Millington, B. & Millington, R., Forthcoming. 'The Datafication of Everything': Towards a Sociology of Sport and Big Data. *Sociology of Sport Journal*.

'The Datafication of Everything': Towards a Sociology of Sport and Big Data

The 2011 film *Moneyball*, an adaptation of the novel of the same name by Michael Lewis (2004), was met with widespread acclaim upon its release. Starring Brad Pitt as Billy Beane, the real life General Manager of Major League Baseball's (MLB) Oakland A's, and Jonah Hill as Beane's assistant 'Peter Brand' (a fictional character presumably based on actual MLB figure Paul DePodesta), the film depicts the small-market A's' struggle to compete with baseball's financial titans – in particular the New York Yankees. As Lewis' subtitle, *The Art of Winning an Unfair Game*, foretells, *Moneyball* centres on the most rehearsed of sport film conventions: the underdog tale, awash with uplifting sentiment. Without the financial means to compete in baseball's open marketplace, Beane's roster for the 2002 season was seemingly destined to fail. Despite a successful season the year prior, star players had become too expensive for the A's to retain, and had thus been replaced with cast-offs from other teams. As Lewis writes, there was a perception as well that Oakland had been fortunate in achieving prior success: "This was the year the luck of the A's was meant to run out" (p. 123).

Predictions of the A's' demise initially appeared well founded, as the team struggled at the outset of the 2002 season. Yet with some in-season manoeuvring, Beane helped spur his team to an American League record 20-game winning streak. A dramatic homerun from first baseman Scott Hatteberg – once unwanted by anyone but Beane – sealed the feat: a true event seemingly built for the big screen.

The A's success in 2002 was thus as spectacular as it was unpredictable. Behind the team's performance on the field, however, laid a second, even more important, underdog tale. Not only had his team pulled off an astonishing run, Billy Beane himself had transcended the alleged myopia of traditional baseball analytics. On one level, this meant questioning the merits of established performance measures whilst privileging others. On Base Percentage (OBP) was deemed more valuable to team success than the statistic Batting Average, for example, and thus a better indicator of individual ability as well. At a deeper level, though, Beane and DePodesta were putting into practice a way of thinking about baseball that for some time had been confined to the game's margins. In this approach, the belief was not just that statistics like OBP were different or even better than traditional measures, it was that data could be collected and analysed more rigorously and comprehensively so as to yield objective insights into the game. Said another way, established baseball statistics were not just faulty, *they were faulty in that they were too subjective*. The Moneyball methodology – the art of winning an unfair game – was to use objective analysis to uncover and then exploit market inefficiencies.

In this paper, we argue that the story of Moneyball is indicative of a wider and indeed highly important trend: the adjoining of sport with the phenomenon termed 'Big Data'. As we endeavour to show herein, statistical movements are far from new, and far from novel in their intersection with sport. In the 'Age of Big Data', however, the quest for human measurement is emboldened in new ways, and to an entirely new extent. Big Data presumes, on the one hand, that data can be collected in relation to virtually all aspects of life. It promises, on the other, that data can be 'crunched' on a scale once unimaginable, and at remarkable speed as well. The outcome for Big Data's proponents is ostensibly 'progress' – including progress in the realm of sport. As we endeavour to show herein, however, progress is not, nor has it ever been, an apolitical construct.

In pursuing these topics, we begin from the presumption that a critical and contextual sociological analysis of sport's intersection with Big Data is at this point overdue. Both Big Data and Moneyball have escaped attention in the sociology of sport, even though they have garnered substantial mainstream interest, and even though the scientific rationalizing of sport has long been a concern among sport scholars (e.g., Bale, 1993; Giulianotti, 2005; Shogan, 1999; Vertinsky, 2003). Herein, we aim to make an initial contribution towards addressing this gap. Specifically, having historicized Big Data and explained its key tenets, the paper presents four overlapping postulates on sport's relationship with Big Data. These are as follows: 1) that sport's recent statistical turn exists in reciprocity with the wider Big Data movement; 2) that the growing presence of advanced analytics in high level sport is widely deemed a progressive trend; 3) that Big Data is growing increasingly impactful *across* the sporting landscape; and 4) that Big Data has its discontents. Taken together, we see Big Data as a dynamic and growing network – one that holds sport as an important component, and one that should inspire caution and concern as much as it has excitement. Our concluding section emphasizes the need for future research on the trends discussed herein.

An Avalanche of Printed Numbers

As said above, Big Data is far from the first statistical movement known to humankind. "Since biblical times governments have held censuses to gather huge data sets on their citizenry," write Viktor Mayer-Schonberger and Kenneth Cukier (2013) in their recent book Big Data, "and for two hundred years actuaries have similarly collected large troves of data concerning the risks they hope to understand – or at least avoid" (p. 14-15). As Ian Hacking (1990) recounts in detail, the 18th and 19th centuries were particularly crucial along these lines. For example, statistical interventions were taken up across Europe at this time, as states pursued elaborate numerical inquiries to better 'know' their populations. The outcome of such activity was a veritable 'avalanche' of printed numbers. Vital statistics – rates of birth, mortality, morbidity, longevity, health, illness, and so on – were especially important (also see Hunt & Wickman, 1994).

Much the same was true at this time across the Atlantic as well. In the United States, for example, the census grew from asking four questions of family households in its original manifestation in 1790 to 13,010 questions, now directed to both families and a vast range of social institutions (farms, hospitals, businesses, and so on), 100 years later (Hacking, 1990, p. 2).

Yet Hacking adds that this numerical deluge was only an outcome: "Behind it lay new technologies for classifying and enumerating, and new bureaucracies with the authority and continuity to deploy the technology" (p. 3). That is to say, in one sense, the statistical ethos in the West gave rise to a technocratic class interested in unearthing the 'truths' of human vitality. At the moment of the French Revolution, for example, France's Marquis de Condorcet advocated for the 'moral sciences' as means for establishing statistical laws on human experience (Hacking, 1990, p. 38). Years later, Belgian statistician Adolphe Quetelet conceived of *l'homme moyen* – the average man – as the standard bearer for statistical normalcy, while Alfred Binet developed enumerating tests of reasoning *en route* to establishing the 'Intelligence Quotient', or IQ. Scientists would take up both concepts with fervour in the years that followed (Gould, 1996). In England, Francis Galton – cousin of Charles Darwin – became a veritable 'apostle of quantification', entrusting that "with sufficient labour and ingenuity, anything might be measured, and that measurement is the primary criterion of a scientific study" (Gould, 1996, p. 107).

At the same time, and as Hacking observes, technologies of classification and enumeration became ever more crucial in pursuing these ends. Perhaps no innovation better exemplifies this than Herman Hollerith's 'Electric Sorting and Tabulating Machine', an early forerunner to the modern computer. With the United States census growing evermore unwieldy, Hollerith began in 1881 to design a mechanical tool for sorting its results. Human census takers would still be deployed by the tens of thousands, but now their results would be inscribed on a punch card. As explained in Joanne Weisman Deitch's (2001) book <u>A Nation of Inventors:</u>

> Each card represented one person and each hole a different statistic, such as age or marital status. The cards were sorted and later read electronically by a press containing pins that penetrated the card only at its holes. Each pin that passed through a hole made electrical contact with a small cup of mercury, closing a circuit and advancing a dial counter by one (p. 42).

In Bruno Latour's (2005) terms, the labour of sifting through and organizing census data had been translated from human to non-human actants. With Hollerith's invention in tow, the 1890 census was completed in one year – 7 years *less* than the survey of 1880. As we shall see, this trend towards rapid number sorting has only intensified in the Age of Big Data.

