitations that have largely been neglected in both industry discourse and, to a lesser extent, academic work.

With the creative industries evidently poised on the cusp of a Big Data revolution, the need to examine this emerging phenomenon has never been more pressing. In the context of these developments, this article adds to and advances a somewhat limited but growing body of research on social media metrics and Big Data by examining how these innovative methodologies are being developed, utilised and integrated specifically within the television industry. In doing so, it recognises and responds to Kosterich and Napoli’s (2016) call for research to more thoroughly examine the formal processes, technologies and institutions that underpin this emerging industrial strategy.

In taking up this challenge, the article is divided into two parts. The first of these provides a broad but necessary overview of the history of ratings in order to contextualise these more recent developments. Following this, I map out the relevant critical terrain through a brief outline of key research on TV ratings, social media metrics and Big Data. Though the article draws on a relatively limited pool of TV studies oriented scholarship on Big Data, it combines this material with a much larger and somewhat untapped body of research that sits at the intersection of computer sciences and cultural studies. Having established the critical role of ratings and how this industry has been conceptualised within academia, the second section of this article utilises the most pertinent of boyd and Crawford’s (2011) “six provocations” in order to examine how the television industry is responding to the challenges of Big Data. Whilst these provocations provide the main critical framework for the article, I also draw on trade press, marketing materials and interviews with producers and television executives. In including these “industrial voices”, this essay provides a more holistic, empirical and critical account of these recent developments that traces preliminary connections between the introduction of new audience measurement practices and the production culture of contemporary television.

**Big Data and television**

TV ratings have long been an integral component of television production yet are often overlooked in scholarly analyses. This is not always a deliberate oversight. The mechanics of ratings are highly complex and can be difficult if not impenetrable to those working in the humanities who are more au fait with textual analysis than statistical analytics. Their complexity and (in)accessibility is further exacerbated by the industry’s growing reliance on Big Data; a model of analysis that requires a combination of sophisticated software and expensive hardware, not to mention a highly specialised skill set. What is more, ratings, as well as the means through which they are attained, are often highly guarded, only of monetary value when their availability is restricted to certain parties. This is especially true of more recent data-driven companies such as Netflix whose entire business model relies upon the limited availability of such data.

As such, television ratings constitute an example of a “transparent intermediary”[[3]](#endnote-3); a term that Joshua Braun (2014) has used to describe the hidden or invisible (infra)structures that underpin and shape the production, distribution and consumption of cultural goods. Whereas Braun’s account focuses on the importance of institutions such as Nielsen and BARB, Michael Lahey’s (2016) more recent analysis of the critical role that application programming interfaces (or APIs) play in the construction of connected viewing experiences represents a more specific and technical example of an equally invisible but influential force in the ecosystem of television, one that Lahey describes in similar terms to Braun as “invisible actors”.[[4]](#endnote-4) Regardless of their different subject matter, both accounts ultimately draw attention to a rich yet often overlooked array of institutions, technologies and protocols that, despite their lack of visibility to consumers, play an integral role in every facet of television culture. Indeed, it is fair to say that most viewers are largely unaware of how television ratings work or their own role within such a system. Yet the content we consume is ultimately defined by this process. According to Braun, the inconspicuous nature of these systems is the very reason that they should be subject to more rigorous scrutiny.

Braun suggests that we should pay closer attention to transparent intermediaries, not only because they are too often overlooked in critical accounts, but because they ultimately ‘facilitate the exercise of structural power’ (2014: 124). However, the need to scrutinise ratings is also motivated by the fact that the industry is currently undergoing a process of significant transformation in which the gathering and analysis of information about audiences is increasingly performed by computer algorithms. Of course, the increasing mechanisation of market research isn’t necessarily a recent phenomenon (see Striphas, 2015). Nevertheless, this trend is especially prevalent today within the realm of television ratings and audience research, where companies such as Nielsen, in the US, and BARB, in the UK, as well as smaller start-ups including BluFin Labs, SecondSync (both of which have since been acquired by Twitter), Canvs and Datasift, are beginning to utilise larger datasets in an attempt to provide more comprehensive and complex portraits of the audience.

This transition has taken place in the broader context of what some critics are calling the “computational” or “algorithmic turn” in media production in which the production and distribution of media is increasingly determined by insights gathered via the harvesting and analysis of large amounts of pertinent data (Napoli, 2014; Striphas, 2015; Uricchio, 2011, 2015). Drawing on Pierre Bourdieu’s (1984) notion of cultural intermediaries, Jeremy Wade Morris (2015) has argued that the organisation and presentation of cultural goods is increasingly driven by algorithms, citing Amazon’s recommendation engine as a key example. This “curation by code”, Morris maintains, has profound implications for the production and consumption of culture more widely. Whereas in Bourdieu’s account the role of the intermediary is performed by an individual (or institution) acting as a cultural gatekeeper, Morris suggests that the role of human agency within the (re)presentation of digital culture is gradually diminishing, with creative and curatorial decisions increasingly delegated to computational processes and complex algorithms. It is important to stress that Morris is not suggesting that the individual has become obsolete as a consequence of the development of curatorial algorithms. Rather, his account makes the simple but valuable observation that we are living in an era in which cultural production is subject to a combination of human (and institutional) *intermediaries* as well as computerised *infomediaries*.[[5]](#endnote-5) It is also important to note that while algorithms are a distinct entity in and of themselves, they are an integral component of Big Data and an increasingly significant part of the mechanics of television ratings today. Algorithms are used to gather and analyse data but, as Morris’ account makes clear, they are also utilised when it comes to producing, organising and recommending cultural goods.

In bringing these two distinct but related lines of enquiry together it would therefore be productive to think of ratings firms such Nielsen as BARB as *transparent infomediaries* as they are highly transparent (Braun, 2014; Lahey, 2016) and increasingly automated (Morris, 2015; Striphas, 2015). Conceiving of the ratings industry as *transparent infomediaries* has critical value as it ultimately encourages us to re-evaluate the function and influence of these institutions whilst drawing our attention to a slow but significant shift in how companies such as Nielsen and BARB now operate – a shift that has profound implications for the production and consumption of television that will be explored below.

