

Changing the Game: Applications of Digital Technologies and Their Impact in the Sports Industry

DISSERTATION
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Management

submitted by

Arne Grüttner

from

Germany

Approved on the application of

Prof. Dr. Andrea Back

and

Prof. Dr. Jan Marco Leimeister

Dissertation no. 5046

Difo-Druck GmbH, Untersiemaun 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, October 23, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

*Today I will do what others won't,
so tomorrow I can do what others can't.*

Table of Contents

Summary	v
Zusammenfassung [German].....	vi
Introduction	1
Publication I.....	10
What We Know and What We Do Not Know About Digital Technologies in the Sports Industry	
Publication II.....	29
The New Window to Athlete's Soul – What Social Media Tells Us About Athletes' Performances	
Publication III.....	51
From Data to Dollar – Using the Wisdom of an Online Tipster Community to Improve Sports Betting Returns	
Publication IV	84
Going Global: Enhancing International Social Media Fan Engagement – Evidence from the German Bundesliga	
Publication V [German].....	113
Konfrontiert mit einer Flut von Kooperationsanfragen: So meistern Sportorganisationen die systematische Analyse und Bearbeitung	
Curriculum Vitae.....	129

Summary

Advances in digital technologies, such as cloud computing, electronic platforms, and artificial intelligence, enable the improvement of existing as well as the innovation of entirely new products and services in the sports industry. For example, over-the-top media services such as DAZN or Amazon's sports streaming initiative are digital technologies that promote a massive shift in the viewing habits of fans – away from traditional television broadcasting and towards tailored live and on-demand streaming sports consumption. However, while the digitalization of the sports industry is favorable in many cases, digital technologies present major challenges to the stakeholders of the sports industry in practice: first, they struggle to understand and unleash the full beneficial impact of digital technologies due to a lack of know-how and missing resources. Second, digital technologies facilitate the integration of new actors that provide the sports industry with novel resources, skills, and competences. As a result, the competitive dynamics are increased, and industry competition may be reshaped.

From an academic perspective, this ongoing digitalization of the sports industry provides unique opportunities to study digitalization-related phenomena that cannot be easily observed in other business contexts. In this vein, a research agenda for the Information Systems (IS) discipline on “why and how should we study sports digitalization” was proposed by Xiao et al. (2017, p. 3). The cumulative dissertation at hand, therefore, addresses both the abovementioned challenges from practice and the calls from researchers. To do so, it is comprised of five publications (I – V) that tackle problems derived from practice and demonstrate theoretical knowledge as well as practical guidance on how digital technologies can be applied beneficially in the sports industry.

Publication I examines the current state of research on digital technologies in the sports industry in the IS discipline through a concept-driven literature review. Based on these initial findings, Publication II applies data mining techniques and a sentiment analysis to investigate the impact of social media on athletes’ performances. In Publication III, state-of-the-art data mining and natural language processing techniques are applied to extract the wisdom of an online community of sports tipsters to improve sports betting returns. Publication IV investigates how sports organizations can use social media to enhance international fan engagement. Lastly, Publication V demonstrates how a web-based information technology (IT) tool supports sports organizations during the analysis and processing of cooperation requests.

Zusammenfassung

Digitale Technologien, wie das Cloud Computing, elektronische Plattformen und die künstliche Intelligenz, ermöglichen die Verbesserung bestehender sowie die Innovation völlig neuer Produkte und Dienstleistungen in der Sportindustrie. Beispielsweise sind Over-the-top-Mediendienste wie DAZN oder Amazons Sport-Streaming-Initiative digitale Technologien, die eine massive Veränderung der Fernsehgewohnheiten von Sportfans fördern. Während die Digitalisierung der Sportindustrie viele Vorteile mit sich bringt, stellen digitale Technologien die Stakeholder der Sportindustrie in der Praxis häufig aber auch vor große Herausforderungen: Erstens haben sie aufgrund mangelnden Know-hows und fehlender Ressourcen Mühe, die positiven Auswirkungen der digitalen Technologien zu verstehen und auszuschöpfen. Zweitens erleichtern digitale Technologien die Integration neuer Akteure, die der Sportindustrie fortschrittliche Ressourcen, Fähigkeiten und Kompetenzen zur Verfügung stellen. Dadurch verändert sich die Wettbewerbsdynamik und der Wettbewerb wird erhöht.

Aus akademischer Sicht bietet diese fortschreitende Digitalisierung der Sportindustrie einzigartige Möglichkeiten, um Phänomene, die im Zusammenhang mit der Digitalisierung auftreten und in üblichen Geschäftskontexten nicht leicht zu beobachten sind, zu analysieren. Daher haben Xiao et al. (2017) eine Forschungsagenda für die Information-Systems(IS)-Disziplin aufgestellt zum Thema warum und wieso wir die Digitalisierung in der Sportindustrie untersuchen sollten. Die vorliegende kumulative Dissertation umfasst fünf Publikationen (I – V), die sich mit aus der Praxis abgeleiteten Problemen befassen und aufzeigen, wie digitale Technologien in der Sportindustrie angewendet werden können.

Publikation I untersucht den aktuellen Stand der Forschung zu digitalen Technologien in der Sportindustrie in der IS-Disziplin mittels einer konzeptgetriebenen Literaturrecherche. Auf der Grundlage dieser ersten Ergebnisse wendet Publikation II Data-Mining-Techniken und eine Sentimentanalyse an, um den Einfluss von Social Media auf die Leistungen von Sportlern zu untersuchen. In Publikation III werden state-of-the-art Data-Mining- und Natural-Language-Processing-Techniken angewandt, um die Weisheit einer Online-Community von Sporttippern zu extrahieren. In Publikation IV wird untersucht, wie Sportorganisationen Social Media nutzen können, um das internationale Fan-Engagement zu verbessern. Publikation V zeigt, wie ein webbasiertes Informationstechnologie(IT)-Tool Sportorganisationen bei der automatisierten Analyse und Bearbeitung von Kooperationsanfragen unterstützt.

Introduction

Advances in digital technologies, such as cloud computing, electronic platforms, and artificial intelligence, enable the improvement of existing as well as the innovation of entirely new products and services in the sports industry (Gruettner, 2019; Ratten, 2017; Xiao et al., 2017). For example, over-the-top media services such as DAZN or Amazon's sports streaming initiative are digital technologies that promote a massive shift in the viewing habits of fans – away from traditional television broadcasting and towards tailored live and on-demand streaming sports consumption. However, while the digitalization of the sports industry is favorable in many cases (Aversa et al., 2018; Yang et al., 2012), digital technologies present major challenges to the stakeholders of the sports industry in practice (Gruettner, 2019; Xiao et al., 2017): first, they struggle to understand and unleash the full beneficial impact of digital technologies due to a lack of know-how and missing resources. Second, digital technologies facilitate the integration of new actors that provide the sports industry with novel resources, skills, and competences. As a result, the competitive dynamics are increased and industry competition may be reshaped (Porter & Heppelmann, 2014; Xiao et al., 2017).