The question remains as to the rationale for these developments. For Hollerith it was certainly to speed the rate of data sorting. Others, though, mounted grander claims. In the Enlightenment tradition, some averred that numbers, together with scientific reasoning, were made for the betterment of democracy. True to the revolutionary moment in which he lived, the aim of Condorcet in France "was to free the people from the ignorance and error upon which political despotism depended" (Rose, 1999, p. 201). The famous nurse Florence Nightingale furthermore used statistics to advocate for improved sanitary conditions in the hospitals in which she worked (Porter, 1986). Statistics were also eagerly adopted by industry at this time. Most notably, Frederick Winslow Taylor surmounted 'scientific management' as a cure to the inefficiencies that in his view ailed the American shop floor. This involved re-configuring and improving machinery. But it also meant parsing workplace activities "to the extent that each task was fragmented into its smallest constituent units which would be timed and measured" (Grint, 2005, p. 178). The division of labour was divided more so than ever.

Yet the facile link between quantification and 'progress' has proved grounds for contestation as well. Patricia Vertinsky (2003) reminds us that early statisticians tended to fervently embrace eugenics, with leading thinkers such as Francis Galton seeing deviance from statistical norms as reason for trepidation over the wellbeing of the body politic. We have seen how in Galton's eyes measurement was to be transposed onto all of human activity. The corollary, though, was that enumerating the body and its habits would foment literal hierarchies of (for example) ability and reproductive potential, and thus determine 'viable' candidates for marriage, reproduction, and other biopolitical endeavours (p. 102). Similarly, though Alfred Binet did not wish his IQ tests to become infallible metrics for ranking mental capacities, his intentions could not prevent many in the late 1800s and early 1900s from mobilizing them precisely in this way (Gould, 1996).

The Birth of the Baseball Box Score

Sport and physical activity, it is crucial to say, were not free from the statistical 'avalanche' that was re-shaping social institutions at this point in time. Pre-modern games were giving way to modern competitions (cf., Bale, 1993) – a transition that lent itself to new forms of analysis.

Baseball presents an apt case in this regard. In the 1800s baseball was evolving like other sports: harmonising rules across space, standardizing equipment, and, most significantly for these purposes, developing new forms of data inscription (see Guttman, 2007). A key figure along these lines was British-born sportswriter Henry Chadwick. Scanning the baseball landscape, Chadwick saw no analogue for the cricket 'box score' (as in Britain), which is to say there was no meticulously conceived alpha-numeric system for recording player performance. Chadwick set out to remedy this: in his scorebook, 'A' would stand for first base, 'B' for second, 'C' for third, 'L' for foul balls, 'K' for strike out, and so on. Players on each team were ascribed numbers 1 through 9. This system was subject to many changes over the years, some in the interest of rendering it more accessible to fans. Yet the important point is that players and their on-field activities were codified so that they could in turn be permanently inscribed in text (cf., Schiff, 2008). Fans, statisticians, and others – even those far afield – were given a new way of following their players and teams of interest.

As Chadwick's biographer Andrew Schiff (2008) writes, it was not an utter coincidence that baseball was quantified in this way in the mid-1800s. "Statistics, in the middle of the nineteenth century, also began to represent something more than just mere quantification: it was a way of proving the truth" (p. 83). Chadwick himself saw statistics as means for revealing indisputable insights into 'America's pastime'. As he wrote in 1864:

> Many a dashing general player, who carries off a great deal of *éclat* in prominent matches, has all "the gilt taken off the gingerbread," as the saying is, by these matter-of-fact figures, given at the close of the season; and we are frequently surprised to find that the modest but efficient worker, who has played earnestly and steadily through the season, apparently unnoticed, has come in, at the close of the race, the real victor (cited in Schiff, 2008, p. 84).

That industry in general was interested in worker efficiency at this time, and increasingly gleaned knowledge of efficiency through formalized measures, is a point not lost on Schiff.

Indeed, the synergy between the increasingly industrialized economy and the increasingly rationalized sporting landscape ran deep. In the late 1800s, a style of play called 'scientific baseball' emerged and was stamped for approval by many of the game's luminaries – Chadwick among them. As Puerzer (2002) recounts, scientific baseball was anathema to "the 'manly game' of slugging and aggressive play" (p. 37). Managers embracing this philosophy valued precision (place hitting or stealing bases) and strategic sacrifice (exchanging 'outs' to advance runners) so as to 'manufacture' runs. As the language of 'manufacturing' foretells, the important point is that baseball managerialism was reflecting industrial managerialism in key ways. As one example, specialized workers were central to scientific baseball, as in the case of 'pinch hitters' who were called upon in specific cases to enhance the odds of a positive outcome. This relationship was furthermore a reciprocal one: "Frederick Taylor, the 'Father of Scientific Management', used a baseball team as a metaphor in the description of a well-working company that he presented before Congress in 1911" (Puerzer, 2002, p. 37; cf., Risker, 1995). Sport was thus reflecting the Taylorist ethos; Taylorism drew on baseball for support.

The Age of Big Data

It is beyond the scope of this paper to chart every step in the transition from these earlier days of statistical analysis to our present-day conjuncture. It is worth highlighting, however, that the mid-1900s first brought the articulation of statistical analysis and the political rationality of welfarism. Even the mighty US census could not fully account for the plight of the unemployed in the Depression years. The Roosevelt administration thus set out to fortify the country's statistical apparatus: "More people were employed in counting and in analysing numbers; more things were counted; more numbers were published" (Rose, 1999, p. 227). Further advances in computing technology – a by-product of wartime investment – were also important in this regard. The switch from mechanical (as in Hollerith's machine) to electronic means of compiling census data promised similar benefits to the earlier switch from human to mechanical labour. The upshot in the eyes of statisticians was savings in terms of time and money, as well as greater accuracy in data processing. In this sense, much as they did in the late 1800s, "the [Census Bureau] again led the way in data processing for the rest of the government and the private sector" (Anderson, 1988, p. 197).

Computing technologies of course only evolved further through the latter half of the 1900s. In turn, it is an increasingly accepted premise that computers, together with drive to know the (consuming) population, have helped spur the arrival of an Age of Big Data. Advances in processing power are crucial in this regard – data is so easily amassed, stored, and evaluated that these procedures can be carried out on a scale once unimaginable. That Big Data has ramifications for sport, and vice versa, is also an increasingly accepted view.

A series of 'V' terms figure prominently in many definitions of Big Data. 'Volume' is the first of these, for as the company IBM – unsurprising proponents of new forms of statistical analysis – remarks, data can now be collected from virtually anywhere: "sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few" (IBM, n.d.a). In fact, before Big Data became a fashionable term, 'The Petabyte Age' was suggested as an appropriate designation for the current moment in time (Anderson, 2008). A petabyte is a quadrillion bytes, and thus the term bespeaks the tremendous storage demands now placed on statisticians. Google, for instance, is said to process 24 petabytes of data per day, while Facebook sees 10 million photo uploads per hour (Mayer-Schonberger & Cukier, 2013, p. 8). The central flaw of 'The Petabyte Age' nomenclature is thus its pending obsolescence: "inevitably, petabytes of data will give way to even bigger bytes: exabytes, zettabytes and yottabytes" (Lohr, 2012). Furthermore, this new information is both registered and relayed at rapid pace: 'velocity' is thus a second 'V' term to go along with 'volume'.

