Examining Human-Data Interaction in U.S. Collegiate Sports

Samantha Kolovson
kolovson@uw.edu
Human Centered Design & Engineering
University of Washington
Seattle, WA, USA

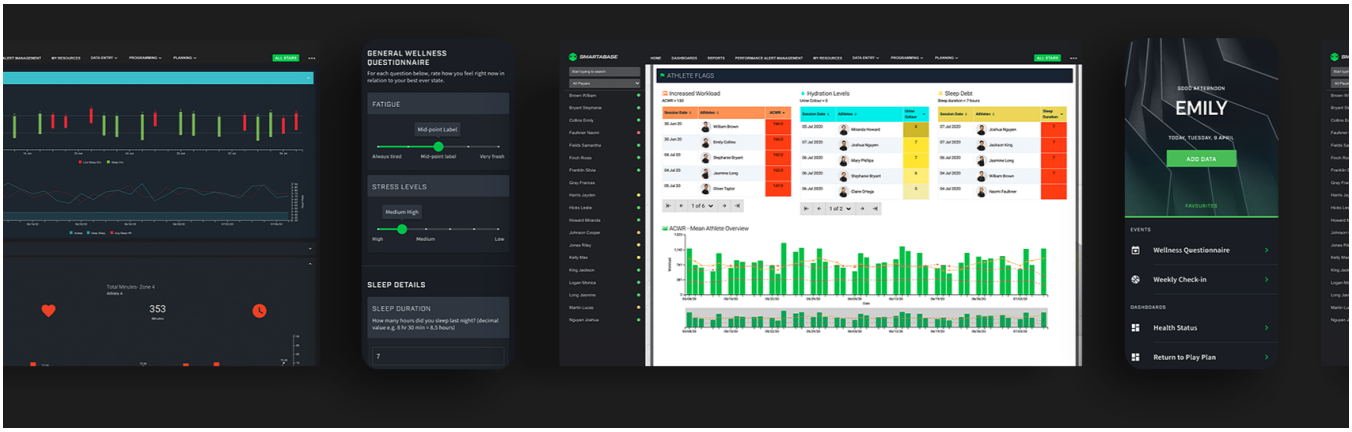

Figure 1: FusionSports' Smartabase [6] is one of the many team management systems available for collecting, analyzing and making use of athlete data.

ABSTRACT

Sports teams, particularly U.S. collegiate teams, are a compelling context for studying human-data interaction (HDI). In my previous and on-going work studying how data should be collected and used by these teams, the HDI tenets—legibility, agency, and negotiability—encompass much of what there is to be concerned about regarding HDI on sports teams. Nuanced social norms and the tensions between different roles and goals on teams involve complex interactions with data as boundary negotiating artifacts. I also suggest a shift towards being data-informed rather than data-driven and discuss themes of trust, accountability, and empowerment.

KEYWORDS

human-data interaction, sports, tracking, boundary negotiating artifacts, agency, surveillance, power

ACM Reference Format:
Samantha Kolovson. 2021. Examining Human-Data Interaction in U.S. Collegiate Sports. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

"Why didn't you get more sleep?" Kyle's coach asks, cornering him before the start of practice. Kyle sighs. What could he possibly say? He can't now go back and get more sleep and anyways he had to study for his midterm exams. Between practice, team meetings, video review, travelling for games, and classes, Kyle was barely finding time to eat never mind do his assignments and study. The only time to get any work done was after dinner, and Kyle, being more passionate about his area of study than some of his teammates, often stayed up until past midnight.

On this particular morning, Kyle's Whoop band [13] reported to his coach that he got five hours of sleep, only 70% of what he needed, and he was only 36% recovered. As Kyle is one of the team starters and the team is preparing to play in the NCAA March Madness basketball tournament, these were unacceptable numbers in both Kyle and his coach's eyes.