From an academic perspective, this ongoing digitalization of the sports industry provides unique opportunities to study digitalization-related phenomena, “which cannot be easily observed in generic business contexts” (Xiao et al., 2017, p. 16). This is in line with Chiasson and Davidson (2005), who call for the consideration of industries in Information Systems (IS) research as institutional contexts by explaining how structures of certain industries, including schemas, rules, norms, and routines, become entrenched as authoritative guidelines. In this vein, Xiao et al. (2017) proposed a research agenda for the IS discipline on “why and how should we study sports digitalization” (p. 3). The reasons why we should investigate the sports industry are based on its manifold special characteristics from which other industries and phenomena can learn and be guided. These characteristics are the complexities embedded in the organizational activities, the heterogeneity of stakeholder groups, the nature of the product consumed, the specific consumers, and the enormous economic, political, and social impact (Gruettner, 2019). The cumulative dissertation at hand addresses both the abovementioned challenges and calls and is comprised of five publications (I – V) that demonstrate theoretical and empirical knowledge on how digital technologies can be applied beneficially in the sports industry.

Digital technologies have been characterized in several ways in the literature. These definitions have in common that they are equipped with an ambivalent ontology (Kallinikos et al., 2013). Digital technologies exist as digital tools and infrastructure, digital platforms, or artifacts with digitized components, applications, or media content (Nambisan et al., 2017; von Briel et al., 2018). They are re-combinable, editable, and distributable (Yoo et al., 2010) – characteristics that allow them to evolve their identity over time (Jarvenpaa & Standaert, 2018; Lehmann & Recker, 2019). Fields of application of digital technologies in the sports industry – defined as the market in which the “products offered to its buyers are sport related and may be goods, services, people, places, or ideas” (Pitts & Stotlar, 2002, p. 4) – are, among many others, analytics solutions to improve the performance of athletes, digital platforms to interact with fans, or pricing predictions for ticket sales (Gruettner, 2019).

The initially conducted literature review on “What We Know and What We Do Not Know About Digital Technologies in the Sports Industry” of this dissertation (Publication I), on the one hand, demonstrates four groups of benefits that can be achieved by the application of digital technologies for the stakeholders of the sports industry. These benefits are (1) improved knowledge processing and creation, (2) process automation, (3) information exchange and digital interaction, and (4) performance improvements through data analytics solutions. On the other hand, the literature review also points out that researchers have treated the sports industry as “just another” type of industry rather than a unique context with its own manifold special characteristics that call for special attention. As a consequence, guidelines for sports practitioners from research on how digital technologies can be applied in the sports industry to achieve the aforementioned benefits are scarce (Gruettner, 2019; Xiao et al., 2017). Publication II to V, therefore, address four problems derived from practice and illustrate how digital technologies can be applied in the sports industry to achieve desired benefits.

Digital technologies and their benefits can be observed and studied on different levels of analysis (Davern & Kauffman, 2000; Gruettner, 2019; Xiao et al., 2017). Example levels of analysis that benefit from the application of digital technologies include individual users, teams and work groups, business processes, firms, and even entire industries. In this dissertation, three levels of analysis in the sports industry are differentiated: (1) an individual level, (2) an internal level, and (3) an external level (Gruettner, 2019). At the individual level, athletes and sports teams benefit from digital technologies. At the internal level, the sports organizations themselves benefit from digital technologies, for instance, their management team, coaching staff, and other

support staff (e.g., doctors and statisticians). The external level includes people, groups, and entities that interact with stakeholders from the individual and internal level in the creation of shared benefits. They are not directly involved in the practice of sports. External stakeholders are, for instance, professional sports leagues, sports federations, sports fans, technology vendors, media companies, player agents, sponsors, and investors.

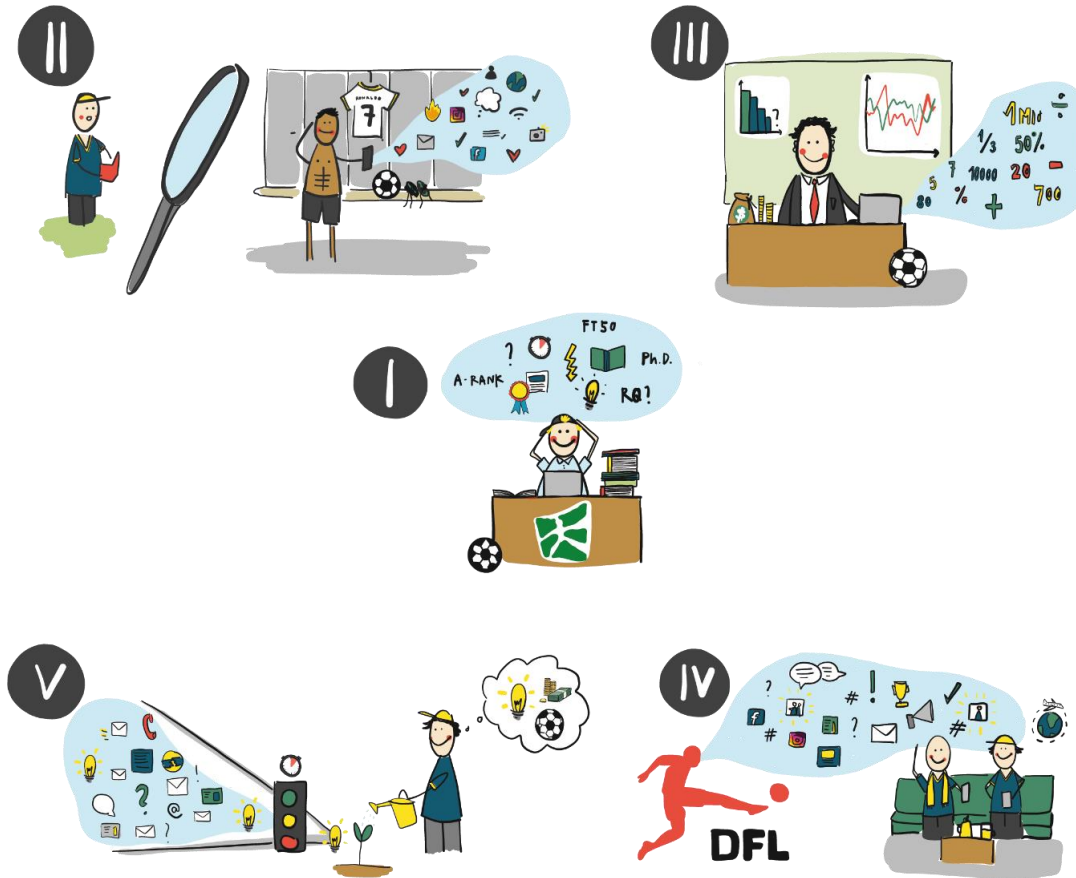
The five publications included in this dissertation analyze selected digital technologies and their accompanied benefits in the sports industry through different theoretical lenses and on different levels of analysis that are described as follows. Publication I examines the current state of research on digital technologies in the sports industry in the IS discipline through a concept-driven literature review. Based on these initial findings, Publication II applies data mining techniques and a sentiment analysis to investigate the impact of social media on athletes' performances. The results are especially relevant for coaches and team managers at the internal level that can identify athletes who use social media excessively. In Publication III, state-of-the-art data mining and natural language processing techniques are applied to extract the wisdom of an online community of sports tipsters to improve sports betting returns. The results of this publication support bookmakers and tipsters at the external level to either protect them from losses or to improve their betting returns. Publication IV investigates how sports organizations (i.e., at the internal level) can use social media to enhance international fan engagement. Lastly, Publication V demonstrates how a web-based information technology (IT) tool supports sports organizations during the analysis and processing of cooperation requests (i.e., at the internal level). Figure 1 illustrates publications I – V in the chronological order of their development.