‘Variety’ is a third commonly cited component of Big Data, and is perhaps the most crucial one of all. In an important turn of phrase, Mayer-Schonberger and Cukier (2013) point to the nascent ‘datafication of everything’. It is not just easily enumerated occurrences and activities – rates of birth and death, for instance – that are of interest to statisticians, but *virtually every aspect of daily life*. ‘Geo-loco’ information, for instance, which refers to one’s (changing) spatial location, is incredibly valuable to industry, as its ‘datafication’ presents the possibility of geographically-targeted advertising. Likewise, the datafying of personal interactions through their inscription on platforms like Twitter can yield precious insights into consumer tastes. The ‘datafication of everything’, then, is distinctively rhizomatic. Data collection is ‘flattened’ out, as it stems from everyday activities and not simply from the purposeful initiatives of institutions like the Census Bureau. Data itself, meanwhile, is molecular, pertaining to subjects deemed divisible in seemingly countless ways (cf., Deleuze, 1995).

The Age of Big Data thus seems to have vindicated Francis Galton’s original supposition that ‘anything might be measured’. And yet the same problem that befell statisticians centuries ago lingers still – namely, how to transform seemingly overwhelming troves of information into an easily digestible form. As such, what has arisen in combination with the ‘datafication of everything’ are technologies for enhanced data processing. These are updates, one might say, to Hollerith’s original tabulating machine.

Perhaps the best case in point along these lines is the company Yahoo’s ‘Hadoop’ apparatus, described by *Wired* writer Cade Metz as “an open source platform designed to crunch epic amounts of data using an army of dirt-cheap servers” (Metz, 2011). The company eBay, for instance, is said to employ Hadoop in combination with 2500 servers in its data processing initiatives. The ideal outcome of arrangements of this kind is the mitigation of ‘dark data’, meaning data that, while easily collected, goes unseen by those who might (literally) capitalize upon it. As Luc Burgelman (2013) writes, whereas a bank or retailer might typically look at customer transactions when considering a promotional event, this is only part of the picture in the Age of Big Data. “The highly valuable data about who these customers are interacting with and what they are saying in social mediums — like how they feel about brands, what they are looking for, where they shop, their personal network, a customer service experience they had as they walked into a bank or store — is left dark.” Like Hadoop, SAP’s HANA database exemplifies this trend. According to Ohlhorst (2012), it was used by the telecommunications company T-Mobile “to mine data on its 30 million U.S. customers from stores, text messages, and call centers to tailor personalized details” (p. 85).

As implied in these examples, there is a discernible consumerist ethos driving Big Data forward. Big Data has certainly been linked to ‘progress’ in medicine, politics, and other realms, as past statistical movements were (e.g., see Arnold, 2013). Yet the desire to know the population in its intimacy now tends to veer towards gleaning information on (for example) taste preferences in the interest of profitability. As Mayer-Schonberger and Cukier (2013) write, Mastercard does not just aggregate and analyze “65 billion transactions from 1.5 billion cardholders in 210 countries in order to divine business and consumer trends.” The next step is key as well: “Then it sells this information to others” (p. 127). As those writing in Foucault’s tradition have argued, political rationalities have long been enmeshed with technologies of enumeration. Recalling the work of Francis Galton, among others, classical liberalism deployed measurement to embolden the state, shape and mould the body, and engender new (rational) ways of knowing. Quantification thus formed a technological apparatus – one that, in its darkest form, was mobilized to prevent the reproduction of ‘undesirables’. Welfarism saw the making of new statistical apparatuses in the quest to balance economic progress and social security (Miller & Rose, 2008). For its part, neoliberalism employs numbers to construct a radically market-oriented society – one where the state’s relevance recedes, and where citizens become, inexorably, consumers. In this sense, the drive to compile and ‘crunch’ vast (‘volume’) and complex (‘variable’) troves of information at breakneck speed (‘velocity’) is not solely a product of innovation and humankind’s longstanding interest in quantification. Big Data is also a product of its political economic context – one that, as Nikolas Rose (1999) says, is bent on knowing the 'soul' of the consumer-citizen.

Sport in the Age of Big Data

To link these wider developments to the sporting context it is helpful to return yet again to the case of baseball. Though 'scientific baseball' retained a place of high prominence for some time (and lingers today under the name 'small ball'), the post-war years brought scepticism about the validity of its basic principles. In 1971, a small group of baseball enthusiasts met in Cooperstown, New York (home of the sport's Hall of Fame) to create a formal organization of baseball historians and statisticians. The result was the formation of the Society for American Baseball Research (SABR), an organization replete with objectives such as fostering the study of baseball and establishing accurate historical accounts (SABR, 2013). As much as these guiding objectives, though, SABR came to be known for its tendency to look sceptically upon established baseball wisdom.

It is here that statistics become relevant. Henry Chadwick's method of quantifying baseball, devised in the 1800s, was undoubtedly influential. As time passed, however, it was increasingly seen as highly *subjective* as well. His strident belief, for instance, that a walk (i.e., four balls to one batter) stemmed from the pitcher's ineptitude, rather than the batter's trained eye, caused him to devalue (in an analytic, and not economic, sense) this means of reaching first base in relation to obtaining the same result via a base hit. Walks were thus not counted in one's total at bats, and were in turn left absent from Batting Average (hits/at bat) – a measure deemed crucial in determining a player's ability (Schiff, 2008, p. 85). Likewise, fielding 'Errors' were recorded in the box score at the statistician's behest. "What is an error?" famed SABR member Bill James would ask years later in response. "It is without exception, the only major statistic in sports which is a record of what an observer thinks *should have been accomplished* … It is, uniquely a *record of opinions*" (cited in Lewis, 2004, p.66, emphasis in original).

In fact, James is a key figure in this narrative. One of the original members of SABR's Statistical Analysis Research Committee, in the late 1970s James began producing his annual *Baseball Abstract* – initially a self-published document featuring statistically-informed analyses of, and at times punchy diatribes against, established baseball knowledge. At one level, this meant attacks on specific measures like Batting Average. As James and other statisticians pointed out, the omission of walks from this measure was highly problematic. Earning walks was a skill, the argument went, not simply a product of pitcher ineptitude, making 'On Base Percentage' a far better indicator of a player's true value. More broadly, James was helping to popularize 'sabermetrics' – a term he coined to honour SABR, and one that speaks to the use of analytics in "the search for objective knowledge about baseball" (Birnbaum, 2013).

As James and others set out to deconstruct and re-build baseball's analytical foundations, they were first stymied by a lack of information upon which to base their work. "All I have is the box scores" James wrote in an early *Abstract* (cited in Lewis, 2004, p. 82), and the box scores were to his reading woefully inadequate. As such, and mirroring the deployment of human census takers a century prior, by 1984 the sabermetric community was dispatching volunteers, hundreds strong, to ballparks across the United States with the aim of collecting data needed for the new scientific study of baseball (Lewis, 2004). The game was in effect witnessing its own 'avalanche' of printed numbers – a trend that in turn engendered a dizzying array of new baseball statistics as well. Measures like On-base Plus Slugging (OPS) and Inside-the-Zone Contact Percentage (Z-Contact%) became the erudite cousins of cruder statistics like Batting Average. If consumers are now 'dividuals', to use Deleuze's (1995) term, meaning entities that are broken down into specific parts, tastes, behaviours, and so on, baseball players are seen in much the same manner in the eyes of the sabermetric community. What is more, discrete measures of player performance are commonly *reassembled* through statistics that conceptualize a player's value as a whole. In re-building baseball analytics then, James and other sabermetricians sought to replace traditional, subjective, and ‘flawed’ measures of ability (e.g., Batting Average and Errors) with more accurate and objective forms of assessment.

