The Ethics of Sports Analytics in Track and Field

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Maryanna Lansing
Spring, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

The Ethics of Sports Analytics in Track and Field

What is Sports Analytics?

Since the events described in the book *Moneyball*, when the Oakland athletics revolutionized baseball through data analytics, sports has never been the same. The next step now in every sport's development and to get an advantage over the other teams is to partake in sports analytics. Sports analytics as defined by Benjamin Alamar and Vijay Mehrotra is "the management of structured historical data, the application of predictive analytic models that utilize that data, and the use of information systems to inform decision makers and enable them to help their organizations in gaining a competitive advantage on the field of play" (Alamar and Mehrotra, 2016, p.1). Originally this "historical data" was just the statistics from a game like points scored, assists, or steals. However, with the development of more advanced wearable sensors and more accessible exercise and physiology labs, the data has changed from just game statistics to biometric data from the athletes.

Today, runners of all levels (professional to just running for fun) can wear watches that provide them with distance, cadence, heart rate, recovery and more. Professional runners and Olympic runners have access to labs that can provide this data and much more such as VO_2 max, stride length, blood oxygen levels, lactic acid levels, etc (Tom Farrell). This data is being used to increase performance by finding an athlete's weak points. However, this information is usually provided to coaches and athletes alike and sometimes coaches, trainers, or athletes themselves can take advantage of this data and abuse it. This paper will explore the research question, *what are the ethics behind sports analytics and biometric data collection in track and field and how are these ethics established?*

Research and Methods

To explore the research question above, three STS methods will be used. First, network analysis is used to understand the relationship between coaches, athletes, trainers, wearable devices, and data. Second, firsthand personal perspectives on the subject will come from interviews from professional runners. Three professional athletes were interviewed: Tom Farrell, a member of the Oregon Track Club Elite; Chase Weaverling, a member of the Reebok training group in Charlottesville, Virginia; and Brigdet Guy, a pole vaulter and a potential Olympic athlete. Chase and Bridget were interviewed in person while Tom was interviewed over the phone. Trainers and coaches were not contacted because this paper focuses more on the athlete's perspective and the effects these new technologies have on the athletes. The questions asked in the interviews are in the appendix of this paper. Third, the last method used is historical case studies on two major events in Olympic Track and field: one, the Case of Caster Semenya and two, the 2016 Russian doping scandal. These cases will be analyzed in the order above because the technology from the Russian doping scandal builds upon the technology from the Case of Caster Semenya. Each of these events provides interesting uses of biometric data and questions the line at which the protection and privacy of the athlete is being compromised.

Sports Analytics in the Real World

In recent years, biometric data collection has increased in the sport of track and field. Originally the advice to get better at running was to run longer distances and at higher speeds, do more drills, lift weights, etc. However, now with new and developed technology like advanced treadmills and quicker lactic acid tests, runners can understand their own specific strengths and weaknesses. From this information, coaches can create specific training methods and loads for

each of their athletes. Also, athletic trainers, those who work with athletes to prevent and treat injuries, can help work on the specific weaknesses of the athletes that could cause injury. Athletic trainers can be more preventative than reactive.

However, even with these positives, there are negatives to this having all of this information on the athlete. For one, the data can be abused and risk the privacy of the athlete. In the case of Caster Semenya, which will be explained in greater detail later in this paper, having so much biometric data on the athlete backfired on her (Wells and Darnell, 2014) because it prevented her from being able to compete. Yet, this data can be ignored or changed. Mary Cain, an American athlete who runs the 1500 meters and the 800 meters, had data available to help her and prevent possible issues she dealt with. However, her coaches decided to ignore it and follow arbitrary policies that they themselves put in place (Cain, 2019). Also, in the case of the Russian Olympic athletes, biometric data can be forged in order to cheat a system that was put in place to make competition fairer.