The findings of each publication contribute in facets to our understanding of how digital technologies can be applied beneficially in the sports industry. Therefore, the dissertation at hand assists researchers and sports practitioners alike and sets the stage for further efforts that investigate applications of digital technologies in the sports industry. Future research should, for instance, inspire and guide sports practitioners to create novel products and services enabled through shared digital technologies in the sports industry. Novel products and services can only be achieved by comprehensively exchanging the gained digital capabilities through digital technologies. For instance, mixed reality smart glass manufacturers such as Google and its HoloLens can partner with TV providers to enhance fans' television experiences and change the way sports are consumed in the future.

Publications I – V, including abstracts as well as general information (authors and publication outlet and status), are presented and summarized on the following pages.

Figure 1

Illustration of Publications I – V in the Chronological Order of Their Development



Publication I

What We Know and What We Do Not Know

About Digital Technologies in the Sports Industry

Abstract

Recent advances in digital technologies (DTs) have transformed various aspects of how the sports industry operates and competes. Common applications of DTs on- and off-site the field of play include, for example, analytics solutions to improve players' performance, pricing predictions for ticket sales, or digital platforms to interact with fans. DTs further facilitate shared digital capabilities, thereby allowing for the integration of new stakeholders and thus giving rise to new digital ecosystems in the sports industry. However, Information Systems' (IS) research on DTs in the sports industry is still nascent. Hence, the role of DTs for the sports industry is not entirely

understood. Therefore, we analyze how DTs shape the sports industry based on a concept-driven literature review. The analysis of 16 publications yields four groups of benefits that can be achieved by the usage of DTs for the stakeholders embedded in the digital ecosystem of the sports industry.

Author(s)

Arne Grüttner

Publication Outlet and Status

Published in the Proceedings of the 25th Americas Conference on Information Systems (AMCIS), 2019. [VHB-JOURQUAL3: D-ranked]

Publication II

The New Window to Athlete's Soul –

What Social Media Tells Us About Athletes' Performances

Abstract

Professional sports has evolved from a game to an organization that has been codified, strategized, and commercialized. One factor that is shaping the sports industry is the pervasiveness of social media. On the one hand, social media is used as a powerful medium for distributing and getting news, engaging in topical discussions, and empowering brands. On the other hand, social media has become a crucial mouthpiece for athletes to interact with peers, share opinions, thoughts, and feelings. However, millions of followers, tweets, and likes later, researchers, practitioners, and athletes alike ask whether social media has an impact on an athlete's performance. We conducted a social media usage and a sentiment analysis of 124,341 Twitter tweets extracted from 31 tennis athletes. We linked these data to 8,095 corresponding match day performances. The results show that high social media usage has a significant negative impact on athletes' performance.

Author(s)

Arne Grüttner, Min Vitisvorakarn, Thiemo Wambsganss, Roman Rietsche, and Andrea Back

Publication Outlet and Status

Published in the Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS), 2020. [VHB-JOURQUAL3: C-ranked]

Publication III

From Data to Dollar –

Using the Wisdom of an Online Tipster Community to

Improve Sports Betting Returns

Abstract

With thousands of (online) bookmakers accepting wagers on sporting events, sports betting has become a billion-dollar business worldwide. Therefore, researchers and practitioners have gathered interest in investigating the “wisdom-of-crowds” effect in online tipster communities to predict the outcomes of sports events. We extracted 1,534,041 tips of 3,484 tipsters from Blogabet.com and used this user-generated content to investigate whether there is wisdom in online tipster communities that can be used to improve betting returns. We applied state-of-the-art data mining and natural language processing techniques and tested our hypotheses using quantitative research methods. Our results demonstrate that there is indeed wisdom in such online tipster communities that can improve sports betting returns. Tipsters won 3.29% more tips than the implied win probability set by bookmakers and produced averaged yields of 3.97%. We further identified four characteristics that are significant indicators for smarter sub-crowds within the overall crowd of an online tipster community.

Author(s)

Arne Grüttner, Thiemo Wambsganss, and Andrea Back

Publication Outlet and Status

In press for publication in the European Journal of International Management (EJIM) – Special Issue on: "International Sports Management". [VHB-JOURQUAL3: B-ranked]

Publication IV

Going Global:

Enhancing International Social Media Fan Engagement –

Evidence from the German Bundesliga

Abstract

The unbreakable boundary between the physical and online world in sports is getting blurry. In turn, the glamour of sports moves to a global digital spotlight. Sports organizations try to seize this opportunity to digitally attract international fans by utilizing social media. However, they currently lack an understanding of the sports content drivers to enhance fan engagement and how to interact with international fans due to cross-cultural effects. We build upon consumer engagement theory and link sports organizations' brand attributes with the concept of visual emotional and informational appeals. Our experiment demonstrated: it is crucial for sports organizations to understand how to adjust visual appeals to maximize content effectiveness, as our analyses yielded evidence that emotional appeals have a significant positive effect on fan engagement. Furthermore, the results showed that the relationship between sports content, visual appeals, and fan engagement is significantly moderated by cross-cultural and media type effects.