*Manufacturing and measuring audiences: the ratings effect*

Given the contemporaneity of the developments discussed above, critics such as Braun (2014), Lahey (2016), and Morris (2015) remain somewhat uncertain as to what the algorithmic turn and the emergence of transparent infomediaries might mean for television and the production of cultural goods more broadly. However, there seems to be very little doubt that transformations in the way that television viewing is measured will have a profound effect upon the industry. Indeed, regardless of their in/visibility, TV ratings play a crucial role within the delicate ecosystem of television and the slightest change in how audiences are measured can have profound consequences. As Karen Buzzard explains:

the business implications of a shift or a change in the currency could be staggering. One truism in media market research is that different methodologies produce different ratings and different portraits of audiences. And these differences could mean millions of dollars won or lost if a new methodology were adopted (2012: 148)

Buzzard’s use of “could” implies a hypothetical argument. However, there are a number of precedents that demonstrate the significant role that ratings play within television culture and the impact that even relatively minor changes can have on production practices. In the early 1970s, for instance, Nielsen adopted a new approach to measurement in which they shifted their focus from the *quantity* of the audience towards the *quality* of the audience, a process that Jane Feuer has described as involving ‘a de-emphasis on numbers and a greater emphasis on ‘demographics’’ (1984: 3).[[6]](#endnote-6) The consequences of this shift were profound. As M.J. Clarke explains, ‘changes in these ratings in the early 1970s that supported more granular, demographic-sensitive data, famously encouraged CBS’s movement from broad-appeal, so-called hayseed comedies to the slick-modern MTM and socially conscious Norman Lear product, in an effort to capture more upscale, urban ratings’, adding that ‘a change in the metrics used to calculate consumption greatly changed the picture of the television market – from one of a tremendously large, undifferentiated audience to one serving elite niches – and, thereby, creative decision making’ (2013: 123). Clarke’s deliberate use of “encouraged” is important here. In reality, there were a number of other factors at play, including the introduction of the Financial Syndication Rules in 1970 as well as significant socio-political shifts in the demographic make-up of U.S. audiences around this time. Nevertheless, this transformation in audience measurement practices significantly contributed to what Todd Gitlin has called the “turn to relevance” (1994) within U.S. prime time television – namely the production of a more relevant style of programming that addressed a range of different niches and targeted more lucrative demographics.

 To date, this is one of the clearest examples of how a shift in the currency of TV ratings has had a tangible impact upon the content and culture of television production. Naturally this raises a number of questions about television today. If the “turn to relevance” was even partly inspired by a shift in Nielsen’s ratings strategy, then the recent adoption of Big Data and social media metrics should be of even greater interest to critics and industry figures alike. Indeed, as noted above, the industry’s rather precipitous adoption of Big Data involves a much more dramatic reconfiguration of the way that audiences are measured and manufactured, particularly when compared to developments in the 1970s.

To suggest that this change in tact will produce a different portrait of the audience is probably an understatement. However, whilst Big Data has become a buzzword of late within the media industries, it is important not to exaggerate its influence. Using Greenwood et al.’s economic model of ‘institutional change’ (2002), Kosterich and Napoli (2016) have produced a detailed account of the processes through which social media metrics have become formally adopted within the industry, concluding that they have not displaced the previous currency, or ‘market information regime’. Rather, ‘Nielsen’s efforts to diversify into social TV analytics’ they note, ‘have clearly been accompanied by a discursive effort to explicitly position social TV analytics as supplementary to traditional Nielsen ratings’ (2016: 12). The same can also be said of the ratings industry in the UK where, at the time of writing, BARB are currently piloting a new system dubbed Project Dovetail. As its name implies, Dovetail is an attempt to combine traditional methods of small group sampling with newer data sets such as those acquired through social media – an approach very similar to the NTTRs and one that even involves the assistance of Nielsen. However, like Nielsen, BARB have been careful to stress the importance of more traditional measurement practices. As the accompanying narration to a promotional video for Project Dovetail explains:

Some argue that these new mountains of Big Data outperform more traditional forms of data. BARB believes that they are complimentary to the strengths of our panel. In fact, we think it is the combination of the two that gives us the best possible way forward for measuring viewing (BARB, 2015).

Whilst organisations such as Nielsen and BARB are clearly working to envelop social media metrics into their existing business models in order to nullify their disruptive potential, Big Data continues to gather momentum as a new paradigm of market research and therefore warrants closer critical attention.[[7]](#endnote-7)

Though the above constitutes a somewhat broad overview of the history of TV ratings, it nevertheless demonstrates the integral role that this industry plays within the precarious ecosystem of television production. At the same time, this brief survey draws attention to more recent and significant changes in the industry – namely the “computational” or “algorithmic” turn – and suggests that a more productive way to understand and approach organisations such as Nielsen and BARB is to conceive of them as *transparent infomediaries* – largely invisible forces yet increasingly automated and ever-more influential.

**Six Provocations**

Having outlined some of the key works on TV ratings and the “algorithmic turn” in the cultural industries, the second part of this article uses several of boyd and Crawford’s (2011) “six provocations” to explore these ideas in more concrete terms. Comprised of a series of critiques of this new approach to research, these provocations draw attention to a number of different challenges associated with Big Data and thus functions as a critical framework that encourages us to question the prevailing industrial consensus that Big Data represents a superior model of knowledge production. Through this critical lens, I trace some preliminary connections between these burgeoning methods of data-driven audience research and emerging trends in the production, promotion and commissioning of television. Whereas much of the literature cited above has sought to map the conceptual terrain around ratings, audience research and/or Big Data, I want to advance these debates by considering some of the more tangible implications of these developments.

In their seminal account, boyd and Crawford (2011) use the following provocations to critique Big Data:

* + 1. Automating research changes the definition of knowledge;
		2. Bigger data are not always better data;
		3. Limited access to Big Data creates new digital divides;
		4. Not all data are equivalent;
		5. Claims to objectivity and accuracy are misleading;
		6. Just because it is accessible doesn’t make it ethical.[[8]](#endnote-8)

Though all of these provocations highlight significant challenges for the industry, some are more pertinent than others when it comes to television ratings and audience research. For that reason, the final section of this paper focuses on the first three of these provocations which, as will become clear below, arguably present the biggest challenge for those working in the television industry today.