In November 2015, Russian Olympic athletes, especially Russian track and field athletes, were charged with cheating in past international competitions through a massive doping scandal (Zorithian, 2016). The Russian track and field athletes were then banned from international competition while an investigation took place and then were banned from the 2016 Olympics in Rio. In 2016, Grigory Rodchenkov admitted to participating in a scheme issued by the Russian government to "help" Russian athletes dope and then pass drug tests (Fogel, 2017). The Oscar winning documentary from Netflix, *Icarus*, goes into detail about Grigory Rodchenkov involvement in the scheme and his escape afterwards. Russian track and field athletes are still currently banned from international competition. Some Russian athletes, who were not part of

the doping scheme, have been invited to compete under the Olympic flag but not for their country (Rachel Axon, p.1).

This past year, Caster Semenya, a South African runner, was banned from international competition. The reason beyond her banning was that her body contained too high levels of testosterone to compete as a woman in international competition. “Most women, including elite female athletes, have natural testosterone levels of 0.12 to 1.79 nanomoles per liter, the I.A.A.F said” (Longman and Macur, 2019, p.3). Caster and her lawyers, based on the opinions of certain scientists, fought that the IAAF was basing these levels on incorrect information and that there is no real evidence that higher testosterone levels gives an athlete that much of an advantage. However, in the spring of 2019, the court ruled that Caster would be banned from international competition.

Actor-Network Theory and Technological Determinism

In the field of science, technology and society (also known as STS), there is a focus on making sure that as engineers develop new technologies, they must think of the benefits and the consequences that these new technologies will have on society or the ethics behind their new developments. Therefore, the benefits and consequences of biometric data collection in the sport of track and field must be analyzed and potentially established further. To accomplish this task, this paper uses two STS frameworks: actor-network theory (ANT) and technological determinism.

ANT is the theory that explains the interactions between actors (who can be humans or non-humans), intermediaries, and the network in which they exist (Cressman, 2009). The intermediaries are connections between actors and they communicate the needs and intentions of

the other actors. The networks are the relationships between the actors. These networks can be large and inclusive or smaller parts of a larger network. However, this is a major criticism of Actor Network Theory. Critics say that you can define the network infinitely. When using this theory, a clear and defined system must be established. Another criticism that this framework faces is that the "researchers simply report what they see and intangible elements like values and norms are not recognized" (Cressman, 2009). This causes a lack of understanding of the experience of each of the actors and how that impacts the system in which they exist. Yet, this issue can be avoided by careful observation and study of the network.

This framework works well with the ethics of biometric data collection because there is a clear and defined network established. The actors in this network are the coaches, athletes, trainers, analysts, and athletes' data. Each of the actors has their own independent relationships that can overlap with the others; however, the most clear and consistent connection between all the relationships is the athlete's data. It is important to study how that data affects each actor and how each actor interacts with that data.

Technological determinism is when society is shaped by technology (Smith, 1994). There is hard determinism and soft determinism. Hard determinism is when technology develops independently from social concerns. Soft determinism is when there is a chance for humans to shape the technology but technology is still the guiding force. The issue of the ethics of biometric data collection is soft determinism. The actors listed above need new technology to update training and rehabilitation methods. However, the new technology has changed and shaped the sport in unexpected ways on its own. Where running is headed as a sport is driven by the new technological development in shoes, training styles, and more.

Small Scale Biometric Data Use vs Large Scale Biometric Data Use

Within the International Association of Athletic Federation (IAAF) and International Olympic Committee (IOC), laws need to be implemented that establish a standard for the use of biometric data from the athletes that compete under both of these groups. Most problems that athletes have faced when it comes to the exposure or misuse of their data has happened on public or large scales like the Russian Doping Incident or the case of Caster Semenya. Due to these cases, old policies must be revisited and new policies established. However, on a personal and smaller level, between athlete and coach, most athletes have a relationship with their coach that is built on trust. This trust between coaches and athletes allows for greater exploration of athletes' data so that the coach and the athlete together improve the athlete's performance. Still, because not every relationship between coach and athlete is well established, the standards established by the IAAF and the IOC need not only apply for the large competitions but also on the smaller levels to protect all parties. The exact nature of these policies is still being discovered. Since the technology is new and still developing, it is hard to apply standards across multiple situations. Athletes and the IAAF and the IOC need to work together to establish the best policies that both keep the sport intact but also protect the athletes' interests and privacy.