Author(s)

Arne Grüttner, Laurenz O. H. Haferbeck, Roman Rietsche, and Andrea Back

Publication Outlet and Status

Submitted to the Proceedings of the 29th European Conference on Information Systems (ECIS), 2021. [VHB-JOURQUAL3: B-ranked]

Publication V [German]

Konfrontiert mit einer Flut von Kooperationsanfragen:

So meistern Sportorganisationen die systematische Analyse und Bearbeitung

Abstract

Die zunehmende Digitalisierung und tiefgreifende Kommerzialisierung der Sportindustrie hat dazu geführt, dass viele Sportorganisationen zu nationalen und globalen Marken mit enormer Reichweite und Popularität geworden sind. Als Folge erhalten Sportorganisationen eine Vielzahl von Kooperationsanfragen (z.B. von Start-

ups). Die sich daraus ergebenden Möglichkeiten zur Verbesserung und Neuerfindung von Produkten, Dienstleistungen und Betriebsabläufen bleiben jedoch oft ungenutzt, da es an Ressourcen und an etablierten Prozessen bei der Analyse und Bearbeitung dieser Kooperationsanfragen mangelt. Wir haben auf Basis der wissenschaftlichen Methode Design Science Research ein IT-Tool konzipiert, das sowohl eine initiale automatisierte Bewertung von Kooperationsanfragen übernimmt als auch nachgelagerte Arbeitsschritte steuert und optimiert.

Author(s)

Arne Grüttner, Thomas Braschler, and Andrea Back

Publication Outlet and Status

Published in Wirtschaftsinformatik & Management – Issue on: "IT und Sport", 2020.
[Practitioner magazine]

References

- Aversa, P., Cabantous, L., & Haeffliger, S. (2018). When decision support systems fail: Insights for strategic information systems from Formula 1. *Journal of Strategic Information Systems*, 27(3), 221–236. <https://doi.org/10.1016/j.jsis.2018.03.002>
- Chiasson, M. W., & Davidson, E. (2005). Taking Industry Seriously in Information Systems Research. *MIS Quarterly*, 29(4), 591–606. <https://doi.org/10.2307/25148701>
- Davern, M. J., & Kauffman, R. J. (2000). Discovering Potential and Realizing Value from Information Technology Investments. *Journal of Management Information Systems*, 16(4), 121–143. <https://doi.org/10.1080/07421222.2000.11518268>
- Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. *Proceedings of the 25th Americas Conference on Information Systems (AMCIS) 2019*.
- Jarvenpaa, S. L., & Standaert, W. (2018). Digital Probes as Opening Possibilities of Generativity. *Journal of the Association for Information Systems*, 19(10), 982–1000.
- Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The Ambivalent Ontology of Digital Artifacts. *MIS Quarterly*, 37(2), 357–370.

- Lehmann, J., & Recker, J. (2019). Offerings that are “Ever-in-the-Making”: Post-Launch Continuous Digital Innovation in Late-Stage Entrepreneurial Ventures. *Proceedings of the 40th International Conference on Information Systems (ICIS) 2019*, 1–17.
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. *MIS Quarterly*, 41(1), 223–238. <https://doi.org/10.25300/MISQ/2017/41:1.03>
- Pitts, B. G., & Stotlar, D. K. (2002). *Fundamentals of Sport Marketing* (2nd ed.). Fitness Information Technology.
- Porter, M. E., & Heppelmann, J. E. (2014). How Smart, Connected Products Are Transforming Competition. *Harvard Business Review*, 92(11), 64–88.
- Ratten, V. (2017). *Sports Innovation Management* (1st ed.). New York, New York: Routledge.
- von Briel, F., Davidsson, P., & Recker, J. (2018). Digital Technologies as External Enablers of New Venture Creation in the IT Hardware Sector. *Entrepreneurship Theory and Practice*, 42(1), 47–69. <https://doi.org/10.1177/1042258717732779>
- Xiao, X., Hedman, J., Ter, F., Tan, C., Tan, C.-W., Lim, E. T. K., Clemenson, T., Henningsson, S., Vatrappu, R., Rao Mukkamala, R., & Van Hillegersberg, J. (2017). Sports Digitalization: An Overview and A Research Agenda. *Proceedings of the 38th International Conference on Information Systems (ICIS) 2017*.
- Yang, L., Su, G., & Yuan, H. (2012). Design Principles of Integrated Information Platform for Emergency Responses: The Case of 2008 Beijing Olympic Games. *Information Systems Research*, 23(3-part-1), 761–786. <https://doi.org/10.1287/isre.1110.0387>
- Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, 21(4), 724–735. <https://doi.org/10.1287/isre.1100.0322>

Publication I

What We Know and What We Do Not Know About Digital Technologies in the Sports Industry



Author(s)

Arne Grüttner

Published in the Proceedings of the 25th Americas Conference on Information
Systems (AMCIS), 2019.

Abstract

Recent advances in digital technologies (DTs) have transformed various aspects of how the sports industry operates and competes. Common applications of DTs on- and off-site the field of play include, for example, analytics solutions to improve players' performance, pricing predictions for ticket sales, or digital platforms to interact with fans. DTs further facilitate shared digital capabilities, thereby allowing for the integration of new stakeholders and thus giving rise to new digital ecosystems in the sports industry. However, Information Systems' (IS) research on DTs in the sports industry is still nascent. Hence, the role of DTs for the sports industry is not entirely understood. Therefore, we analyze how DTs shape the sports industry based on a concept-driven literature review. The analysis of 16 publications yields four groups of benefits that can be achieved by the usage of DTs for the stakeholders embedded in the digital ecosystem of the sports industry.

Keywords: digital ecosystem; digital technology; literature review; sports digitalization; sports industry

Introduction

In recent years, sports has evolved from an activity of game to an activity of organization that has been codified, strategized, professionalized, and commercialized (Davenport, 2014; Xiao et al., 2017). With an estimated total market value of over 500 billion dollars, ranking as one of the top business markets globally, the sports industry has a huge economic as well as social impact (PricewaterhouseCoopers [PwC], 2018). One major development shaping the sports industry is widespread digitalization. Digital technologies (DTs) such as cloud computing, electronic platforms, and artificial intelligence have transformed various aspects of how the sports industry operates and competes (Davenport, 2014; Xiao et al., 2017). Common applications of DTs on- and off-site the field of play include, for example, analytics solutions to improve players' performance (e.g., Cordes and Olfman (2016)), pricing predictions for ticket sales (e.g., Mignerat and Audebrand (2010)), or digital platforms to interact with sports consumers, so-called fans, more closely (e.g., Wulf et al. (2015)).