 *Automating research changes the definition of knowledge*

The first of boyd and Crawford’s six provocations echoes the concerns of critics such as Striphas (2015) and Morris (2015) who have expressed a similar degree of wariness regarding the increasing automation of culture. Comparing the emergence of Big Data to the development of Fordist regimes of production in the first half of the 20th Century, boyd and Crawford argue that ‘just as Ford changed the way we made cars – and then transformed work itself – Big Data has emerged [as] a system of knowledge that is already changing the objects of knowledge, while also having the power to inform how we understand human networks and community’ (2011: 3). What is at stake, according to boyd and Crawford, is more than simply a change in the processes through which knowledge is acquired, but rather a fundamental redefinition of knowledge itself. In other words, the adoption of Big Data will not only deliver new insights, but will also foster a research culture more attuned to the specific properties and affordances of such a methodological approach. Indeed, boyd and Crawford argue that Big Data privileges real-time analyses, which in turn will determine the kinds of questions asked as well as the outcomes of said research (2011: 4).

There is evidence of this privileging of real-time analysis in the very design of some of the most popular hardware and software configurations currently utilised in the market research industry. For instance, the intense volume and velocity of Big Data – two of Laney’s (2001) infamous “three Vs” – has encouraged the development of analytics software such as Apache Storm and Apache Kafka, both of which are designed to analyse information in real-time, and both of which are used extensively in social media metrics. Depending on the particular software and server configuration, they can analyse information on-the-fly, without data ever being committed to disk. These tools therefore prioritise the analysis of data in the present over data from the past, making it difficult if not impossible to perform retrospective analyses due to the technical limitations and prohibitive costs associated with data retention. As the volume and velocity of data continues to grow at a rate that exceeds the development of affordable storage space, it is safe to say that real-time analytics will become an even more prevalent form of research in the future.

While boyd and Crawford argue that Big Data privileges real-time analysis, and therefore the production of a particular form of knowledge centred on real-time analytics, the situation is more complicated than this. Services such as Twitter (who are used by Nielsen for their NTTRs) utilise software packages such as Apache Storm to perform real-time analytics, which are necessary for providing features such as “trending topics”. At the same time, however, Tweets are retained and remain available for retrospective analyses – an approach known as batch processing (which is performed on clusters stored using cloud services such as Amazon’s S3). In short, Twitter (and Netflix for that matter) have adopted a more complex approach to Big Data, in that they both combine real-time analysis (which underpins certain key features including Netflix’s recommendation engine) with retrospective batch analysis (which allows analysts go back and review historical data): a form of data-processing known as Lambda architecture. Even so, the preservation of content via batch processing is often limited to a relatively short period of time so that space can be made for the relentless stream of new incoming data. Despite the addition of retrospective batch analysis to the arsenal of companies such as Nielsen and Netflix there is still a methodological emphasis on real-time or very recent data analytics, which in turn privileges and produces certain forms of knowledge and user experiences. In relation to the examples discussed above, it may be too pre-emptive to say that the definition of knowledge itself is changing, as boyd and Crawford (2011) argue. However, these examples certainly indicate that certain methods and forms of knowledge are privileged by the design, affordances and limitations of these emerging technologies.

But what does the increasing automation of audience research mean for television? For one thing, if Big Data, or more precisely social media metrics, are more conducive to real-time analysis, this could lead to a greater investment in content that can more effectively deliver this kind of measurable data, such as live programming and event television (see Sørensen, 2015). Indeed, NTTRs indicate that live programmes perform much better than their scripted counterparts when it comes to generating social data. This, in turn, potentially changes the criteria of what might be considered a success. The 2015 MTV *Video Music Awards*, for example, dominated the NTTRs for all television broadcasts during the week commencing the 24th of August 2015. According to the figures, the *VMAs* generated tweets from just over 2,200,000 unique authors producing a total of 21,300,000 tweets, resulting in almost 680,000,000 impressions (in other words, the number of times a *VMA* related Tweet was seen across the wider social media sphere – whether or not these were actually read or had any positive economic effect is another matter altogether).[[9]](#endnote-9) To put the social media success of the *VMAs* into some perspective, we need only consider the second most popular series or special that week,[[10]](#endnote-10) *WWE Monday Night RAW*, which generated just 66,000 Tweets – roughly 97% fewer Tweets than the *VMAs*. Even the most popular sports broadcast for that week (*MLB Baseball: Chicago Cubs at Los Angeles Dodgers*) generated a comparably meagre 102,000 Tweets from unique authors – again, a small fraction of the social media activity generated by the *VMAs.*

Despite the *VMAs* dominant performance in social media ratings, these figures appear to contradict Nielsen’s more established method of measuring TV viewership which revealed that the broadcast itself was only the fourth most-watched cable series that week with just over 5,000,000 viewers – more than 3,000,000 short of the top cable broadcast that week, the pilot episode of *Fear the Walking Dead* which proved to be stiff competition in more ways than one.[[11]](#endnote-11) In fact, if we take into account the viewing figures that week for major networks, syndicated networks and cable, the *VMAs* didn’t even feature in the top 20. This discrepancy between live viewing figures and social media ratings raises a number of important questions related to the currency of these different approaches. As Kosterich and Napoli’s (2016) study suggests, networks and advertisers still place most of their faith in traditional ratings. Nevertheless, as Big Data continues to make further inroads within the market research culture of the television industry, it is likely that networks, producers and advertisers will feel more inclined to invest in or design programming that can generate high levels of social media activity.