In this particular case, the actors that have a main part in this system of Olympic track and field are the coaches, trainers, the data, the technology to collect the data, and the athletes. According to the article "Tracking U.S. Professional Athletes: The Ethics of Biometric Technologies," "With athlete contracts worth tens of millions of dollars and lucrative championships at stake, professional sport has long strived to optimize the physical performance

of its employed athletes, who must perform under conditions of stress, including fatigue, overexertion, overtraining, and sleep deprivation" (Karkazis and Fishman, 45). Though runners' contracts are not usually worth tens of millions, there is still a high demand for them to perform well for their country in international competition. Coaches are supposed to train their athletes to succeed when needed, while athletes are trying to be as successful for as long as possible, staying healthy throughout their career. Trainers are supposed to keep their athletes healthy so that they can perform and train. The technology helps collect the biometric data from these athletes to help each of the human actors succeed in their roles and goals and helps to analyze such data. What if one of the actors uses the data improperly? According to Kathryn Henne in "I Feel Like a Lab Rat," "The ability to track, document, and share data on various aspects of athletes' biological and physical characteristics becomes particularly problematic in practice: Not only are there significant power differentials in U.S. professional sport that often disadvantage athletes, there is also an absence of regulation aimed at protecting athletes' data and its use" (Henne, 2017, p.63). The athlete is not always protected from the use of their data and it can work against them. This lack of protection is significant because it shows a breakdown of the network. If an athlete's data is working against them, they will be less likely to share their data with the other human actors (coaches, trainers, analysts) in the network. Without access to the data, coaches, trainers, and analysts can do or improve on their jobs.

Athletes have varying interests in the collection of their data. Some are afraid of the data working against them but even more than that it can cause mental health issues for some. It can be too much information. The athlete could see a dip in performance over a couple days that could be due to a number of factors, but just looking at the prediction or the value, it can psych the athlete out. The athlete can respond negatively and hurt their performance more.

Furthermore, like Chase Weaverling, a professional long-distance runner in the Charlottesville area, said, “The cool thing about running is that it's an unpredictable sport. My training leading up to my race in Atlanta would have shown that I wouldn't perform well. Yet, on the day I ran a good race. I kept 4:50 pace for 13 miles.” However, those who are interested have problems that could be solved through data analytics.

The athletes in the Oregon Track Club Elite have multiple tests run on them to better understand themselves as runners and to solve problems they face in their careers. One problem solved is immediate in workout feedback. The coaches take a simple blood sample to test the lactic acid levels of the runners in a workout. This helps both the coaches and the athletes, as it allows the coaches to modify the workout if an athlete is working too hard or not hard enough for what is needed that day and it allows athletes to know when they can work harder or need to better rest (Farrell). The club also works with centers to analyze their gait to see what needs strengthening and what type of shoes they should be wearing. They also measure oxygen use and take blood tests to see levels of vitamin D, iron, hydration, and more are in a good range (Farell).

The analytics that Tom Farell’s club participates in are the types of analytics that Chase would like done. He would like to know how well he is recovering and whether or not he is overtraining (Weaverling). That way he can know where the line between just right and overdoing it is before reaching it. However, not all coaches are interested in this type of analysis and provide their athletes the opportunities for this type of analysis. Chase’s last two coaches are not very scientific. They believe the best way to get better at running is to focus just on running and just run (Weaverling). This disagreement about the usefulness of data causes a tension between two parts of the network: athletes and coaches. Athletes, like Chase, want to and do trust their coaches but they also want and soon, as technology progresses, will need to partake in

biometric data research to gain a competitive edge. However, if their coaches will not use data to improve/adjust training methods, then an athlete is likely to explore how data can help them and will lose a competitive advantage.