DTs enable the convergence of heterogeneous knowledge and information into new products and services (e.g., 3D printing, data analytics, or mobile computing) (Nambisan et al., 2017). DTs further facilitate shared digital capabilities that can either be independently customized for a company's own ecosystem or foster simultaneous use by multiple companies (Tan et al., 2017), thereby allowing for the integration of

new stakeholders and thus giving rise to new digital ecosystems that operate on shared DTs (Senyo et al., 2019). The partnership of the National Football League (NFL) and the Chinese Internet giant Alibaba is a prominent example of a digital ecosystem in the sports industry that is only made possible by the rise of DTs. More precisely, the NFL and Alibaba share their digital capabilities to broadcast NFL games, not only on traditional broadcasters in the U.S. but also live on a digital platform in Asia (Forbes, 2019). However, traditional stakeholders of the sports industry invest in DTs to continue digitalizing their digital ecosystems. These stakeholders face challenges such as large investments, entry barriers, and missing know-how due to high-levels of complexity, which are embedded in the sports industry (Davenport, 2014; Xiao et al., 2017). For instance, new actors, such as data providers and livestreaming services, are becoming critical constituents of a new digital ecosystem providing the sports industry with new resources, skills, and competences. Guidelines from the Information Systems (IS) research are scarce because researchers have not investigated how new DTs and the entrance of new actors change the ecosystem dynamics (Senyo et al., 2019; Xiao et al., 2017). Hence, IS research lacks an analysis of the role of DTs in the specific context of the sports industry. Therefore, we aim to answer the following research question (RQ):

RQ: *Which role do DTs play for the digitalization of the sports industry?*

To answer this question, we conduct a systematic concept-driven literature review that provides an overview of how IS researchers put DTs into perspective in the sports industry. We analyze DTs' role for the sports industry by identifying the benefits that stem from their usage. To provide explanations how DTs can be beneficially leveraged among the stakeholders of the sports industry to enable shared digital capabilities for researches and practitioners alike, we establish logical links from the various stakeholders embedded in the digital ecosystem to DTs' role.

This paper is structured as follows: In the next section, we introduce sports digitalization as an academic discipline in IS research and provide an understanding of the digital ecosystem of the sports industry. We describe our research approach in Section 3. In Section 4, we present the findings covering the role of DTs for the sports industry. We then discuss our findings and avenues for future research as well as limitations of our study in Section 5. Finally, we conclude our paper in Section 6.

Theoretical Background and Related Work

Sports Digitalization as an Academic Discipline

There is no consensus on how sports should be defined in academia. A common definition typically entails characteristics of competitiveness, a ‘non-hostility’ nature, physicality (no matter to what extent), and also conformance of predefined rules (Wright, 2009). Following the work by Loy (1968), sports is also an organizing activity driven by institutional logics. We further define the sports industry as the market in which the products offered to its buyers are sports-related and may be activities, goods, services, people, places, or ideas (Pitts et al., 1994). While DTs in the sports industry are pervasive in practice, little academic research has been conducted on DTs in the sports industry in the IS discipline (Shah et al., 2015; Tan et al., 2017; Xiao et al., 2017). However, the massive transformation of the sports industry, triggered by digitalization, the growing public interest, and the large market potential associated with it, recently led to a surge in the level of scholarly interest in IS. Therefore, an IS community around sports has been established.

One of the most decisive articles for studying the sports industry in the IS discipline was published by Xiao et al. (2017). In their completed research article, Xiao et al. (2017) provide an understanding of “why and how should we study sports digitalization in the IS discipline?” (p. 3). We share the opinion of the authors that the sports industry, as a unique context with distinctive characteristics, calls for special attention rather than being treated as just another type of industry. The reasons for the uniqueness of the sports industry are manifold:

- the complexity embedded in the organizational activities – e.g., the structure of sports organizations as non-profit organizations and their competitive and secretive nature (Xiao et al., 2017);
- the heterogeneity of stakeholder groups – e.g., the different working methods and mindsets of traditional stakeholders, such as sports associations, and new stakeholders, such as data providers (Babiak & Wolfe, 2009; Tan et al., 2017);
- the nature of the product consumed – e.g., sports is a rather an intangible than a tangible product;
- the specific consumers – e.g., fans are mainly driven by emotions such as passion and social values rather than by rational evaluations (Babiak & Wolfe, 2009);

- and the enormous economic, political, and social impact – e.g., not only fans, sponsors, or investors spend their time and money in the sports industry, but governments also push into it to use the radiance of sports as a political instrument that demonstrates power and influence.

In sum, although the use of DTs is salient in the sports industry, IS research has not yet paid extensive attention to it. We argue that dismissing the sports industry as just another empirical context will translate to missed opportunities for comprehending an IT-driven phenomenon that might display interesting dynamics due to the uniqueness of the context at a theoretical level. This is in line with Chiasson and Davidson (2005), who call for the consideration of industries in IS research as institutional contexts by explaining how structures of certain industries, including schemas, rules, norms, and routines, become entrenched.

Digital Ecosystem of the Sports Industry

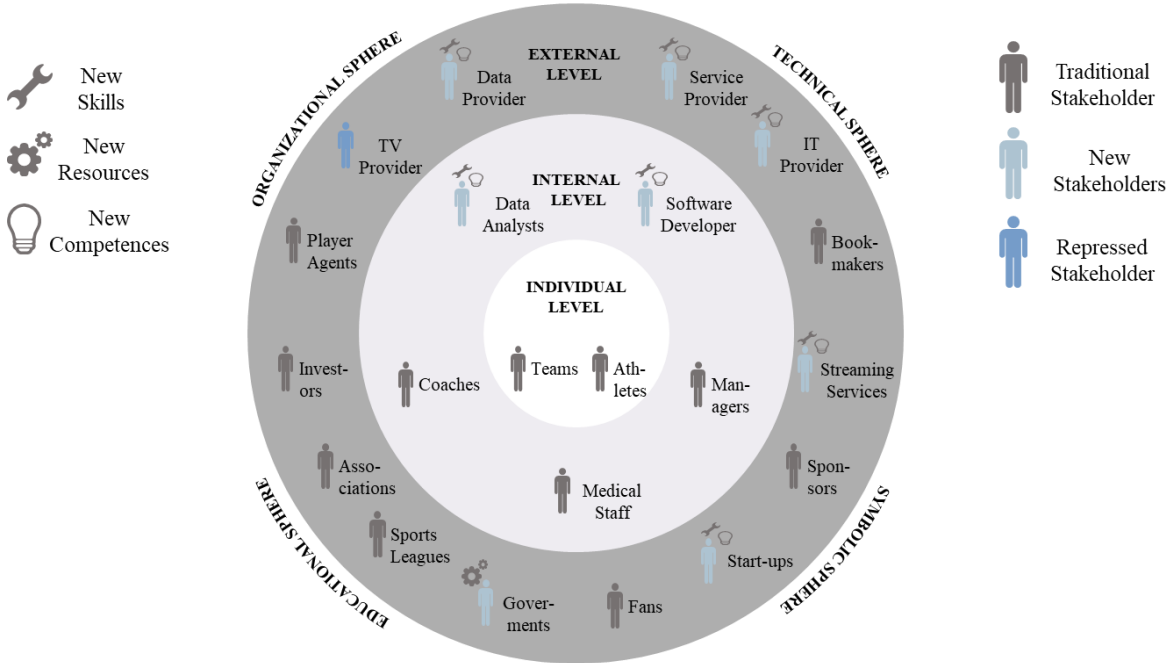
A digital ecosystem is a collaborative environment made up of different entities that co-create value through information and communication technologies. In a digital ecosystem, companies work cooperatively and competitively to support new products, satisfy customer needs, and eventually incorporate innovations (Senyo et al., 2019). There is no well-founded and uniform description of the digital ecosystem of the sports industry described in literature (Holland, 2015; Xiao et al., 2017). Therefore, based on related work and prior studies on the business value of DTs, we developed a conceptual framework that helps to understand the composition of the digital ecosystem of the sports industry. The framework distinguishes between three types of stakeholders embedded in the sports industry, in accordance with Bower and Christensen's (1995) disruptive innovation theory: (1) traditional stakeholders who have been in the industry for a long time, (2) new stakeholders, such as innovators and disruptors (e.g., Information Technology (IT) stakeholders that provide the sports industry with new recourses, skills, and competences), and (3) repressed stakeholders that fear to be replaced by the new stakeholders entering the industry. This framework is depicted in Figure 1 and explained as follows.