Although the impact of Big Data is, somewhat ironically, difficult to measure, research shows that networks and advertisers are spending more on sports, event television and other genres of live programming. ‘Call it the eventization of TV’, one journalist recently explained, ‘at a time when nearly half of all U.S. homes have DVRs, networks are shelling out an estimated $7 billion for rights to air NFL games, awards shows are popping up all over the dial, and there doesn't seem to be a major cable network that isn't exploring a foray into topical late-night’ (Guthrie and Rose, 2013). The same article continues: ‘advertisers, too, are clamoring for such opportunities in a fractured, ad-skipping environment, shelling out $444 million on awards shows and live non-sports events in 2012, up 22 percent compared with five years ago’ (ibid.) According to this account, the growing investment in live programming is largely attributed to the threat of time-shifting and other ad-skipping technologies. However, it could be argued that the popularity of social media metrics, and their privileging of live/real-time analysis, is also encouraging the increasing “eventization of television”.

If Big Data is privileging real-time analytics, which in turn encourages greater investment in live programming, what does this mean for scripted television? As Mike Proulx and Stacey Shepatin observe, ‘scripted dramas tend to produce lower volume backchannels during show airings’ (2012: 118). In other words, people are less likely to Tweet during the types of programmes that demand greater viewer engagement. It is worth exploring this point further as these are precisely the kinds of viewers that ad-supported networks are keen to attract: people highly engaged in their programming and, by extension, their advertising. However, because of their increased level of engagement in content these kinds of viewers are more likely to be poorly represented in social ratings. As a consequence, networks might then assume that scripted drama is not producing sufficient enough “buzz”; a failure to provide data that not only translates to free promotion, but also forms the basis of the insights gathered through social media metrics.Of course, it would be wrong to suggest that Big Data’s methodological emphasis on real-time analytics will signal the end of scripted programming. On the contrary, many television networks and producers have explicitly sought to increase social media engagement in fictional programming in ways that strategically reinforce viewer engagement whilst also soliciting their feedback. AMC, for example, have developed a number of second screen apps for their major scripted series, grouped together under the “Story Sync” initiative. Though second screen applications appear to be more conducive to live television genres (Lee and Andrejevic, 2014), the continued investment in companion apps for scripted dramas provides further evidence of a growing appetite for viewer data regardless of the genre.

In light of the above examples we can conclude that, contrary to the claims of boyd and Crawford, we are not so much witnessing the wholesale emergence of a new epistemological regime or a redefinition of knowledge, but rather we are seeing the adoption, exploitation and coercion of new tools to reinforce existing practices and industrial lore, as demonstrated in Kosterich and Napoli’s (2016) account of the institutionalisation of social media metrics. For example, the announcement of Nielsen’s NTTR service in 2013 caused much consternation across the industry (Watercutter, 2013). Advertisers, network executives, journalists amongst others all speculated as to what form these would take and how they might unsettle or disrupt the industry. In the end, the numbers – at least those that are made available to the general public – appear to simply reinforce Nielsen’s longstanding practice of ‘counting eyeballs’ (Gitlin, 1994: 49). Nevertheless, the introduction of the NTTRs and Project Dovetail suggests that the industry is beginning to change tact. This is significant because, as several critics have already pointed out (Buzzard, 2012; Braun, 2014; Lee and Andrejevic, 2014; Kosterich and Napoli, 2015), the implications of a shift in how audiences are measured would be profound. In explaining why there is a considerable amount at stake in how the ratings industry operates, Braun argues that:

the politics involved are the politics of representation and the risk of misrepresentation or, worse, omission and invisibility. The worry here is that groups whose viewing activities are not accurately recorded will not be sought after as audiences. Their interests and views may therefore be less readily represented in the content of media, and thereby omitted from the public agenda (2014: 136-137)

And while social media ratings have the potential for a fairer and more representative system of audience measurement, there is also the possibility that we are simply seeing the development of a culture in which viewers are being ‘interpellated into ever more convenient, instrumental, and commercially viable social identities’ (ibid.) In the context of this analysis, it could be argued that the allure of data generated through social media renders certain demographics, particularly those who are less prolific users of these platforms, entirely invisible altogether. If anything then, the use of social media metrics (i.e. “indirect data”) is very limited as it magnifies the potential for misrepresentation by privileging certain types of audiences (younger, more technologically savvy) and viewing behaviours (such as real-time commentary and interactions) at the expense of others.[[12]](#endnote-12)

*Bigger data are not always better data*

b bbboyd and Crawford are not alone in their assertion that a profound epistemological shift is taking place in the wake of Big Data. Andrejevic (2014), for example, has express a similar set of concerns in response to the automation of research, arguing that the excess (or volume, to use Laney’s (2001) terminology) of information associated with Big Data, and its subsequent need for computational processing, has produced a culture more concerned with the results of research rather than what those findings might actually mean; a quality that he describes as ‘knowing without understanding’ (2014: 21). In many ways, Andrejevic’s concerns resonate with boyd and Crawford’s claim that there is:

an arrogant undercurrent in many Big Data debates where all other forms of analysis can be sidelined by production lines of numbers, privileged as having a direct line to raw knowledge. Why people do things, write things, or make things is erased by the sheer volume of numerical repetition and large patterns (2011: 4).

When considered in the context of audience research, these arguments suggest that the industry’s growing reliance on Big Data may lead to a culture of production in which ‘correlation supersedes causation’ (Kitchin, 2014: 4). However, this kind of logic is true of how the ratings industry has always operated. Historically, organisations such as Nielsen and BARB have been almost exclusively concerned with questions of quantity (*how many* people are watching, *who* is watching) as opposed to questions of quality (*how* and *why* they are watching). Though Big Data and social media metrics have the potential to reveal new insights pertaining to the latter, this may not work to the advantage of the ratings industry as the breadth and depth of such data will ultimately produce too many competing demands. As one television ratings executive explained when asked about the prospect of a move towards such large scale research: ‘a census is not [necessarily] a good thing. When they take your blood, they don’t take all of it’ (Bachman in Buzzard, 2012: 148). This perspective speaks to Andrejevic’s observation that a ‘paradox of an era of information glut emerges against the background of the new information landscape’ in which, ‘to inform ourselves as never before, we are simultaneously and compellingly confronted with the impossibility of ever being fully informed’ (2013: 2). In other words, the more information there is, the harder it can be to make sense of.