To put these developments in other terms, as time passed baseball statisticians sought to increase the *volume* and *variety* of information about the game and, in turn, the volume and variety of advanced statistics as well. With online communication, so too was the *velocity* at which data travelled dramatically enhanced. In the 1980s and 1990s, however, sabermetricians still stood at the gates of baseball’s formal management structures. It is true that the 1900s saw occasional attempts at reimagining the game’s analytical foundations by teams themselves. In the late 1940s, Brooklyn Dodgers General Manager Branch Rickey adhered to the exhortations of statistician Allan Roth to employ On Base Percentage in assessing players’ on-field talents. Likewise, in the 1960s and 1970s, baseball scouting director Jim McLaughlin decried typical forms of player evaluation in his work for the Cincinnati Reds and Baltimore Orioles. His computerizing of game data in the interest of rationalizing player analysis was an act ahead of its time (Perry, 2006). Nonetheless, it was not until Billy Beane, Paul Depodesta, and the Oakland A’s embraced the sabermetric approach that the statistical revolution at the game’s periphery invaded the core.

‘Moneyball’ is effectively the pursuit of objective knowledge combined with the quest to exploit market inefficiencies. To be sure, for sabermetricians who embraced the former there were many cases of the latter in baseball’s pre-2000 conjuncture. Salaries, after all, were based on what they saw as flawed analytics (Hakes and Sauer, 2006). As said at the outset, Beane and Depodesta found great success at the margins, replacing conventional ‘sluggers’ with sabermetric darlings like Scott Hatteberg – he of the high On Base Percentage. Personnel changes were furthermore met with new strategic designs. For proponents of ‘small ball’ or ‘scientific baseball’ (which sabermetricians would deem deeply unscientific, and perhaps then ironically named), advancing runners on the base paths by sacrificing outs (e.g., by bunting) was good practice. For the likes of Beane, it was in most cases unconscionable. As said by baseball writer and analyst Dave Cameron (2012), the sacrifice bunt should be more or less excised from the Manager's playbook: "It is not always the wrong move, but it is used far too often and in too many situations where swinging away is more likely to produce a positive result." Indeed, if objective analysis is the epistemological basis of sabermetrics, the notion of probability lies at its ontological core. That is to say, the search for 'truth' through new analytics is effectively a search for *probable truths* (where is this batter most likely to hit the ball?) in the interest of leveraging probabilities so as to score or prevent runs (how can we align our defense so as to account for this batter's tendencies?). This, in Ian Hacking's (1990) terms, amounts to the 'taming of chance'. And though it might be said that leveraging probabilities was central to baseball strategy in past eras as well – even Henry Chadwick's to some extent – 'Moneyball' brings a new intensity and a new arsenal of analytical tools to this pursuit.

With time, and with the success of Beane's 'Moneyball' approach, sabermetrics infiltrated front offices across the Major Leagues. Now institutionalized, individual teams and the league at large can devote human and financial resources towards compiling and analysing ever more data. At the very least they can consult with the myriad companies that now render such services. Take, for example, PITCHf/x, a service that records the flight of pitches "to within an inch of accuracy":

> It has been installed in all 30 Major League Baseball (MLB) stadiums, and currently tracks pitches for every MLB game. The system utilizes three tracking cameras and a central pitch tracking system in every stadium. Each tracking camera records the pitch from the time it leaves the pitcher's hand until it crosses the plate, and then it sends this information to the tracking system to calculate and store the digital record of the pitch – including speed, location, and trajectory. The pitch data can then be streamed in real time for television broadcast effects, consumer applications, performance analysis, or other forms of entertainment and/or evaluation (Sportvision, 2013).

Likewise, sport franchises are now installing in-stadium cameras so as to exhaustively track player movement during games (e.g., see Lowe, 2013). The idea, then, is evidently to 'datafy everything': to collect unstructured data through camera tracking, even if all potential uses of said data are not yet known. Of course, this also generates the possibility that important data will be lost in a sea of information – that is, the possibility for 'dark data'. Technology writer Barry Eggers (2012) suggests that teams will therefore soon turn to number-crunching apparatuses like the aforementioned

Hadoop. He furthermore imagines the merits of doing so: "By having his data scientist run a Hadoop job before every game, [San Francisco Giants manager] Bruce Bochy can not only make an informed decision about where to locate a 3-1 Matt Cain pitch to Prince Fielder, but he can also predict how and where the ball might be hit, how much ground his infielders and outfielders can cover on such a hit, and thus determine where to shift his defense."

Unstructured data, motion tracking, 'data scientists': this is sport in the Age of Big Data. Human and increasingly sophisticated non-human 'actants' are now conspiring to bolster the volume and variety of data available to baseball teams, as well as the velocity with which this data moves (cf., Latour, 2005). Baseball is in turn seen and played in new ways.

Four Postulates on Sport and Big Data

With the above in mind, in the remainder of this paper we present four initial postulates on sport in the Age of Big Data. The aim in this regard is to adumbrate what appear to be basic principles underlying sport's newfound statistical turn while recognising that further theorizing and empirical research into each is still required.

Postulate 1: That sport's statistical turn exists in reciprocity with the wider Big Data movement.

This is evident from our preceding discussion, but it is worth stressing here precisely how this synergistic relationship between sport and other sectors has unfolded. In this regard John Urry's (2000) writing on mobility is instructive. Following Castells (1996), Urry sees the concept of 'network' as more apt than 'society' at characterising contemporary experience. Whereas society connotes boundaries and regional specificity, networks are "dynamic open structures" (p. 192). Moreover, they allow for the manifestation of 'flows' or 'global fluids': "heterogeneous, uneven and unpredictable mobilities of people, information, objects, money, images and risks" (p. 194).

Sport's relationship with Big Data strikes us as precisely dynamic and open in this way. On the one hand, sport's statistical turn has drawn from its wider environment. Prior to Billy Beane and the Oakland A's adoption of advanced metrics, for example, financial traders Ken Mauriello and Jack Armbruster set out to bring a Wall Street approach to the game of baseball through their company AVM Systems. Skilled in the analysis and exchange of securities derivatives – that is, fragments of stocks and bonds – Mauriello and Armbruster applied a similar logic of deconstruction to baseball's on-field events. The baseball diamond became a mathematical matrix: hits were not singles, doubles, triples, or home runs, but data points mapped on a fragmented grid. As Lewis (2004) recounts, Mauriello and Armbruster were thus shunting the conventional box score with even greater fervour than sabermetrician Bill James: "In AVM's computers the game became a collection of derivatives, a parallel world in which baseball players could be evaluated more accurately than they were in the real world" (p. 133). In this regard, people, ideas, and technologies 'flowed' from the financial sector to the sporting landscape.

But there is also an *outward* movement at play: sport itself affects its wider contexts. Writing in *The New England Journal of Medicine*, Phillips, Greene, and Podolsky (2012) advocate for 'Moneyball medicine', an articulation that combines good medical practice with cost-effective decision-making. From their perspective, not only does the Oakland A's rigorous approach to statistical analysis map nicely onto the concept of evidenced-based medicine, so too does their thriftiness: "in health care, we have been spending as if we had the budget of the [New York] Yankees — while all signs suggest we'll soon be operating more like the [A's]" (p. 1583). Perhaps unsurprisingly, these ideas have also flowed into business. Wolfe et al. (2007) draw out numerous lessons regarding the impacts of Moneyball on industry, such as the need for an 'innovation champion' not unlike Billy Beane to counter 'anti-innovation' (and presumably then unproductive) business environments. *People* can evidently 'flow' outwardly as well. The analyst Nate Silver, for example, first came to fame by devising an algorithmic system for predicting player performance in baseball, and has more recently been lauded as a seemingly clairvoyant psephologist – one focused on probabilities in politics in much the same way that sabermetricians are in baseball (see Silver, 2012).