On the one hand, according to Davern and Kauffman (2000), the potential value of DTs can be observed at several levels of analysis, at which flows of DTs value become discernible for the investing firm (i.e., from micro to macro level). Examples of different levels of analysis where value can accrue (directly or as consumers of the output) include individual users, teams and work groups, business processes, firm level, and even

industry level. For the purpose of this study, we differentiate between three levels of analysis in the sports industry, based on the article by Caya and Bourdon (2016): (1) an individual level, (2) an internal level, and (3) an external level. At the individual level, we find athletes and teams who benefit from the application of DTs. The internal level of analysis consists of stakeholders within a sports organization. For instance, their management team, coaching staff, and other support staff, such as doctors and statisticians. External level stakeholders are people, groups, and entities that are not directly concerned with the transformation of potential value into realized value creation of a sports organization. Stakeholders at this level interact with stakeholders of the individual and internal level in the creation of shared value. External stakeholders are, for instance, professional sports leagues, sports federations, sports fans, technology vendors, media companies, player agents, sponsors, and investors.

On the other hand, the framework divides the digital ecosystem of the sports industry into four spheres with regard to the institutional activities of sports, which are described in Loy (1968): (1) the organizational sphere, (2) the technical sphere, (3) the symbolic sphere, and (4) the educational sphere. The organizational sphere describes the organizational aspects of the sports industry in terms of teams, sponsorship, and government. Activities in this sphere deal with the administration of sports and the pursuit of both business outcomes and sports outcomes. Therefore, organizational activities not only encompass activities in direct relation to sports production, but they also include business activities that ensure targeted outcomes (e.g., customer satisfaction and revenue). The technical sphere describes “the material equipment, physical skills, and body of knowledge which are necessary for the conduct of competition and potentially available for technical improvements in competition” (Loy, 1968, p. 8). Activities in this sphere, for example, include skills and knowledge possessed by coaches to enhance the technical equipment. The symbolic sphere of sports includes elements of secrecy, display, and ritual. Sports consumers are a major part of the symbolic sphere of sports, both as observers and as participants, depending on their roles and level of engagement. Finally, the educational sphere of sports deals with the activities of acquiring the above-mentioned skills and knowledge of the technical sphere, thus focusing on those activities related to the transmission of skills and knowledge.

Indeed, the stakeholder levels and institutional activity spheres are closely related to each other and sometimes overlap. These overlaps reflect the high levels of complexity contained in the digital ecosystem of the sports industry.

Figure 1*Stakeholders Within the Digital Ecosystem of the Sports Industry***Research Method****Literature Review**

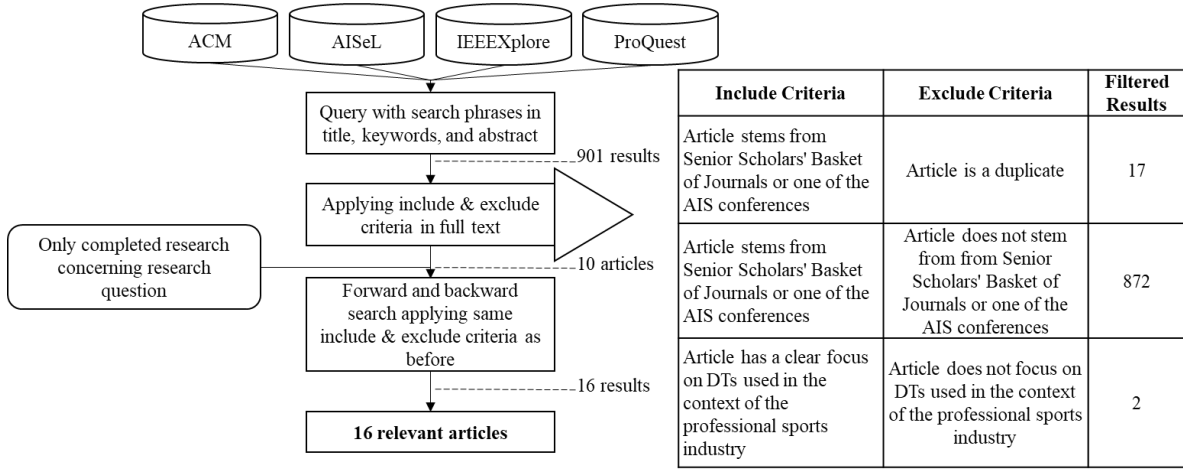
Research lacks a comprehensive overview that synthesizes the role DTs play in the context of the sports industry. Therefore, we performed a concept-driven literature review, relying on an adjusted five-phase identification and selection process that was originally derived by vom Brocke et al. (2009).