Despite the qualitative opportunities inherent in social media analytics, the ratings industry is still primarily concerned with how many people are watching rather than why they are watching – the latter of which is an approach that Lev Manovich has described as ‘cultural analytics’ (2009) – and in this respect Big Data has an obvious advantage. Nevertheless, boyd and Crawford maintain that the biggest strength of Big Data – its size – is also one of its main weaknesses. Indeed, the volume and variety of Big Data produces a larger number of variables that can amplify the margin of error. In detailing this particular frailty, boyd and Crawford note that ‘large data sets from Internet sources are often unreliable, prone to outages and losses, and these errors and gaps are magnified when multiple data sets are used together’ (2011: 5).

While the problem of data excess has been the subject of a number of recent scholarly studies, the notion that bigger data are not always better data is hardly new – especially in the context of television ratings. For instance, as John Ellis observed in 2000:

Across the day, the evening, the week and the month, the level of detail provided by BARB is extraordinary and even perhaps counterproductive. For the schedulers have to undertake a considerable degree of interpretation in order to deal with the figures. From the plethora of detail, they first construct a narrative of the audience for themselves (2000: 136-137)

Clearly, information excess is a problem that has plagued the television industry since at least the early 2000s. Even so, the volume of data generated at the beginning of the millennium pales in comparison to the rate of information produced today. Since Ellis published his account, the data generated, tracked and available has expanded exponentially – perhaps best exemplified by CISCO’s study that found internet traffic had grown from 100gb per hour in 1997 to 16,000 gb per second in 2014: an increase of more than 57,000,000% (Cisco, 2015).[[13]](#endnote-13) Naturally, this proliferation of data changes the prospects and possibilities of television audience research. On the one hand, this sudden explosion of information generates a near limitless pool of data from which numerous correlations can be drawn and converted into potential economic gains. On the other hand, however, this data deluge ultimately produces a dissonance in consumer demand; a digital cacophony of desires and preferences that can never be fully satiated.In other words, platforms such as Twitter and Facebook are generating data that might be considered *too* granular, with market researchers, network executives and television producers more inclined to construct their own preferred, apophenic narratives in an attempt to make sense of this overwhelming information.

 If information excess can be counter-productive when it comes to audience research, then the new breed of data-driven services such as Netflix arguably face a greater challenge than ratings firms and traditional broadcast networks. As Tricia Jenkins notes, ‘using programs such as Hadoop, Pig, Python, Cassandra, Hive, Presto, Teradata and Redshift, Netflix is able to process 10+ petabytes of data along with 400+ billion new events on a daily basis in order to learn about its users’ viewing habits’ (forthcoming). These “events” refer to user generated data including the time, location, and device used to access the service, as well as a plethora of other interactions such as pausing, rewinding, re-watching, search history, and so forth. Based on the figures cited by Jenkins, Netflix subscribers generate 2.8 trillion events per week – a figure that will have grown exponentially following the SVOD provider’s rollout to more than 130 new territories in early 2016 (Netflix, 2016). Though Netflix produces an unprecedented level of data, the lack of access to this information makes it difficult for onlookers to ascertain the role that data plays in the creative process. However, CEO Reed Hastings offered some insight in a recent interview, insisting that Netflix ‘start[s] with the data […] but the final call is always gut. It’s informed intuition’ (in O’Brien, 2016). Such an approach to creativity is evocative of the *paint-by-numbers* motif employed by pop artists such as Andy Warhol in the early 1960s which came to symbolize – and was ultimately used to critique – the increasing mechanization of popular culture at the time. Though Reed insists that Netflix’s approach is more *TV-with-the-assistance-of-numbers* than *TV-according-to-the-numbers*, Big Data is nevertheless occupying a more prominent role within contemporary television culture. Yet given the various challenges associated with data excess outlined above, early adopters should clearly be wary of the supposed truism that bigger is better.

*Limited access to Big Data creates new digital divides*

In addition to the problems of data excess, another common misconception surrounding Big Data is that it is an inherently more democratic paradigm of research (see Baack, 2015). Such a fallacy is based on the widespread assumption that individuals, businesses, and/or governments all have equal access to data. However, as has been pointed out numerous times before (boyd and Crawford, 2011; Andrejevic, 2013, Zelenkauskaite and Bucy, 2016), this utopian version of data equality does not reflect reality. Social media data is rarely free or accessible to the general public. Indeed, popular services such as Twitter and Facebook use APIs to restrict access to their data “firehose”, with permission usually only granted to those who are able to pay. Individuals and smaller organisations therefore face significant economic barriers as the cost of data access is often too prohibitive. Citing a recent independent academic study that used social media data, Dixon et. al noted that in order to obtain data from Twitter the researchers in question ‘paid thousands of dollars and signed a contract that prevented them from sharing data with others’ (2015: 298).[[14]](#endnote-14) Conversely, larger and more established organisations such as Nielsen and BARB have ongoing and long-standing contracts with Twitter through which they are able to access large quantities of data (Nielsen, 2012). The high cost of access to data thus ultimately works in the favour of established firms such as Nielsen and BARB who already have the necessary economic power and industrial alliances firmly in place.

It is worth noting that in the context of the television industry this data/digital divide has always existed, with firms such as Nielsen and BARB restricting their data to a limited number of (paying) clients. Rather than increasing this divide, it could be argued that Big Data has created new opportunities to bridge this data gap. Indeed, despite the increasing barriers of access to information for individuals and smaller companies, there is evidence to suggest that networks and producers are finding ways to complement if not bypass their long-standing relationships with ratings providers by turning to Big Data and social media analytics on a more ad hoc basis.A notable example of this occurred when the producers of *Being Mary Jane* (BET, 2013 - ) decided to conduct their own research using Adobe Social; a relatively inexpensive and widely available social analytics software programme. In taking this step, the network was surprised to discover a high degree of interest in one of the series’ more peripheral characters, Avery (played by Robinne Lee). According to the network’s senior director of social media, this discovery led them to ‘amp up [their] coverage of [the character] from a content perspective’ (Lespinasse, in Kuchinskas, 2014). While it was far too late to change the direction of main narrative in order to reflect this discovery, the network was still able to re-cut commercials and create other original promotions as a way to capitalise on these insights. Although this particular example refers to a promotion rather a programme that has been affected by the availability of social media analytics, it still represents an important development. As Jonathan Gray has demonstrated in his study of media paratexts, *Show Sold Separately* (2010), ancillary texts such as promotions play an important role in framing our understanding and interpretation of the primary text.