The fundamental point here is one that sport scholars have long made: that sport reflects and contributes to its wider conditions. Moreover, in these ethno-, techno-, and ideoscapes (cf., Appadurai, 1990), Big Data is revealed to be more than a temporal qualifier (as the phrase 'the Age of Big Data' portends) or a collection of 'V' terms. It is, moreover, a dynamic, integrated network along which 'flows' can be relayed.

Postulate 2: That the growing presence of advanced analytics in high level sport is widely deemed a progressive trend.

Progress is of course a supple term. For those in front office positions in elite sport – General Managers like Billy Beane, for instance – it has a discernible economic inflection. Decisions regarding which amateur players to draft or which professional 'free agents' to sign have significant financial ramifications. Big Data in this sense can 'tame chance' by revealing insights into sport's labour market in the interest of avoiding 'burdensome' contracts. As baseball's Oakland A's learned, exploiting market inefficiencies to build a cost effective roster can have the cascading effect of bolstering fan attendance, and thus revenue as well (Hakes & Sauer, 2006).

But progress when it comes to sport and Big Data also means simply the pursuit of purportedly *better* knowledge. To reveal new insights – to make the invisible visible through data – is to allegedly excise the vagaries of subjective analysis. This point is made explicit by National Basketball Association (NBA) executives Daryl Morey and Sam Hinkie:

> In reality, the referendum on whether using objective analysis improves decision-making is long over. Industries can remain insulated for a time, but the advantage of augmenting decision-making with data is such that adoption becomes near ubiquitous over time. For example, in baseball, what started as a small movement with the Oakland A's has become routine with all 30 teams using analysis to one degree or another. While the storytelling genius of Michael Lewis

> turned baseball's adoption of analytics into a fascinating yarn, the phenomenon is actually just the mundane manifestation of the march of progress (Morey & Hinkie, 2011).

Morey and Hinkie are hardly the first figures to link statistical analysis and human progress in this way. This is a lingering Enlightenment rationality. What is significant, however, is that advanced analytics are deemed here to have reached what Žižek (1999) terms a 'post-political' state: the debate centres on how, and not whether, they should be employed within and beyond industry. 'Better knowledge' evidently is said to have cascading effects as well. For example, IBM has developed software to register biometric and psychological player data to help in the avoidance of injuries (see IBM, n.d.b).

The claim that advanced analytics are progressive in nature is therefore underpinned by a claim to epistemological objectivity. The tensions inherent to such assertions are likely familiar to sociologists of sport. The thesis at the core of Andrews' (2008) assessment of the current state of kinesiology is that an epistemological hierarchy has arisen with the scientizing of this discipline (also see Andrews et al., 2013). Liberal capitalism, with its bent towards rationality and productivity, finds 'epistemic corroboration' in positivist objectivity. "Both are constituents," Andrews writes, "and simultaneously constitutors, of a particular understanding of modernity, centered around linear evolutionary assumptions pertaining to the inevitable progress of human civilization through the advancement of empirically grounded – often a euphemism for quantitatively driven and objectively reasoned – science" (p. 48). Andrews here is effectively responding to Ingham and Donnelly's (1990) question, posed nearly 20 years earlier in light of requests for academic knowledge to be more socially applicable, as to "whose knowledges will be found to be acceptable in the marketplace of physical culture?" (p. 61). If, as Andrews avers, purportedly objective knowledge has reached the summit of kinesiology's epistemic hierarchy, it seems that in the Age of Big Data quantitative insights into sport are now growing more powerful as well.

And yet, in baseball and perhaps in organized sport writ large, so too is there a risk of overstating the influence of objective analyses (or, at least, the pursuit of objectivity) in relation to knowledge codified as subjective. Traditional, qualitative analyses are still valued in baseball. Many teams continue to rely on the 'old-guard' of baseball scouts to evaluate talent, particularly when it comes to projecting success amongst high school or college prospects. Here the localized knowledge of scouts with case-by-case accounts of 'intangible' factors is seen as a suitable companion to advanced analytics. So too can scouts combine qualitative and quantitative insights as well – for example, by turning player performance, seen with the 'naked eye', into scores on a hard numerical scale (Silver, 2012, p. 100). For the aforementioned analyst Nate Silver (2012), Big Data effectively means amassing *more* data, whether qualitative or quantitative. 'Progress' lies in large part in what you do with it thereafter. "This is the essence of [Billy] Beane's philosophy: collect as much information as possible, but then be as rigorous and disciplined as possible when analysing it" (Silver, 2012, p. 100). Indeed, it is hard to imagine a future whereby the observational skills of coaches, executives, and players, among others, will be erased from the analytics 'toolkit'. Silver's vision of productive co-existence is thus not unlike Vertinsky's (2009) view that the purported 'gap' between kinesiology's sub-disciplines is not as yawning as typically presumed, and that the divides that do in fact exist can be mended with reflexive care.

Postulate 3: That Big Data is increasingly impactful across the sporting landscape.

Our analysis thus far has centred mainly on elite sport, and men's sport in particular. This is where Big Data has to date been most discussed. We have focused even more precisely on baseball. As *Moneyball* author Michael Lewis wrote in 2009, however, "[t]he virus that infected professional baseball in the 1990s, the use of statistics to find new and better ways to value players and strategies, has found its way into every major sport." Soccer, American football, and basketball are all pertinent examples.

Without overstating Big Data's reach, as per Postulate 2, it is important to emphasise its impacts *across* sport, and not just among those comprising its new technocratic class. To do so is to first point beyond executives like Billy Beane and towards the experiences of professional and elite amateur athletes. For example, if analytics can lessen the risk of injury – or, at least, if managers and coaches put faith in this notion – it will surely impact on training regimens in turn. As a case in point, Australia's New South Wales Waratahs rugby team has taken up IBM's analytical approach to injury prevention, in part by affixing GPS trackers to individual players during matches and training. This allows the measurement and monitoring of levels of intensity, collisions, and fatigue. Said the team's Athletic Development Manager: "IBM's predictive analytics technology gives us a very objective, sensitive and reliable measure of predicting each player's limit and their injury risk and allowing us to modify training accordingly" (IBM, 2013). Once again, progress operates on many levels: improved health through better knowledge; improved performance; and, in that injuries are said to be a hindrance on merchandise and ticket sales, and thus revenue, financial success as well.

To look across the sporting landscape is furthermore to consider Big Data's potential impacts on recreational sport and fitness. Running parallel to the adoption of new technologies for data collection among professional franchises has been the proliferation of personal tracking devices for monitoring everyday behaviour. At the centre of such activity is the Quantified Self (QS) community: self-tracking devotees who find merits in registering and analysing everything from sleep, to food intake, to exercise, often with the help of mobile application software, or 'apps'. "Running programmes, for example," writes Deborah Lupton (2012) in discussing these technologies, "can be downloaded to one's smartphone or tablet computer, which are able to record the number of kilometres run each session, the route taken, automatically report these details to one's followers on social media sites, suggest new routes and remind the user that she or he has not run for a few days" (p. 231). In keeping with Silver's (2012) vision of data's variability, Melanie

Swan (2013) notes that both quantitative and qualitative data is commonly registered by QS enthusiasts (though, true to the name ‘Quantified Self’, the latter is often turned into the former). Swan makes two additional points relevant to these purposes.

