Wuz You Robbed? Concerns With Using Big Data Analytics in Sports

Dov Greenbaum, Radzyner Law School and Yale University

Human enhancement in sports triggers a near-universal and visceral negative reaction. Loland (2018) posits that this reaction relates to an age-old distinction that athletics makes between the nontrivially defined "natural" and the "unnatural." Perhaps, however, Joe Jacobs best articulated a rawer reason for this pervasive *weltanschung*: Jacobs, when his boxer, Max Schemrling lost to Jack Sharkey in a highly contested fight, famously exclaimed, "We wuz robbed!" (Thomas 1994). A statement that still echoes as truthful today.

We argue that this manifest injustice in sports—unfair human enhancement—doesn't stem simply from the other side somehow gaining an unfair advantage, but rather, it arises specifically when that unfair advantage, when employed during actual gameplay, causes the one side that ostensibly played (more) fairly to lose what should have been theirs to win. It is telling that the oft-used phrase regarding fairness in sport is "a level playing field"; as long as the actual game seems fair, other unfairness off the field might be allowable.

To wit: When it is simply training tools that result in training benefits, they tend to be allowed, as described by Loland elsewhere (Loland 2009). This can be seen most cogently in the continued tolerance of expensive apparatuses such as cryosaunas (Lombardi 2017) or hypoxic/altitude tents (Staff 2006). Doping, however, with the drug erythropoietin (EPO), with its short half-life, provides the same result, but is problematic: It is injected near the time of competition (like many other drugs favored by dopers) (Gilford 2016), and it or its immediate effects remain in the bloodstream during play (Audran 1999).

Further, allowable unfair training includes those instances where financial wherewithal and independence to train afford only a subset of players the best trainers, training tools, and training opportunities. However, only when that financial wherewithal translates into actual differences on the field, that is, during the playing of the sport itself, will we find organized sports stepping in and disallowing that unfairness. For example, this was the case with the relatively unaffordable and unavailable Speedo LZR bathing suits that would only be accessible to a subset of competing athletes (Zettler 2009).

Notwithstanding the preceding distinction, there remain a few unfair and unnatural advantages that are still allowed even during gameplay. One of these, the use of big data and predictive analytics, both in real time during the game and during training, is trending toward even greater inclusion in sports (Link 2018). And while their primary purpose can be seen to enhance training, their application during actual competitive play can provide superhuman analyses and feedback through the use of expensive proprietary technologies, and/or as interpreted by a small cadre of professionals. Imagine, a baseball catcher being fed very specific information about the batter—in a tie game, facing a full count (three balls, two strikes), at an away game, on a chilly day, in the ninth inning, with one runner on base, there is a 93.58% likelihood that he will swing on the next pitch—and providing that information to the pitcher, rather than the catcher simply employing intuition to signal to the pitcher to throw high and away. In short, data analytics could be seen as a clear competitive advantage, allowing those players and teams that can afford it to play better and smarter.

Why then is it still allowed? Unlike the Speedo LZR suit, and similar technologies, data analytics would arguably still not violate the spirit of the sport. Modern competitive sport is effectively the collection and analysis of data—the counting of goals, fouls, and baskets—and the comparison of one data set against another—one team against its rivals. Moreover, more so than most other human recreational endeavors, sports is an excessively statistical game, where every relevant number has always been tabulated and analyzed by both rabid fans and professional teams. To disallow analytics would be unthinkable.

Nevertheless, while data analytics and big data may not be as problematic for the organizers of competitive sports as the enhancements described by Loland, they raise a number of other often-unappreciated ethical concerns.

Professional sports teams often invest millions annually into dedicated analytics departments that employ advanced predictive analytics and other areas of artificial intelligence in seeking to squeeze out a competitive advantage from every data point. However, in collecting data points from every possible source—from sports specific statistics, to general physiological data, to even genetic information (Sela 2016)—there is a real concern that the data collectors will stumble across incidental findings, for example, actionable data with health-related relevance to either the athlete themselves or their extended family, such as hypertrophic cardiomyopathy or a predisposition

Address correspondence to Dov Greenbaum, Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Ave, Bass 432 New Haven, CT 06520, USA. E-mail: dov.greenbaum@yale.edu

to Alzheimer's disease. Further, even when these data are only correlative rather than causative, they will be perceived as predictive by the athlete, the athlete's team, and/or the relevant management, and they could be used, with or without justification, to assume a particular fate for the athlete or a family member, perhaps ending a career prematurely, or creating problems for insurance and employment (Greenbaum 2014).