*Being Mary Jane* is also a significant example as it demonstrates signs of a potential shift in the power dynamics between ratings companies and their clients. In this instance, the network was able to bypass Nielsen, discovering something that the market research giant might never have known or considered to tell them. Of course, networks and producers have always conducted their own independent research, but the quality and relative availability of social media data is allowing them to do this more efficiently and cost effectively than ever before, not to mention on a larger scale than was previously possible. Though critics such as Andrejevic are justifiably concerned when it comes to the problem of data access, the example of *Being Mary Jane* is evidence of the more democratic side of Big Data.

Although access to data remains a barrier for many – in particular producers, writers but also academics (Zelenkauskaite and Bucy, 2016) – the technologies, software and various protocols that underpin this emerging industry are much more open to critical scrutiny. One key reason for this is due to the influence and involvement of the open source community which has played, and continues to play, a pivotal role in the development of Big Data. The Netflix “tech blog”, for instance, regularly publishes detailed posts about the key hardware and software developments that help deliver the company’s streaming service. While the tech blog openly details Netflix’s infrastructure and therefore constitutes an invaluable resource when it comes to understanding the mechanics of Big Data, the data itself continues to remain elusive. As such, the tech blog clearly has limited critical value and ultimately stands as further evidence of Andrejevic’s data divide.

Data access remains an issue when it comes to private organisations such as Netflix, yet is less of an issue in the case of public service broadcasters. The BBC, for instance, publishes a monthly report about its iPlayer usage which contains a wide array of facts and figures. However, these reports are an example of what Stefan Baack describes as an ‘interpretative monopoly’ (2015) in which the information is only accessible in its final processed form, rather than as raw data, and which may therefore reflect the biases of the relevant institution. In this example, the raw data is processed and presented in a way that ultimately serves to justify the economic and cultural value of the iPlayer by focusing primarily on issues of popularity. Were the raw data readily available, who knows what other stories could be told, or insights unearthed.

These examples suggest an insider/outsider data divide between those working within the industry (i.e. network executives, writers, directors, etc.) and those on the outside (i.e. journalists, critics, scholars, etc.) However, the line in the data divide is not so clearly drawn. In many instances those involved in the production of a television programme are not privy to these large datasets. To take one recent example: in an interview in 2015, Beau Willimon, the creator of *House of Cards*, admitted that, ‘I have no idea how many people watched the show on Netflix. [Neflix] have never given me any data whatsoever. All they say is, “Well, we’re doing well and we’d like another season.” And that’s really all I need to know.’ (in Howard, 2015). Willimon’s comments seem to corroborate Hastings’ belief that the numbers are important but do not determine the creative process. Willimon’s comments also draw attention to an interesting distinction between the production cultures of broadcast television and SVOD, in which the latter generates more data but the former makes this information more openly available (for example, through the daily and weekly ratings published via Nielsen and BARB or sent directly to broadcasters). What this distinction might mean for the production of television is unclear but certainly warrants further investigation.

**Conclusion**

This article has sought to advance a largely theoretical discussion around Big Data and television ratings by paying closer attention to the mechanics of this emerging research paradigm and through offering more concrete examples of how the industry is responding to these developments. And while we are only in the embryonic stages of the industry’s engagement with Big Data, it is possible to draw some initial conclusions from the examples and provocations discussed above. Firstly, a closer analysis of the mechanics of Big Data suggests that its bias towards real-time analytics encourages the production of, or increased investment in, certain genres (see Sørensen, 2016) and potentially changes the terms of what constitutes a success – though the wider implications of this remain unclear. Secondly, in challenging the credulity of Big Data advocates such as Abraham and Fairhead, I have argued that this model of research may not be the most appropriate way to analyse television audiences, with the scale of data ultimately increasing the chances of apophenia. However, it is worth noting that contrary to claims that Big Data is creating a culture of ‘knowing without understanding’ (Andrejevic, 2014: 21), there is little evidence to suggest that this is the case within television production, as evidenced by Willimon’s and Hastings’ comments. Thirdly, and finally, an analysis of current developments in the ratings industry provides significant evidence of a data divide that is transforming the production dynamics of television in complex and unexpected ways. As I have argued above, the data divide is simultaneously reinforcing existing power structures (particularly in the case of transparent infomediaries such as Nielsen and BARB) but is also opening up new opportunities for networks, producers and writers, in some instances making it easier for them to gain a direct line to viewers. At the risk of constructing my own apophenic narrative of the ratings industry that simplifies its response to these developments, one pattern that emerges is that whereas Nielsen and BARB are using Big Data in order to compliment and reinforce their *quantitative* approach to audience research, it appears that some networks and producers are using it to pursue their own *qualitative* research, subsequently feeding this data back into the creative process. Either way, it is clear that the move towards Big Data and social media metrics is having an effect on the industry that is comparable to, if not greater than, the changes that took place in the US in the early 1970s.

While this article has focused on just a few of boyd and Crawford provocations, there is much more that could and should be said about those that were omitted from this account. For example, the issue of ethics will become more pronounced once planned reforms to EU data regulation come into force in 2018 (European Commission, n.d.). Indeed, this proposed reform will have significant implications for how data can or cannot be used by those working within television (and elsewhere), with a move towards stronger privacy for consumers that may limit the availability, spread and influence of data, which may damage the democratic potential of Big Data and reinforce the data divide.

Though the full implications of this emerging research paradigm remain unclear, it is safe to say that Big Data will play a pivotal role in shaping the future of television – whatever that may be.