The first is that health and fitness data tracking on a personal level presents challenges not unlike ‘bigger’ Big Data initiatives. Personal data sets, Swan notes, can easily grow beyond a manageable level in the process of self-assessment. Employing language popularized by Silver (2012), she adds that this can in turn make it difficult to discern ‘signal’ from ‘noise’ – that is, important from irrelevant information. A second key point is that self-tracking is increasingly a mainstream phenomenon, and not merely the prerogative of the QS community proper. Almost two thirds of the US adult population are now tracking weight, diet, and exercise; a third partake in the even more intimate assessment of matters such as blood sugar, blood pressure, headaches, or sleep patterns (Swan, 2013, p. 86). Young people too are evidently not exempt from these trends. Big data interventions are finding a place in youth sports as well – for example, through the development of software for performance tracking aimed at young people specifically (e.g., see Martin, 2014).

Finally, to look across the sporting realm is to consider consumption too. In one sense fans now have new means for evaluating and discussing their players and teams of interest. For example, the NBA has made available an online search engine containing nearly every statistic in league history. The engine is powered by SAP’s HANA platform – highlighted above for its use by T-Mobile in the process of mining customer information – and can accordingly handle 4.5 quadrillion combinations of data (Beck, 2013). Volume, in other words, accompanies variety. The English soccer club Manchester City has taken this a step further. They not only released extensive player and team data in the interest of communicating with fans, they added the explicit goal of engendering the next wave of analytics and analysts. “We will work directly with those of you who came up with good concepts,” the team said in a press release, “and also connect you to others who are working in the same research area” (Furnas & Lezra, 2012). As data analysis becomes more valuable to sport franchises, the boundaries between sport consumer and sport (intellectual) labourer grow evermore blurry.

In another sense, fans are also the *objects* of surveillance in the Age of Big Data – a reality that extends pre-existing trends in many ways. Surveillance of fans is of course far from new. CCTV cameras have long been deployed in city spaces (e.g., during sporting mega-events) and stadiums in the interest of assuring 'proper' fan comportment (e.g., see Giulianotti & Klauser, 2011; Silk, 2004; Sugden, 2012). At the same time, market research is part and parcel of consumer culture. In the early years of television, for example, media and advertising companies adopted methods like circulation audits to know consumer preferences, as well as theories from the social sciences to understand their motivations (Leiss et al., 2005). Now, however, the same mobile communication technologies that are helping recreational athletes and fitness enthusiasts track their own behaviour are also helping to conflate surveillance and market research in the name of profitability. The technology company Cisco, for example, recently introduced a 'solution' (or system) called StadiumVision® Mobile in a small selection of high profile sport venues in America and Europe. Capitalizing on the proliferation of web-enabled mobile devices like smartphones at sporting events, the system delivers both video (e.g., shots of team benches during time-outs) and data feeds (e.g., trivia contests) to fans in a highly efficient manner. The corollary, though, is that teams are given even more information upon which to base their branding decisions, target marketing, promotional activities, and so on (Cisco, 2013; also see Hsu, 2012). This is not simply surveillance, but an accelerated form of *dataveillance* – the monitoring and assessment of consumer-produced data (cf., Clarke, 1988). And in that data is personalized, voluminous, and rapidly exchanged, the imperative is less about theorizing consumer behaviour than it is about pursuing 'Big', empirical accounts of consumer activity. As said in one report on fan monitoring at sport events: "With the right data, a team now can understand the demographic makeup of its fans with far more precision than ever before" (Dorsey, 2013).

Postulate 4: That Big Data has its discontents.

In a first, specific sense, concerns arising with Big Data's arrival pertain to privacy, and thus follow directly from the heightened ease of surveillance described above. Legal scholar Paul Ohm (2012) suggests that consumer databases, if continuing on their current growth trajectory, will soon hold the potential to link every consumer to at least one secret they hold dear. The arrival of Big Data will only exacerbate this problem: "In the absence of intervention, soon companies will know things about us that we do not even know about ourselves. This is the exciting possibility of Big Data, but for privacy, it is a recipe for disaster" (Ohm, 2012). A recent study by *The Wall Street Journal* lends credence to this claim. Focused on 101 popular Smartphone apps, including those targeting sport participants and consumers, their analysis revealed the probing manner in which these technologies are 'watching' their users – often surreptitiously (Thurm & Kane, 2010). The ESPN ScoreCenter app, ostensibly made for sports fans, was found to collect data on its users' Phone ID number, location, and username/password so as to in turn share this information with Disney, the app owner. Other apps shared similar data directly with marketing firms.

The general point that follows from this specific concern is that there is reason to be wary of the facile link between data collection, objectivity, and human 'progress'. To be sure, Big Data does not seem to carry forward from past statistical movements the desire to eradicate 'lesser' populations. Big Data, and sport's dalliance with Big Data, is – as Paul Rabinow and Nikolas Rose (2006) say in a different context – about capitalism and liberalism, not eugenics, "at least in so far as eugenics has acquired an inescapably negative meaning in our contemporary culture" (p. 211). Big Data does, however, redound in a context already rife with anxiety over corporate (as well as state) surveillance. The concern, then, is that the economic variant of progress – the drive to know consumers more intimately in the name of profitability – risks overwhelming the variant that rests on the pursuit of 'better' knowledge.

Even more broadly still, and acknowledging's Silver's (2012) observation that both qualitative and quantitative data remain relevant in the sporting context, the ability of numerical analyses, more so than explicitly subjective ones, to elide the circumstances from which they arise has inspired critical commentary as well. In other words, the clear epistemological delineations described in Postulate 2 are to some extent fabrications to begin with – though they are not always taken as such. "All researchers are interpreters of data" say danah boyd and Kate Crawford (2012) in their writing on Big Data – even researchers that strive towards objective insights. To this assertion we would add that all sport statisticians are interpreters of data as well. Indeed, interpretation both follows from *and* precedes data collection. As Gitelman and Jackson (2013) write, and as described above, data collection has now become an 'always-everywhere' proposition. But while data are commonly seen to be 'collected', 'entered', 'compiled', 'stored', 'processed', 'mined', and 'interpreted', "[l]ess obvious are the ways in which the final term in the sequence – interpretation – haunts its predecessors" (p. 3). Objectivity from this view is *contextual*, "it comes from somewhere and is the result of ongoing changes to the conditions of inquiry, conditions that are at once material, social, and ethical" (p. 4). But the concern that Gitelman and Jackson mount is that is that our zealousness in the quest for more data will translate into faith about the neutrality and autonomy of numbers. For example, under such conditions, a question arises as to what will prevail: an athlete's (subjective) claim to fatigue, or an automated evaluation that yields a competing verdict. This is a question of power dynamics – one that renders Andrews' (2008) concerns about the state of knowledge in kinesiology relevant once again. Andrews' observations, after all, pertain to the epistemological *hierarchies* that can be forged and fortified through 'progressive' forms of objective analysis. The extent to which explicitly subjective insights not only remain, but remain *salient* in relation to 'powerful' numbers is a matter requiring further attention.