According to vom Brocke et al. (2009), journals are selected in the first phase. It is recommended to focus on articles published in scholarly journals or at renowned conferences as these articles are usually peer-reviewed before publication (Rowley & Slack, 2004; Webster & Watson, 2002). Due to those recommendations, only articles stemming from the Senior Scholars' Basket of Journals or articles published at one of the conferences organized by the Association for Information Systems (AIS) (i.e., AMCIS, ECIS, HICSS, ICIS, PACIS) are chosen for further analysis. In the second phase of the data collection process, we selected databases for examination. This study only includes databases related to the IS discipline: ACM Digital Library (ACM), AIS Electronic Library (AISeL), IEEEExplore, and ProQuest. The third phase requires the construction of a search query that includes search phrases and is executed on the databases chosen in the second phase. It is commonly recommended to use a set of

search phrases that are as precise as possible to exclude results covering topics or research questions that do not contribute to the research issue (Rowley & Slack, 2004). Thus, potentially relevant articles have to match the following search phrases for title, abstract, or keywords: (baseball; basketball; e-sport*; football; soccer; sport*) AND (analytics; digitalisation; digitalization; “information technology”).

The first search phrase for the search query includes the combination of different expressions of sports, such as short form, singular and plural expressions as well as expressions for major sports disciplines. The sports disciplines are added because articles deemed to be relevant for this study are missing when using only expressions for sports. The second search phrase connected to the first search phrase by an AND operator contains expressions related to digitalization and IT. The combination of both search phrases ensures that relevant articles are found, and the application examines the role of DTs for the sports industry.

The above-mentioned search phrases are each transformed to the specific syntaxes of the literature databases. The search resulted in a total of 901 articles (the numbers are as of 21 September 2018). These articles’ full texts are assessed with regard to the include and exclude criteria (see, Figure 2). After applying the include/exclude criteria, 10 relevant articles remained. In the fifth phase, the 10 remaining articles from the fourth phase are accumulated by identifying further relevant studies using the approach, as suggested by Webster and Watson (2002), of searching forward and backward. The Web of Science, as recommended by Webster and Watson (2002), is used to search forward. Additionally, Google Scholar is used, since researchers’ experience showed that Google Scholar provides a more comprehensive impression on the actual number of citations. These articles are evaluated using the method described in the fourth phase and if identified as relevant, they are added to the pool of results. This search reveals another six relevant articles. The literature review process is finished after one-step of forward and backward search of the initially identified 10 articles. As a result, we identified 16 articles in total. Figure 2 summarizes the results of the literature review process. For further analysis, a concept matrix is used, which is explained in Section 4.

Figure 2*Overview of Literature Selection Process***Data Analysis**

We aim to analyze DTs' role in the sports industry, which stakeholders benefit from the implementation of DTs, and how DTs can be beneficially leveraged among them to create shared digital capabilities. Therefore, to enable a comparison between the identified articles and to find patterns in the application of DTs, we analyzed our identified articles by five consecutive steps, which are described in the following.

First, we performed a qualitative content analysis of the articles and coded the context of the study including the DTs described, the type of sports (if mentioned), whether the article comprises empirical data, and the place of publication. Second, to identify practical applications of DTs in the sports industry only, we excluded the articles that have no empirical context from our sample for further analysis. As a result, the sample size was reduced to 12 articles. Third, in these 12 articles, we collected information about the mentioned stakeholder levels, the institutional activity spheres affected by the DTs, and whether the DTs have a supportive, enabling, or replacing role in the context of the article. Fourth, we developed a concept matrix that is based on the findings of the previous steps. This concept matrix is used to find patterns in the application of DTs to show which role DTs play in the context of the sports industry. The applied approach is to analyze DTs usage based on written case material. Specifically, we analyze which stakeholder levels have initiated the implementation of the DTs described in the article, the institutional activity spheres affected by the implementation, and DTs' role (as classified below). Following the logic of gradually decomposing complex concepts to make them more easily accessible, these impacts, altogether, form a cause-and-effect structure. By coding the impacts evident in the case material, the role of DTs is

investigated and structured to build logic links between the stakeholder levels and DTs' role via each institutional activity sphere. We then counted the occurrences of each logical link found in the concept matrix to identify patterns in the application of DTs. These patterns may help us gain an understanding of how DTs can be beneficially leveraged among stakeholders in the sports industry to gain shared digital capabilities. As a last step, we used the identified logical links and collected information about the mentioned benefits of the DT described in the articles and merged similar benefits into groups. The benefits were clustered into benefit groups if they were logically related to the same subject. We followed the theoretical approach of clustering proposed by Jankowicz (2003).

Results

Overview of Selected Articles

Table 1 provides an overview of the 16 articles identified. There is no article which stems from one of the journals included in the Senior Scholars' Basket of Journals. All identified articles stem from one of the AIS conferences, which underlines the high degree of topicality, as conferences have shorter review cycles than journals. 12 articles have an empirical context. The other four articles either develop conceptual frameworks for the usage of DTs (e.g., Wilkerson and Gupta (2016)) or provide an overview and a research agenda for sports digitalization (e.g., Xiao et al. (2017)). The remaining 12 articles rely on empirical evidence in their analyses, differ in context, and cover a broad variety of topics, although eight out of the 12 articles deal with data analytics solutions. In these articles, various applications of DTs in the sports industry are discussed. For example, ranging from algorithms to predict players' performance (e.g., Cordes and Olfman (2016)) through the development of a practice-based research network system to gain a better understanding of sports-related injuries (e.g., Lam et al. (2016)) to an analytics dashboard for improving decision-making in ocean race sailing (e.g., van Hillegersberg et al. (2017)).

Table 1*Overview of Identified Articles*

Article	Context of the Article	Empirical	Published
Caya and Bourdon (2016)	Development of a conceptual framework that identifies value creation from business intelligence use in competitive sports.	No	HICSS
Cordes and Olfman (2016)	Design science research approach to predict athletic performance with a genetic algorithm in football.	Yes	AMCIS
Fohrholz and Glaschke (2016)	Capacity and pricing predictions of coach vendors including a case study about the European Soccer Championship 2016.	Yes	HICSS
Hanisch and Hanisch (2007)	Development of a speech-enabled human-computer interface that ensures injury tracking in elite sporting clubs.	Yes	PACIS
Holland (2015)	Teaching case that investigates the impact of the Internet and social media on the sports market to develop a strategy for sports clubs.	Yes	ECIS
Lam et al. (2016)	Development of a practice-based research network system to gain a better understanding of sports-related injuries.	Yes	AMCIS
Loucopoulos and Kavakli (2016)	Research on enterprise capability modelling challenges to address dynamic requirements using a sports event as an example.	No	AMCIS
Mignerat and Audebrand (2010)	Investigation of the roles and actions of institutional entrepreneurs in the selection and implementation of IT for sporting events.	Yes	ICIS
Morgan and Ravindran (2017)	Teaching case that describes the use of business analytics to target baseball-free agents.	Yes	AMCIS
Shah et al. (2015)	Development of an analytics platform for professional sports teams (i.e., soccer) using a design science research methodology.	Yes	ICIS
Tan et al. (2017)	Case study to assess the differences between IT-enabled capabilities in sports (i.e., FC Bayern Munich) and traditional businesses.	Yes	AMCIS
van Hillegersberg et al. (2017)	Development of an analytics dashboard for improving decision-making in ocean race sailing.	Yes	AMCIS
Wilkerson and Gupta (2016)	Development of a framework that uses analytics to improve sports injuries prevention.	No	AMCIS
Wulf et al. (2015)	Teaching case that illustrates how value can be generated by social media using the example of a soccer club (i.e., FC Bayern Munich).	Yes	ECIS
Xiao et al. (2017)	Overview and research agenda of sports digitalization in the IS academic literature stream.	No	ICIS
Xu and Yu (2015)	Case study that uses sentiment analysis to detect players' pre-game moods (i.e., basketball) to predict their on-court performance.	Yes	HICSS