Bridget Guy, a professional pole vaulter, shares a different perspective. Both Tom and Chase are long distance runners; for them internal data matters the most in improving their performance and recovery. However, for Bridget Guy, it goes beyond just the internal biometric data. She did perform a gait analysis and speed analysis with the USA team in order to better understand how she is gaining speed as she is coming down the runway (Guy). After she runs down the runway she must worry about angles and physics as she is attempting to thrust herself through the air on a pole. She brought attention to the fact that athletes have more data besides just biometric data that is being used.

These professional runners have a good relationship with their coaches. They trust that their coaches do or would use their data properly and to benefit the athlete. Bridget said, "I consider Coach Wilson (Mario Wilson, the jumps coach at the University of Virginia and Bridget's coach) both a mentor and a friend. I am lucky that he has been there for me both inside and outside of track." Both Bridget's coach and Chase's coach do not participate in data collection; however, both athletes, if given the opportunity, would like to participate in data collection and data analysis. Still, they trust their coaches train mainly without extra knowledge from data.

Still, the technology has been developing within the running world and it is not slowing down. For example, Yuanlong Liu and Robert W. Schutz, researchers at the University of British Columbia, developed mathematically based models in 1998 to predict future world-best performances in men's track and field (Liu and Schutz, 1998). Using world record performances

plus the best performances from 1990-1992, through least square estimation, random sampling, and Monte Carlo simulation, Liu and Schutz developed an exponential model for men's 1500 meters with time as the dependent variable. This model not only predicted future world records but also ultimate performances in the event (Liu and Schutz, 1998). This model is not based on biometric data, but with new technology, now it is possible to create a similar type of model for an individual athlete predicting his/her best performance.

The legality of this issue has been overlooked as the technology is developing faster than consequences can be analyzed. According to the article "The Legality of Biometric Screening of Professional Athletes", "Several aspiring professional athletes have seen their careers cut short by biometric screening" (Roberts, Cohen, Deubert, and Lynch, 2017, p.65). One of those athletes is Caster Semenya. Her career did not end because of prediction models but rather because with the ability to better track testosterone better, the IAAF has set certain standards for how much testosterone females can have in order to compete as a female. These standards are to establish what is considered "female" in fair competition, but females can have higher than normal testosterone and still be female. Caster Semenya was banned from international competition because her levels were too high, deeming it unfair for her to compete in female competition. "In a 2-to-1 decision, the court ruled that restrictions on permitted levels of naturally occurring testosterone were discriminatory but that such discrimination was a "necessary, reasonable and proportionate means" of achieving track and field's goal of preserving the integrity of female competition" (Longman and Macur, 2019, p.2). Therefore, the IAAF's policies are discriminatory, but are deemed reasonable in the goal of fair competition.

The technology of monitoring this level is trying to create fair competition; however, it is also defining who is what gender in very numerical terms that may not be fair to achieve.

According to the article, “Out of Bounds? A Critique of the New Policies on Hyperandrogenism in Elite Female Athletes”, “In practice, the policies do not concern all androgens, but focus specifically on testosterone. As such, women with naturally high endogenous levels of testosterone, primarily though not exclusively women with intersex traits, or what are also called disorders of sex development (DSD), are presumed to have an advantage over women with lower levels of testosterone” (Karkazis, Jordan-Young, Davis, and Camporesi, 2012, p.3). If a female has DSD or just higher levels of testosterone, it may not be their fault they can meet the standards set for this hormone. Her higher levels of testosterone could just be naturally produced by her body.

However, for Caster, this was a major invasion of privacy. “Semenya was reportedly subjected to a two-hour examination during which doctors put her legs in stirrups and photographed her genitalia” (Karkazis, Jordan-Young, Davis, and Camporesi, 2012, p.4). This case shows a very dark side of data use and even more than that how far behind in policies for data use in athletics is. The International Associations of Athletics Federations (IAAF) used old methods and were very public with all information they collected on Caster Semenya causing her to require trauma counseling (Karkazis, Jordan-Young, Davis, and Camporesi, 2012). However, because it was so publicly handled and the IAAF was criticized greatly for how it handled the situation, the IAAF did revisit their policies and established new ones (Wells and Darnell, 2014). Track and field policies are trying and starting to catch up with the technology.