The rest of this scenario could go several ways: In the best case, Kyle's coach could be understanding and modify his training so that he does not get injured from lack of recovery. In the worst case, Kyle's coach could bench him for the start of the tournament or threaten his scholarship. Somewhere in the middle, Kyle and his coach could just continue with training as planned or Kyle's coach punishes him and/or the team with sprints (which is counterproductive as Kyle is fatigued).

Either way, this situation raises several questions that recall the HDI tenets: *Does Kyle have the choice to share his data with his coach? Should sleep data be collected by college athletic teams? How should sleep data be represented to athletes? To coaches? How can and should sleep data be used? Do Kyle and his coach undestand how the data is collected? How reliable are these 70% and 36% values?*

I will discuss the three tenets of Human-Data Interaction (HDI) in the context of tracking technology use by college sports teams

Conference'17, July 2017, Washington, DC, USA

in the United States. Team roles and especially power dynamics in college sports differ from professional sports in the U.S. or around the world. However, findings, thoughts, and discussion from this context could apply to other levels of sports and more widely to workplace contexts. In addition to relevant literature from HCI, sociology, and related fields, my discussion will draw on my own lens as a former student-athlete and my prior work, "Personal Data and Power Asymmetries in U.S. Collegiate Sports" [9]. The paper describes the nuances of this context in detail, the important points of which are:

- **Data:** In this context data are collected by a variety of sports tracking technologies, through athlete self-report, or staff observation. These data appear in the context as different boundary negotiating artifacts [10], but it is important to remember these data or artifacts are often are about an athlete's body.
- **Roles:** student-athletes, coaches, athletic trainers (ATs), and strength and conditioning coaches (SCCs). I often refer to coaches, ATs and SCCs together as staff.
- **Goals:** The team sports context is unique because the team is able to agree on at least one goal—winning. The other goals—improving performance, maximizing performance, preventing injury, and rehabilitating injury—are prioritized differently among the roles and can be situation-dependent. As the use of data becomes more involved in realizing these goals, data is increasingly used in negotiations between roles around goals.

2 DISCUSSING THE THREE TENETS OF HDI IN TEAM SPORTS

2.1 Legibility

The definition of this tenet holds up in a team context though it is important to understand that though different roles might have the same needs regarding understanding data collection and processing (including making algorithms transparent), they may have different needs regarding being able to view and reflect on the data. Additionally, in a team context, the data collected is about athlete, an athlete's body, or an aggregation of all the athletes on the team. So it may be "my" data, "your" data, or "our" data all at once.

Legibility is a concern for all the roles on a college team. If a student-athlete owns their own tracking device, their concerns may be similar to those identified by research in personal informatics and the quantified self movement. However, in the team context, athletes are aware that data is being collected but less aware or unaware of how it is being processed and used by their coaches and other staff members. This is due partly to not always having access to the data themselves. Furthermore, coaches and staff members described not having the tools (education, money, or time) to understand the data they are collecting about their athletes. One staff member described understanding the value that was being calculated by a system but he was unclear about how it was calculated so he was unable to use it.

There are an ever-expanding number of tools being built for collecting and analyzing data about athletes that all claim to be able to provide great insights, improve player and team performance, prevent injuries, and so on, but if athletes and coaches are struggling with legibility than they will not be able to realize those goals in practice.

As companies struggle to make their users data more legible to them, they are also inventing new measures such as "recovery score" and "strain" from Whoop or "body battery" from Garmin. These measures are calculated based on scientifically valid measures like resting heart rate (RHR), heart rate variability (HRV), and others data like time spent in different stages of sleep. However, the algorithms these companies use to combine RHR, HRV, sleep, etc. is proprietary and there is still a lot of work to be done to validate them.

This raises two other needs for the data related to legibility: accuracy and validity. Given that some data are subjective, accuracy may not always be the goal, but in team sports or any other context, stakeholders need information about the accuracy or validity of the data or algorithms processing the data. In a situation where machine learning is being used to help diagnose cancer [3] this is critical, but as athletes and coaches are increasingly relying on tracking data to make decisions ensuring accuracy and validity in the sports context could make a difference in an athletes wellbeing.