Conclusion: Towards an Empirically- and Theoretically-Informed Sociology of Sport and Big Data

What this paper has ventured to show is that the burgeoning pursuit of advanced statistical analyses in sport reflects and reinscribes the general pursuit of such activity in the wider conjuncture. With this reciprocal relationship between sport and Big Data, new realities are evidently arising for ownership, management, athletes, and consumers alike. Of course, amassing and evaluating large data troves has been done before, even in sport. Henry Chadwick once devised the baseball box score with the general spirit of measuring labour in mind; Frederick Taylor returned the favour by drawing an analogy between the shop floor and the baseball diamond. Yet with the advent of technologies like tracking cameras and Hadoop, and with the steady 'flow' of analysts, investors, and ideas across the socio-political-economic landscape, the volume and variety of data only grows, as does their speed of movement. In this context, athletes and fitness enthusiasts are newly evaluated while fans become both 'fanalysts' (fan-analysts) and the targets of corporate surveillance in novel ways. Amidst all of this, 'progress' is still commonly deemed both the propelling force and manifest outcome of number 'crunching'. Now, however, progress means profitability at the margins and 'better' knowledge as much as ensuring the population's biological survival. It evidently means progressing towards more invasive forms of data collection as well.

We follow boyd and Crawford (2012) in emphasizing the need for further research on Big Data in general. "The era of Big Data has only just begun," they say, "but it is already important that we start questioning the assumptions, values, and biases of this new wave of research" (p. 675). Yet here we conclude by re-asserting the need for a robust research agenda on *sport* and Big Data in particular, especially as the tensions between quantitative and qualitative epistemologies continue to be negotiated in sport and beyond. The Moneyball phenomenon in baseball, for example, which we have characterised as closely linked to Big Data's arrival, has been left mostly for disciplines like sport management and economics to scrutinise. There remains a need for scholars in fields like (physical) cultural studies and sociology to fully document and weigh the implications of the Moneyball ethos in particular and sport's affiliation with Big Data in general.

This research in one sense might take a critical approach. As Rose (1999) writes, "numbers determine who holds power, and whose claim to power is justified" (p. 197). Sport's adoption of new forms of quantitative analysis clearly has power-effects. Future work might thus examine in detail the specific surveillance modalities or amalgamated 'surveillant assemblages' (Haggerty & Ericson, 2000) that now govern the activities of fans and players alike, or the ways that progress is potentially undermined by (for example) the proprietary holding of analytic technologies and formulas. But in scrutinizing sport's "new technologies for classifying and enumerating" and "new bureaucracies" for deploying the technology (Hacking, 1990, p. 3), Foucault's (1980) reminder that power is a productive, at times pleasurable force should be heeded as well. It has been said, for example, that the strongest relations of surveillance are those that have been entered into wilfully (cf., Andrejevic, 2004); there are sure to be 'benefits' to fans in (for example) receiving targeted marketing in sport stadiums or devoting unremunerated labour to the analytics movement. Future research should thus be inclined towards further *mapping* the sport/Big Data landscape as much as it is towards critiquing it.

We are reluctant to provide methodological prescriptions to scholars who have to a great extent embraced 'methodological contingency' as a way of accounting for the manifest complexities of lived experience. That said, apart from flexibility in the use of research methods, in the spirit of Science and Technology Studies (STS) we would argue for research into the role of non-human actants – together with humans – in the arrival of Big Data in sport. Among Latour's contributions to STS is the view that social relations are better conceived as socio-technical *networks* – a point that nicely aligns with our own assessment of the 'flows' related to sport and Big Data in Postulate 1. Non-humans are actively situated across these networks, and as such are a core component of the 'conditions of possibility' for human activity – whether in sport or beyond. A full review of Latour's methodological exhortations is not possible in this space (see Latour, 2005), but a key imperative involves the 'tracing' of networks, in part with a view towards "how objects 'mediate' and facilitate various courses of action" (Kemple & Mawani, 2009, p. 243). From this perspective, then, the (non-human) cameras designed for recording every player and fan movement are as worthy of attention as the experiences of players and fans themselves.

Taken together, a sociology of sport and Big Data need be both theoretically- and empirically-informed. It is only fitting that the 'datafication of everything' be met with data collection and analysis on this trend in itself.

References

Anderson, C. (2008). 'The end of theory: The data deluge makes the scientific method obsolete'. *Wired*. Retrieved June 22, 2013 from http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

Anderson, M.G. (1988). *The American census: A social history*. New Haven: Yale University Press.

Andrejevic, M. (2004). *Reality TV: The work of being watched*. Lanham: Rowman & Littlefield Publishers.

Andrews, D. (2008). Kinesiology's inconvenient truth and the Physical Cultural Studies imperative. *Quest*, *60*(1), 45-62.

Andrews, D., Silk, M., Francombe, J. & Bush, A. (2013). McKinesiology. *Review of Education, Pedagogy, and Cultural Studies, 35*(5), 335-356.

Appadurai, A. (1990). Disjuncture and difference in the global cultural economy. *Theory, Culture & Society, 7*(2), 295-310.

Arnold, A. (2013). 'Big Data: For profit or for good?' *Huff Post*. Retrieved July 5, 2013 from http://www.huffingtonpost.com/alicia-arnold/big-data-for-profit-or-fo_b_2641367.html.

Bale, J. (1993). *Sport, space and the city*. London: Routledge.

Beck, H. (2013). 'Advanced statistics added to revamped NBA site'. *New York TImes*. Retrieved June 29, 2013, from http://www.nytimes.com/2013/02/15/sports/basketball/nbas-site-to-feature-updated-statistics-database.html?_r=2&.

Birnbaum, P. (2013). 'A guide to sabermetric research: The basics'. Retrieved June 29, 2013, from http://sabr.org/sabermetrics/the-basics.

boyd, d., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, & Society, 15*(5), 662-679.

Burgelman, L. (2013). 'Attention, Big Data enthusiasts: Here's what you shouldn't ignore'. *Wired*. Retrieved June 29, 2013, from http://www.wired.com/insights/2013/02/attention-big-data-enthusiasts-heres-what-you-shouldnt-ignore/.

Cameron, D. (2012). 'The sacrifice bunt: The real rally killer'. *FanGraphs*. Retrieved July 2, 2013 from http://www.fangraphs.com/blogs/the-sacrifice-bunt-the-real-rally-killer/.

Castells, M. (1996). *The rise of the network society*. Oxford: Blackwell.

Cisco. (2013). 'Cisco introduces StadiumVision Mobile; Ground-breaking delivery of live video to mobile devices in sports & entertainment venues around the world'. *The Network: Cisco's Technology News Site*. Retrieved June 29, 2013, from http://newsroom.cisco.com/release/1134249.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, 31(5), 498-512.

Deitch, J.W. (2001). *A nation of inventors*. Carlisle: Discovery Enterprises, Ltd.

Deleuze, G. (1995). *Negotiations, 1972-1990*. New York: Columbia Press.

Dorsey, B. (2013). Sports teams grow up with Big Data and predictive analytics. *SEAT*, 76–78. Retrieved June 29, 2013, from http://www.wealthengine.com/sites/default/files/ckfinder/files/JamesDeanInterviewSEATSP13.pdf.

Eggers, B. (2012). 'The most important offseason acquisition for the San Francisco Giants could be Hadoop'. *AOL Tech*. Retrieved June 18, 2013 from http://techcrunch.com/2012/11/16/baseball-and-big-data-how-hadoop-and-the-san-francisco-giants-are-taking-on-stats/.

Foucault, M. (1980). *Power/knowledge: Selected interviews & other writings 1972-1977*. New York: Pantheon Books.

Furnas, A., & Lezra, G. (2012). 'Make way for the soccer geeks'. *The Atlantic*. Retrieved June 29, 2013, from http://www.theatlantic.com/technology/archive/2012/08/make-way-for-the-soccer-geeks/261634/

Gitelman, L. & Jackson, V. (2013). 'Introduction'. In L. Gitelman (Ed.), *'Raw data' is an oxymoron* (pp. 1-14). Cambridge: MIT Press.