A similar case of trying to create fair competition, but on a much larger scale, is the Russian doping scandal. Since Russia showed that there is a clear and manageable way of tricking drug tests, scientists have been trying to develop new markers to tell if an athlete is doping. However, these markers are based on the suspicion that athletes are taking drugs, not

hard proof. “This is an unfair position for athletes, whose reputation is significant for deriving their living. They must defend themselves against shadows of suspicion without any real tools at their disposal” (Gleaves, 2017). Again, trying to create fair competition could actually lead to harmful consequences for the athletes. Rather than taking the time to better develop the technology before putting it to use, the technology is ruling the decisions and society is jumping to conclusions trying to benefit the sport but hurting the athletes.

The Case of Caster Semenya and the Russian doping scandal both show that developing technology is changing and defining the sport. In other words, technological determinism is driving the sports industry. New technology that can better measure hormone levels in athletes stopped Caster Semenya’s career. However, alongside the technology that stopped Caster is the technology that can cheat the system. The Russian doping scandal showed that technology can be used to cheat the system and the sport. All of this technology is forcing track and field to change. More rules must be put in place to control the use of new technology and to create fair competition. Track and field must use new technology to improve as a sport but also stay true to the sport is and has always been.

There were some limitations in researching this topic. The topic of biometric data use and sports analytics is a fairly new topic so most research has been done either on other sports who have more established data analytics or just on the data itself. Many coaches, trainers, or athletes have not been studied or asked about the effects of biometric data collection on them. Also, track and field is a sport that is set in the mindset that technology is going to help performance. Most coaches are still in the mindset that to run faster an athlete needs to run more and harder, to jump higher, lift more weights and jump more, etc. This means that athletes, coaches, etc. do not have

a lot of experience working with data and do not understand both the benefits and disadvantages that come with data use.

Further study on this subject would include interviewing athletes from different and more event groups (sprinters, long jumpers, high jumpers, and throws). Also, an interesting experience would be to work with one athlete, collecting and modeling his/her data, and seeing the effect it has on the athlete, the coach, and others who the athlete works with. This would allow the research to see where the data helps, where the data psyches out the athlete, and where the data starts to incriminate the athlete's privacy.

The Standard of Biometric Data Use

What are the ethics behind sports analytics and biometric data collection in track and field and how are these ethics established? It seems that overall, most athletes have worked on and developed good relationships with their coaches, trainers, etc. Therefore, on a personal and small level, data collection is useful and the athletes are protected. This allows for the coaches, trainers, and the athletes to benefit from the data and use it to their advantage. Protection laws might need to be put into to make sure that no one crosses a line and the athlete is not in jeopardy. However, on a larger and more public scale, the IAAF and IOC have struggled in deciding the best course of action. This is why these two organizations must work together to establish new policies and laws to better handle situations like the Russian Doping Incident and the Caster Semenya Case. By working together, the IAAF and the IOC would not only be setting up a precedent across all sports on the use of sports analytics, but also, they would be paving the way for the development of track and field in the future through a safe environment for the athletes.

Appendix:

Interview Questions asked of the Professional Athletes:

What does your average training look like?

How is your relationship with your coach?

Do you record any biometric data, like heart rate, lactic acid, etc ? If so what?

If not, if you had the opportunity to collect such data, would you like to?

Do you/would you give your coaches access to this data and do you/would you trust that they would use it in your best interest?

What do you want your data to be used for?

Do you think an athlete's biometric data should be used to predict performance?

Specific to Tom Farrell:

What work do you do with professional runners?

What wearable devices do the professional runners you work with wear?

Besides just looking at the distance and time, do professional athletes themselves or have anyone do analysis on data collected from the wearables?

Do professional athletes take part in any studies such as sleep or gait analysis and if so can you describe some of these studies?

Have you or the professional runners worked with specific centers and if so which ones?