**Notes**

1. Laney’s (2001) original formulation of the “three Vs” of Big Data referred to what he perceived as the three main challenges of this emerging industry: the volume, velocity and variety of information. Since then, others have added to Laney’s “three Vs”. For instance, Eileen McNulty (2014) suggests that in addition to volume, velocity and variety, data analysts also face the problem of variability, veracity, visualisation, and value. [↑](#endnote-ref-1)
2. For instance, Big Data is widely recognised as playing a pivotal role in the discovery of the Higgs boson particle in 2013. [↑](#endnote-ref-2)
3. It should be noted that Braun uses “transparency” in the definitional sense of something that is invisible or difficult to observe, rather than something which is readily exposed to scrutiny. As such, I would suggest that *opaque intermediaries* might be a more accurate way to describe the ratings industry today. [↑](#endnote-ref-3)
4. In fact, Lahey cites Braun’s concept of “transparent intermediaries” in his own analysis of APIs. [↑](#endnote-ref-4)
5. Though Morris recognises that the (re)presentation of culture today is the product of both intermediaries *and* infomediaries, it would be misleading to suggest that the latter is solely a computational process. Algorithms are ultimately designed by individuals tasked with achieving specific goals, economic, cultural or otherwise. [↑](#endnote-ref-5)
6. It is interesting to note that this development was omitted from Nielsen’s interactive timeline. [↑](#endnote-ref-6)
7. For a more detailed account of how Big Data is transforming scientific paradigms, see Kitchin (2014). [↑](#endnote-ref-7)
8. It should be noted that this is not the order that the provocations appeared in boyd and Crawford’s original account. [↑](#endnote-ref-8)
9. It’s worth noting that social ratings only include Twitter activity from within the U.S. As a live global broadcast, it’s fair to assume that the overall number of Tweets (i.e. including those outside of the U.S.) would have been significantly higher. [↑](#endnote-ref-9)
10. At the time of writing, social TV ratings by Nielsen are divided into two groups: specials and series, and sports events. [↑](#endnote-ref-10)
11. This figure includes live viewers and those that watched the programme via their DVRs within 24 hours of the original broadcast. [↑](#endnote-ref-11)
12. As I suggest here, “Indirect data” (namely data generated *about* television such as Tweets and social posts) exacerbates the problem of misrepresentation as it privileges certain types of audiences and audience behaviours. At the same time, however, it often lacks contextual information about the user such as age, sex, profession, etc. This is less of an issue for “direct data” (namely data generate *through* interactions with television) such as engagements registered through Netflix as these can be linked to specific demographic data due to the existence of a billing relationship with the customer, as well as the option for different login identities for various household members which can provide further granular data about viewers and viewing habits. However, “direct data” is still susceptible to misrepresentation as these login identities do not require specific demographic data (only the “account holder” must provide this information) nor can interactions generated through services such as Netflix shed light on a range of viewing behaviours, including how many people are watching at any given time. In either case, this potential for misrepresentation of the audience is somewhat off-set by the fact that organisations such as Nielsen and BARB use this data to complement their existing research strategies. [↑](#endnote-ref-12)
13. Though these figures don’t discriminate between general traffic and data that would be usable for the purposes of audience research, they do act as a barometer for the proliferation of data more broadly. [↑](#endnote-ref-13)
14. It’s worth noting that in 2014 Twitter set up a series of awards that granted a number of academic projects access to their firehose. However, it’s also worth noting that they only permitted access to six projects out of a possible 1,300 applications and that the initiative was a one-off. See Nielsen (2014).

**References**

Abraham D (2014) *GETIF 2014 – The James MacTaggart Memorial Lecture: David Abraham.* Available at: <https://www.youtube.com/watch?v=Ok0P7E57UN4> (accessed 1 August 2016).

Andrejevic M (2013) *Infoglut: How too much information is changing the way we think and know.* New York: Routledge.

Andrejevic M (2014) Big Data, Big Questions | The Big Data Divide. *International Journal of Communication* Vol 8 (17): 1673-1689.

Baack S (2015) Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data and Society.* July-December: 1-11.

Balnaves M and O’Regan T (2011) *Rating the audience: the business of media*. London: Bloomsbury Academic.

BARB (2015) Why is BARB planning a hybrid future? [Online Video]. Available at :<http://www.barb.co.uk/project-dovetail/> (accessed 1 August 2016).

Bermejo F (2009) Audience manufacture in historical perspective: from broadcasting to Google. *New Media and Society* 11(1-2): 133-154.

Bourdeiu, P (1984) *Distinction: a social critique of the judgement of taste.* London: Routledge & Kegan Paul.

boyd D and Crawford K (2011) Six provocations for Big Data. *Social Science Research Network*. Available at: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431> (accessed 20 June 2016)

Braun J (2014) Transparent Intermediaries: Building the Infrastructures of Connected Viewing. In: Holt J and Sanson K (eds) *Connected viewing: selling, streaming and sharing media in the digital era*. New York; London: Routledge, 124-143.

Buzzard K (2012) *Tracking the audience: the ratings industry from analogue to digital.* New York; London: Routledge.

Carlson M (2006) Tapping into TiVo: Digital video recorders and the transition from schedules to surveillance in television. *New Media and Society* 8(1): 97-115.

Cisco (2015) The Zettabyte Era–Trends and Analysis. Available at: <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html> (accessed 1 June 2016)

Clarke M J (2013) *Transmedia television: new trends in network serial production*. London: Continuum.

Dixon J C, Singleton Jr. R A and Straits B C (2015) *The Process of Social Research.* New York; Oxford: Oxford University Press.

Ellis J (2000) *Seeing things: television in the age of uncertainty.* London: I. B. Tauris.

European Commission (n.d.) Reform of EU Data Protection Rules. Available at: <http://ec.europa.eu/justice/data-protection/reform/index_en.htm> (accessed 1 August 2016).

Feuer J (1984) MTM Enterprises: an overview. In: Feuer J, Kerr P and Vahimagi T (eds), *MTM ‘quality television’.* London: BFI, 1-32.

Gitlin T (1994) *Inside Prime Time.* Hoboken: Taylor and Francis.

Gray J (2010) *Show sold separately: promos, spoilers and other media paratexts.* New York: New York University Press.

Greenwood R, Suddaby R and Hinings C R (2002) Theorizing Change: The Role of Professional Associations in the Transformation of Institutionalized Fields. *Academy of Management Journal* 45: 58-80.