The concerns with incidental findings, and their externalities, are further exacerbated by the fact that those who are collecting and perhaps disclosing data to the athletes are likely not health professionals (e.g., coaches or inquisitive fans), and may not necessarily be adequately trained to divulge such information.

Another concern relating to the collection of an athlete's information relates to privacy for the athlete and their extended families that may share many of the relevant physiological and genetic data that were collected. Further, as more and more data are collected, they can be correlated with publicly available (nonprivate) data regarding the athlete. The power of analytics is such that even these nonprivate data can be used to further infer private information even when there was no consent to disclose such information. Moreover, as teams and athletes themselves push to collect this information, fans and others will similarly seek out this information, either through overt or even surreptitious video or through the collection of DNA from simply a discarded water bottle or a sweaty glove (Greenbaum 2013).

The data necessary for big data analytics are collected from various varied sources, including the athlete, publicly available information, and the relevant sports organizations. Much of big data in sports is made possible by wearables that include headsets and computer interfaces, sensors embedded in the field, in the sporting equipment and in the clothing, wristbands or smart watches worn by the athletes, and cameras positioned to capture everything. Is the data traffic between devices encrypted to protect the athlete's information? If that information is hijacked, can it be used to unscrupulously gamble on the player?

Given these varied sources, various parties may vie for ownership of the inputs and resulting outputs. Can a player who was traded demand that the relevant analytics follow him or her to the next team? Can an athlete legally limit his or her opponent's access to helpful data?

Although there are many unanswered questions relating to ethical and legal concerns in the use of big data in sports, there are a number of options that can be implemented immediately to preempt many of the most serious concerns. These include an effort to educate teams, athletes, and their support staff as to the privacy and health concerns associated with big data. This education can also include the creation of clear and relevant consent forms that outline the risks and repercussions associated with collecting data. Leagues and sports organizations, without outlawing the use of this technology, can set best practices and other guidelines, not for the benefit of the sport per se, but for the benefit of the individual athletes who greatly benefit from such guidance. Above all, it is important that the players not be robbed of their future, their privacy, or any ownership that they may have in the data. ■

REFERENCES

Audran, M., et al. 1999. Effects of erythropoietin administration in training athletes and possible indirect detection in doping control. *Medicine and Science in Sports and Exercise* 31:639–45. doi:10.1097/00005768-199905000-00003.

Gilford, B. August 2016. Bill the scientific american guide to cheating in the olympics. *Scientific American*. Available at https://www.scientificamerican.com/article/the-scientific-american-guide-to-cheating-in-the-olympics/

Greenbaum, D. 2013. If you can't walk the walk, do you have to talk the talk: Ethical considerations for the emerging field of sports genomics. *The American Journal of Bioethics* 13 (10):19–21. doi:10.1080/15265161.2013.828121.

Greenbaum, D. 2014. If you don't know where you are going, you might wind up someplace else: Incidental findings in recreational personal genomics. *The American Journal of Bioethics* 14 (3):12–14. doi:10.1080/15265161.2013.879946.

Link, D. 2018. *Data analytics in professional soccer*. Wiesbaden: Springer Vieweg

Loland, S. 2009. The ethics of performance-enhancing technology in sport. *Journal of the Philosophy of Sport* 36:152–61. doi:10.1080/00948705.2009.9714754.

Loland, S. 2018. Performance-enhancing drugs, sport, and the ideal of natural athletic performance. *American Journal of Bioethics* 18 (6):8–15.

Lombardi, G., E. Ziemann, and G. Banfi. 2017. Whole-body cryotherapy in athletes: From therapy to stimulation. An updated review of the literature. *Frontiers in Physiology* 8:258. doi:10.3389/fphys.2017.00258.

Sela, A. B., et al. 2016. Sports is fleeting, but genetics is forever: Some legal and ethical implications of genetics in college sports. *Virginia Sports and Entertainment Law Journal* 16:1.

Staff. 2006. WADA holds off on ban of hypoxic chambers. *Washington Post* September 17. Available at: http://www.washingtonpost.com/wp-dyn/content/article/2006/09/16/AR2006091600685.html

Thomas, R. M., Jr. 1994. Jack Sharkey, boxing champion, dies at 91. *New York Times*, August 19, A25.

Zettler Patricia J. 2009. Is it cheating to use cheetahs?: The implications of technologically innovative prostheses for sports values and rules. *Boston University International Law Journal* 27: 367–403.