From a personal standpoint, this is the tenet I have least expertise to study and find least interesting, but at a larger societal and global level, making data legible is crucial to moving into our digital future.

2.2 Agency

College athletes in the U.S. do not have a lot of agency. They lack the players' associations that exist in the professional leagues that can negotiate on behalf of the athletes' best interests. Nevertheless, as with any team or coach-athlete relationship, the norms are such that the coach needs to monitor their athletes training to be effective and so athletes seem to opt-in by default.

Athlete monitoring for risk management supports both team and individual goals as it ideally keeps athletes healthy and results in improved performance [8, 9]. This support of athlete's goals creates an "Illusion of Voluntariness" where athletes "agree" to surveillance because they think it will benefit them in the long run.

I take the concept of an "Illusion of Voluntariness" comes from Campbell & Carson's Panopticon.com [4]. They say that people willingly engage in self-surveillance for two reasons: 1) They think they will benefit; 2) They are subtly coerced by the threat of exclusion.

I think a similar illusion exists in sports. In my work, student-athletes 1) exhibited positive reactions to performance tracking because they believed it added value to their training, and 2) "volunteer" their information willingly because they want to play their sport. In other words, there is a subtle threat of being excluded from their team.

Ethicists Karkazis and Fishman similarly describe that the pressure to perform and stay healthy profoundly influences an athlete's 'autonomous' decision to accept risks, which previously incentivised "playing through the pain," but now incentivises submitting to invasive and ongoing surveillance [8].

Since there is no opt-out, athletes exert their agency in the ways that are left to them. For example, reporting 7 hours of sleep when they got 5. However, as more automated forms of tracking increase,

athletes loose the ability to manage how they are presenting themselves. Sleep data could be automatically reported to a coach when an athlete wakes up with no input from the athlete at all.

My current and upcoming research projects will address the HDI tenet of agency, including opt-in and opt-out and how we might give athletes greater control data collection and use.

2.3 Negotiability

Nissenbaum's contextual integrity holds that norms discussed in the previous section also determine information flow and privacy [12]. Given the norms of athlete monitoring, information flows freely. As data types are changing and getting more personal, either the norms need to catch up or the design of tracking technologies needs to support more nuanced exchanges of information.

The nuances of sports team dynamics are such that it is hard to manage data in a way that supports the goals of the team and individual roles. Ackerman [1] defines this gap between the nuances of social activity and what designers are able to support technically as the social-technical gap. Ackerman specifically calls out the highly flexible, nuanced, and contextualized nature of human activity that makes up CSCW's base set of findings and that information sharing, roles, and social norms in technical systems needs to be similarly flexible, nuanced, and contextualized.

My upcoming research project will address the social technical gap by searching for design patterns that athletes and coaches find suitable for their goals and needs for privacy. Athletes might feel comfortable with data being automatically shared if it was represented to a coach in an anonymous, aggregated, or abstracted way so that coaches could use the data to inform decisions about the team but not target them individually. This could also be combined with the agency tenet so that athletes could decide when to make their data anonymous or not.

Though negotiability is aimed at allowing users to re-evaluate decisions as contexts change, this tenet also calls to mind the ability to use data to negotiate in interpersonal contexts. Decisions about when to share data and using data to negotiate could go hand-in-hand. For example, a student-athlete could choose to hide their data if it is "bad" on game day because they want to play but in another situation they could choose to share their "bad" data to convince their coach that they or the team needs rest.

3 TEAM SPORTS RAISES ADDITIONAL HDI THEMES

3.1 Data-driven versus Data-informed

As evident across all three tenets, data is being used increasingly to drive decision making in sports teams. Though data analytics are not new to sports, the data are no longer only publicly available numbers quantifying an athlete's performance but are about an athlete's body. Thus the data may be more subjective and there is potential for harming an athlete's wellbeing. Furthermore, athletes have an expertise about their bodies and coaches have an expertise about their sport that the data cannot replace. I along with Gamble et al. suggest re-framing such that we shift from being data-driven to data-informed [7]. If employed well, data could complement both athlete and coach expertise.