Guthrie M and Rose L (2013) Networks Scramble to Boost Live Programming as DVRs Shift Audiences. *The Hollywood Reporter* [online]. Available at: <http://www.hollywoodreporter.com/news/networks-scramble-boost-live-programming-423121> (accessed 22 June 2016).

Howard A (2015) ‘House of Cards’ Creator Beau Willimon: “I Have No Idea How Many People Have Watched the Show on Netflix”. *The Hollywood Reporter* [online]. Available at: <http://www.hollywoodreporter.com/news/house-cards-creator-beau-willimon-801280> (accessed 2 August 2016)

Jenkins T (forthcoming) Netflix’s geek-chic: How one company leveraged Big Data to change the entertainment industry. *Jump Cut.*

Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data and Society* 1(1): 1-12. Available at: <http://bds.sagepub.com/content/1/1/2053951714528481> (accessed 1 May 2016).

Kitchin R. and McArdle, G. (2016) What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data and Society* 3(1): 1-10. Available at: <http://bds.sagepub.com/content/3/1/2053951716631130> (accessed 1 May 2016).

Kosterich A and Napoli P. M. (2016) Reconfiguring the Audience Commodity: The Institutionalization of Social TV Analytics as Market Information Regime. *Television and New Media* 17: 254-271.

Kuchinskas S (2014) How Social Media Analytics Are Changing TV Promos. In: *Contently*. Available at: <https://contently.com/strategist/2014/04/21/how-social-media-analytics-are-changing-tv-promos/> (accessed 20 May 2015).

Lahey M (2016) Invisible actors: web application programming interfaces, television, and social media. *Convergence: The International Journal of Research into New Media Technologies* 22: 426-439.

Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. Report, Gartner, USA, February. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 15 March 2016).

Lee H J and Andrejevic M (2014) Second-Screen Theory: From the Democratic Surround to the Digital Enclosure. In: Holt J and Sanson K (eds) *Connected viewing: selling, streaming and sharing media in the digital era*. New York; London: Routledge, 40-61.

Manovich L (2009) Cultural Analytics: Visualizing Cultural Patterns in the Era of “More Media”. *Domus*. Spring.

Manovich L (2012) Trending: The Promises and Challenges of Big Social Data. In Cold M. K. (ed) *Debates in the Digital Humanities*. Minnesota: University of Minnesota Press. Available at: <http://dhdebates.gc.cuny.edu/debates?id=2> (accessed 1 August 2016).

McNulty E (2014) Understanding Big Data: The Seven V’s. *Dataconomy.com*. Available at: <http://dataconomy.com/seven-vs-big-data/> (accessed 22 May 2015).

Meehan E (1990) Why we don’t count: The commodity audience. In: Mellencamp P (ed) *Logics of Television: Essays in Cultural Criticism.* Bloomington; Indiana: Indiana University Press, 117-137.

Morris J W (2015) Curation by Code: Infomediaries and the Data-Mining of Taste. *European Journal of Cultural Studies* 18(4-5): 446-463.

Napoli P M (2014) On Automation in Media Industries: Integrating Algorithmic Media Production into Media Industries Scholarship. *Media Industries Journal* 1(1). Available at: <http://www.mediaindustriesjournal.org/index.php/mij/article/view/14/60> (accessed 11 Nov 2015).

Netflix (2016) Netflix is Now Available Around the World. Available at: <https://media.netflix.com/en/press-releases/netflix-is-now-available-around-the-world> (accessed 1 August 2016).

Nielsen (2012) Nielsen and Twitter Establish Social TV Rating. Available at: <http://www.nielsen.com/us/en/press-room/2012/nielsen-and-twitter-establish-social-tv-rating.html> (accessed 11 Nov 2015).

Nielsen (2013) Nielsencelebrates 90 years of innovation. Available at: <http://www.nielsen.com/cn/en/insights/news/2013/nielsen-celebrates-90-years-of-innovation.html> (accessed 3 August 2016).

Nielsen (2014) Twitter #DataGrants selections. Available at: <https://blog.twitter.com/2014/twitter-datagrants-selections> (accessed 10 August 2016).

O’Brien C (2016) Netflix reed Hastings says decisions on programming based on both data and gut instincts. *Venturebeat.* Available at: <http://venturebeat.com/2016/01/18/netflixs-reed-hastings-says-decisions-on-programming-based-on-both-data-and-gut-instincts/> (accessed 1 August 2016).

Proulx M and Shepatin S (2012) *Social TV: How Marketers Can Reach and Engage Audiences by Connecting Television to the Web, Social Media, and Mobile.* Hoboken: Wiley.

Reevell P (2015) Safety in Numbers? In: K7 Media. Available at: <http://k7media.co.uk/2015/02/23/safety-in-numbers/> (accessed 1 August 2016).

Sørensen I E (2016) The revival of live TV: liveness in a multiplatform context. *Media, Culture & Society*. 38(3): 381-399.

Stone M L (2014) Big Data for media. Report, Reuters Institute for the Study of Journalism (University of Oxford), UK, November.

Striphas T (2015) Algorithmic Culture. *European Journal of Cultural Studies* 18(4-5): 395-412.

Uricchio W (2011) The Algorithmic Turn: Photosynth, Augmented Reality and the State of the Image. *Visual Studies*26: 25-35.

Uricchio W (2015) Recommended For You: Prediction, creation, and the cultural work of algorithms. *The Berlin Journal* 28, 6-9. Available at: <https://www.yumpu.com/en/document/view/38991825/spring-2015-issue-28> (accessed 1 August 2016).

Watercutter A (2013) Twitter Ratings Are Here, But No One Knows What They Really Mean. *Wired.* Available at: <http://www.wired.com/2013/10/twitter-tv-ratings/> (accessed 22 Nov 2016).

Zelenkauskaite A & Bucy E P (2016) A scholarly divide: Social media, Big Data, and unattainable scholarship. *First Monday: Peer reviewed journal on the Internet.* Available at: <http://firstmonday.org/ojs/index.php/fm/article/view/6358/5511#1a> (accessed 24 Jan 2017). [↑](#endnote-ref-14)