Giulianotti, R. (2005). *Sport: A critical sociology*. Cambridge: Polity Press.

Giulianotti, R. & Klasuer, F. (2011). Introduction: Security and surveillance at sport mega events. *Urban Studies, 48*(15), 3157-3168.

Gould, S.J. (1996). *The mismeasure of man*. New York: W.W. Norton & Company.

Grint, K. (2005). *The sociology of work: Introduction*. Cambridge: Polity Press.

Guttmann, A. (2007). *Sports: The first five millennia*. Amherst: University of Massachusetts Press.

Hacking, I. (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Haggerty, K.D., & Ericson, R.V. (2000). The surveillant assemblage. *British Journal of Sociology, 51*(4), 605-622.

Hakes, J.K. & Sauer, R.D. (2006). An economic evaluation of the *Moneyball* hypothesis. *Journal of Economic Perspectives*, *20*(3), 173-186.

Hsu, J. (2012). '“Smart” stadium goes high-tech to serve fans'. *NBC Today*. Retrieved June 29, 2013, from http://www.today.com/id/46559840/ns/today-today_tech/t/smart-stadium-goes-high-tech-serve-visiting-fans/#.Uc5JVT7wKVd.

Hunt, A. & Wickman, G. (1994). *Foucault and law: Towards a sociology of law as governance*. London: Pluto Press.

IBM. (n.d.a). 'What is big data?'. Retrieved from http://www-01.ibm.com/software/data/bigdata/.

IBM. (n.d.b). 'Data is a game changer'. Retrieved from http://www-03.ibm.com/innovation/us/sports/index.html.

IBM. (2013). 'IBM predictive analytics reduces player injury and optimises team performance for NSW Waratahs rugby team'. Retrieved from http://www-03.ibm.com/press/au/en/pressrelease/42613.wss.

Ingham, A.G. & Donnelly, P. (1990). Whose knowledge counts? The production of knowledge and issues of application in the sociology of sport. *Sociology of Sport Journal, 7*(1), 58-65.

Kemple, T. & Mawani, R. (2009). The sociological imagination and its imperial shadows. *Theory, Culture & Society*, *26*(7-8), 228-249.

Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network Theory*. Oxford: Oxford Press.

Leiss, W., Kline, W., Jhally, S., & Botterill, J. (2005). *Social communication in advertising. Consumption in the mediated marketplace. Third edition*. New York: Routledge.

Lewis, M. (2004). *Moneyball: The art of winning an unfair game*. New York: W.W. Norton & Company.

Lewis, M. (2009). The no-stats all-star. *The New York Times*. Retrieved June 4, 2013 from http://www.nytimes.com/2009/02/15/magazine/15Battier-t.html?pagewanted=all&_r=0.

Lohr, S. (2012). 'How Big Data became so big'. *The New York Times*. Retrieved June 23, 2013 from http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html.

Lowe, Z. (2013). 'Light, cameras, revolution'. *Grantland*. Retrieved July 5, 2013 from http://www.grantland.com/story/_/id/9068903/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution.

Lupton, D. (2012). M-health and health promotion: The digital cyborg and surveillance society. *Social Theory & Health*, *10*(3), 229–244.

Mayer-Schonberger, V. & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.

Metz, C. (2011). 'How Yahoo spawned Hadoop, the future of Big Data'. *Wired*. Retrieved June 29, 2013, from http://www.wired.com/wiredenterprise/2011/10/how-yahoo-spawned-hadoop/

Miller, P., & Rose, N. (2008). G*overning the present: Administering economic, social, and personal life*. Cambridge: Polity.

Morey, D., & Hinkie, S. (2011). 'Moneyball and the Houston Rockets: The Rockets front office on the stats movement and how Michael Lewis' book impacted the way they've built a franchise'. *Grantland*. Retrieved June 29, 2013, from http://www.grantland.com/story/_/id/7001767/moneyball-houston-rockets.

Ohlhorst, F.J. (2012). *Big Data analytics: Turning Big Data into big money*. Hoboken: John Wiley & Sons, Inc.

Ohm, P. (2012). 'Don't build a database of ruin'. *Hardvard Business Review*. Retrieved June 29, 2013, from http://blogs.hbr.org/cs/2012/08/dont_build_a_database_of_ruin.html.

Perry, D. (2006). 'Extra innings. Can stats and scouts get along?' In J. Keri (Ed.), *Baseball between the numbers: Why everything you know about the game is wrong* (pp. 369-374). New York: Basic Books.

Phillips, C.J, Greene, J.A., & Podolsky, S.H. (2012). Moneyball and medicine. *New England Journal of Medicine, 367*, 1581-1583.

Porter, T.M. (1986). *The rise of statistical thinking: 1820-1900*. Princeton: Princeton University Press.

Puerzer, R.J. (2002). From scientific baseball to sabermetrics: Professional baseball as a reflection of engineering and management in society. *NINE: A Journal of Baseball History and Culture, 11*(1), 34-48.

Rabinow, P., & Rose, N. (2006). Biopower today. *BioSocieties, 1*(2), 195-217.

Risker, C.D. (1995). Frederick Taylor's use of the baseball team metaphor: A historical perspective on scientific management and baseball. *NINE: A Journal of Baseball History and Social Policy Perspectives, 4*(1), 1–11.

Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge: Cambridge Press.

SABR. (2013). 'SABR's Founding Members: The Cooperstown 16'. *Society for American Baseball Research*. Retrieved June 29, 2013, from http://sabr.org/about/founders.

Schiff, A.J. (2008). *The father of baseball: A biography of Henry Chadwick*. North Carolina: McFarland & Company, Inc.

Shogan, D. (1999). *The making of high performance athletes: Discipline, diversity, and ethics*. Toronto: University of Toronto Press.

Silk, M. (2004). A tale of two cities: The social production of sterile sporting space. *Journal of Sport & Social Issues, 28*(4), 349-378.

Silver, N. (2012). *The signal and the noise. The art and science of prediction*. London: Penguin.

Sportvision. (2013). 'PITCHf/x'. *Sportvision, Inc.* Retrieved June 29, 2013, from http://www.sportvision.com/baseball/pitchfx.

Sugden, J. (2012). Watched by the Games: Surveillance and security at the Olympics. *International Review for the Sociology of Sport, 47*(3), 414-429.

Swan, M. (2013). The quantified self: Fundamental disruption in Big Data science and biological discovery. *Big Data*, *1(2)*, 85-99.

Thurm, S. & Kane, Y.I. (2010). 'Your apps are watching you'. Retrieved June 2012 from http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html.

Urry, J. (2000). 'Mobile sociology'. *The British Journal of Sociology, 51*(1), 185–203.

Vertinsky, P. (2003). Embodying normalcy: Anthropometry and the long arm of William Sheldon's somatotyping project. *Journal of Sport History, 29*(1), 95-133.

Vertinsky, P. (2009). Mind the gap (or mending it): Qualitative research and interdisciplinarity in kinesiology. *Quest*, *61*(1), 39-51

Wolfe, R., Babiak, K., Cameron, K., Quinn, R.E., Smart, D.L., Terborg, J.R. & Wright, P.M. (2007). Moneyball: A business perspective. *International Journal of Sport Finance, 2*(4), 249-262.

Žižek, Slavoj. 1999. *The Ticklish Subject*. London: Verso.