What type of supplements do you and the professional athletes take? Any nutritional studies?

What problems do athletes face when working with their data?

What current challenges are athletes facing that analytics could help?

What types of data are used?

Works Cited

Alamar, B., & Mehrotra, V. (2016, July 14). Beyond 'Moneyball': Rapidly evolving world of sports analytics, Part I. Retrieved January 30, 2020, from http://analytics-magazine.org/beyond-moneyball-the-rapidly-evolving-world-of-sports-analytics-part-i/

Cain, M. (2019, November 7). I Was the Fastest Girl in America, Until I Joined Nike. *New York Times*. Retrieved from https://www.nytimes.com/2019/11/07/opinion/nike-running-mary-cain.html

Cressman, D. (2009). A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation, 1–17.

Farrell, T. (2019, October). Personal interview

Gleaves, J. (2017). Biometrics and Antidoping Enforcement in Professional Sport. *American Journal of Bioethics*, *17*(1), 77–79. https://doi.org/10.1080/15265161.2016.1251644

Guy, B. (2020, February). Personal interview

Hayhurst, C. (2019, August). Data Analytics Helps College Coaches and Athletes Optimize Training and Performance. *EdTech Magazine.* Retrieved from: https://www.richmond.com/

Henne, K. (2017). "I Felt Like a Lab Rat": The Importance of Power and Context in Understanding Biometric Technologies. *American Journal of Bioethics*, *17*(1), 63–65. https://doi.org/10.1080/15265161.2016.1251659

Karkazis, K., & Fishman, J. R. (2017). Tracking U.S. Professional Athletes: The Ethics of Biometric Technologies. *American Journal of Bioethics*, *17*(1), 45–60. https://doi.org/10.1080/15265161.2016.1251633

Karkazis, K., Jordan-Young, R., Davis, G., & Camporesi, S. (2012). Out of Bounds? A Critique of the New Policies on Hyperandrogenism in Elite Female Athletes. *American Journal of Bioethics*, *12*(7), 3–16. https://doi.org/10.1080/15265161.2012.680533

Longman, J., & Macur, J. (2019, May 1). Caster Semenya Loses Case to Compete as a Woman in All Races. *New York Times*. Retrieved from https://www.nytimes.com/2019/05/01/sports/caster-semenya-loses.html

Rachel Axon, & USA TODAY. (n.d.). Fallout continues over Russian doping. *USA Today*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=J0E045955021918&site=ehost-live&scope=site

Roberts, J. L., Cohen, I. G., Deubert, C. R., & Lynch, H. F. (2017). The Legality of Biometric Screening of Professional Athletes. *American Journal of Bioethics*, *17*(1), 65–67. https://doi.org/10.1080/15265161.2016.1251647

Russell, J. S., & Browne, A. (2018). Performance-enhancing drugs as a collective action problem. *Journal of the Philosophy of Sport*, *45*(2), 109–127. https://doi.org/10.1080/00948705.2018.1440355

Shuster, S., & Hartog, E. (2016). Can the Games Get Clean? *TIME Magazine*, *188*(6), 63. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=117072401&site=ehost-live&scope=site

Smith, M. R. (1994). Technological Determinism in American Culture, 1–35.

Weaverling, C. (2020, February). Personal interview

Wells, C., & Darnell, S. C. (2014). Caster Semenya, Gender Verification and the Politics of

Fairness in an Online Track & Field Community. *Sociology of Sport Journal*, *31*(1), 44–65. https://doi.org/10.1123/ssj.2012-0173

Yuanlong Liu, & Schutz, R. W. (1998). Prediction Models for Track and Field Performances. *Measurement in Physical Education & Exercise Science*, *2*(4), 205. https://doi.org/10.1207/s15327841mpee0204_2

Zorthian, J. (2016). Russian Doping. *TIME Magazine*, *188*(5), 16. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=116950092&site=ehost-live&scope=site

(2017, August 4). Retrieved January 31, 2020, from https://www.netflix.com/search?q=icarus&jbv=80168079&jbp=0&jbr=0