3.2 Trust

Trust between you and a data collector is important but currently not possible in most contexts. In sports teams, the trust between an athlete and a coach is a major factor in data collection and use. With less trust comes an environment with toxic uses of data [5]. And like other contexts, there is also trust of the company providing the tracking device. University athletic departments sign contracts with companies like Adidas and Nike to supply gear. How should these contracts be drawn up for tracking technology in terms of sharing the athlete's data with the company?

3.3 Accountability

Data are often used for accountability in a team sports context. Data could also be used similarly in future workplace contexts. What will be key is finding the level of detail needed to share the data to achieve the accountability requirement. For example, is it sufficient for the system to report that an athlete has completed a workout or slept well, or are more fine-grained details needed.

3.4 Empowerment

Like Mortier et al. [11] I think users of a system need to be empowered through legibility, agency, and negotiation. Currently, in both a team sports context and beyond, users may be employing strategies of resistance such as Obfuscation [2]. However, if student-athletes were empowered with the tools to understand and use their data and decide how and when it is shared, they might learn about their own behaviors, understand how their decisions affect their performance, and be empowered to decide when to act in ways that fit their athletic, academic, or social goals.

REFERENCES

[1] Mark S Ackerman. 2000. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human–Computer Interaction* 15, 2-3 (Sept. 2000), 179–203.

[2] Finn Brunton and Helen Nissenbaum. 2015. *Obfuscation: A User's Guide for Privacy and Protest.* The MIT Press.

[3] C J Cai, S Winter, D Steiner, L Wilcox, and others. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. In *Proc. ACM Human Computer Interaction 3, CSCW.*

[4] John Edward Campbell and Matt Carlson. 2002. Panopticon.com: Online Surveillance and the Commodification of Privacy. *J. Broadcast. Electron. Media* 46, 4 (2002), 586–606.

[5] Jori Epstein and Daniel Libit. 2020. Texas Tech women's basketball players describe toxic culture: 'Fear, anxiety and depression. *USA Today Sports* (Aug 2020). https://www.usatoday.com/in-depth/sports/ncaaw/big12/2020/08/05/marlene-stollings-texas-tech-program-culture-abuse-players-say/5553370002/

[6] FusionSport. 2021. Smartabase—FusionSport. https://www.fusionsport.com/smartabase/

[7] Paul Gamble, Lionel Chia, and Sian Allen. 2020. The illogic of being data-driven: reasserting control and restoring balance in our relationship with data and technology in football. *Science and Medicine in Football* 4, 4 (Oct. 2020), 338–341.

[8] Katrina Karkazis and Jennifer R Fishman. 2017. Tracking U.S. Professional Athletes: The Ethics of Biometric Technologies. *Am. J. Bioeth.* 17, 1 (2017), 45–60.

[9] Samantha Kolovson, Calvin Liang, Sean A Munson, and Kate Starbird. 2020. Personal Data and Power Asymmetries in US Collegiate Sports Teams. *Proceedings of the ACM on Human-Computer Interaction* 4, GROUP (2020), 1–27.

[10] Charlotte P Lee. 2007. Boundary negotiating artifacts: Unbinding the routine of boundary objects and embracing chaos in collaborative work. *Comput. Support. Coop. Work* 16, 3 (2007), 307–339.

[11] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2014. Human-Data Interaction: The Human Face of the Data-Driven Society. (Oct. 2014).

[12] H Nissenbaum. 2004. Privacy as contextual integrity. *Wash Law Rev.* (2004).

[13] Whoop. 2021. Whoop—The World's Most Powerful Fitness Membership. http://whoop.com/