Linkage between Stakeholder Level, Institutional Activity Sphere, and Role of Digital Technology

To improve clarity about the role of DTs for the sports industry and to understand which benefits yield from the use of DTs, it is analyzed how DTs achieve their role via each stakeholder level and institutional activity sphere (see, Table 2). Therefore, we excluded the articles that have no empirical context from our sample. As a result, the sample size

is reduced to 12 articles at this stage. The articles are analyzed by searching for logical connections between the role of the DTs described in the article and the involved stakeholder levels via an institutional activity sphere. The role of DTs described in the articles is divided into an enabling, a supporting, and a replacing role. The enabling role describes DTs that create new organizational capabilities. The supporting role includes the DTs that support and simplify existing organizational capabilities. The replacing role is defined as DTs that replace existing capabilities such as the automation of manually performed activities.

While examining the connection between the involved stakeholder level, the affected institutional activity sphere, and the role of the DTs, the data gathered shows that the linkage between the internal stakeholder level, the technical activity sphere, and the enabling role of the DTs in use is one of four dominant linkages. Additionally, the internal stakeholder level, the organizational sphere, and the enabling role form another link. These connections are followed by the linkage between the internal level stakeholder, the symbolic activity sphere, and the enabling role. The last observed pattern is the linkage between the internal stakeholder level, the educational sphere, and the enabling role of DTs. The detailed results of the observed patterns are explained in the following subsections.

Table 2

Mapping of Stakeholder Level, Institutional Activity Sphere, and Role of Digital Technology

Article	Stakeholder Level			Institutional Activity Sphere				Role of DT		
	Individual	Internal	External	Organizational	Technical	Symbolic	Educational	Enabling	Supporting	Replacing
Cordes and Olfman (2016)		X			X			X		
Fohrholz and Glaschke (2016)			X	X					X	
Hanisch and Hanisch (2007)		X		X	X			X		
Holland (2015)		X		X		X		X		
Lam et al. (2016)		X		X	X			X		
Mignerat and Audebrand (2010)		X		X	X					X
Morgan and Ravindran (2017)		X			X				X	
Shah et al. (2015)	X	X			X		X	X		
Tan et al. (2017)		X	X	X	X	X	X	X	X	X
van Hillegersberg et al. (2017)		X			X		X	X		
Wulf et al. (2015)		X		X		X		X		
Xu and Yu (2015)		X			X			X		

Internal Stakeholder Level, Technical Activity Sphere, and Enabling Role of Digital Technology

Table 2 shows that seven articles build a linkage between the internal stakeholder level, the technical activity sphere, and the enabling role of the DTs described. The high number of occurrences emphasizes that many DTs described in the IS literature are implemented around the technical activity sphere in the sports industry. In general, while the technical sphere is closely related to the organizational sphere, DTs at this linkage enable an improved knowledge processing and in turn facilitate knowledge creation. For instance, some authors describe the development and implementation of injury tracking systems to gain a better understanding of sports-related injuries in professional sports (e.g., Hanisch and Hanisch (2007) or Lam et al. (2016)). Other authors develop algorithms to understand and predict players' performances (e.g., Cordes and Olfman (2016) or Xu and Yu (2015)).

Internal Stakeholder Level, Organizational Activity Sphere, and Enabling Role of Digital Technology

Additionally, five out of 12 articles establish a connection between the internal stakeholder level, the organizational activity sphere, and the enabling role of the DTs. In contrast to the linkage described in the aforementioned section, DTs at this linkage are used to ensure both sports outcomes and business outcomes (e.g., customer satisfaction or revenue). More specifically, DTs allow process automation, thereby supporting and replacing existing workflows and administrative processes. For example, Mignerat and Audebrand (2010) investigate the role of e-ticketing technologies, which replace paper-based tickets. Likewise, Morgan and Ravindran (2017) focus on business analytics to propose recommendations to target baseball players, which was a manually performed task in the past.

Internal Stakeholder Level, Symbolic Activity Sphere, and Enabling Role of Digital Technology

According to the analyzed articles, authors draw a linkage between the internal stakeholder level, the symbolic activity sphere, and the enabling role of the DTs in use. DTs at this linkage facilitate information exchange and digital interaction (Holland, 2015; Tan et al., 2017; Wulf et al., 2015). Information exchange and digital interaction are especially important in the sports industry, where many fans live outside the actual place of the venue, in remote areas. DTs close this distance gap and, for example, enable

sports organizations to transmit information as well as emotions digitally. As a result, the individual player, the team, and sports organizations can be brought closer to the fans (Tan et al., 2017). In consequence, DTs at this linkage foster a better fan accessibility and a closer relationship building. Three of the identified articles draw this linkage.

Internal Stakeholder Level, Educational Activity Sphere, and Enabling Role of Digital Technology

The analysis shows that three out of 12 articles denote a linkage between the internal stakeholder level, the educational activity sphere, and the enabling role of the DTs. At this linkage, the DTs described in the IS literature are data analytics solutions for tactical information, performance data, and physical actions that are turned into sports accomplishments on the field of play. For instance, DTs enable the collection of large amounts of performance data. This data is then analyzed in real-time by advanced data analytics techniques and displayed in dashboards or on platforms to gain valuable insights for coaches before and in real-time during the game (Shah et al., 2015; Tan et al., 2017). In sum, the benefits of DTs at this linkage ensure (real-time) performance monitoring and in turn performance improvements.

Discussion, Future Research, and Limitations

To elucidate how DTs shape the digital ecosystem of the sports industry, we identified logical links that disclose the interrelation between the various stakeholder levels embedded in the digital ecosystem of the sports industry, the four institutional activity spheres of sports, and the different roles DTs can play. By investigating these links, four patterns emerged that yield to benefit groups. The major benefits are an improved knowledge processing and creation, an enhanced process automation, digital information exchange and a closer digital interaction, and (real-time) performance monitoring and in turn, performance improvements. However, the results of the analysis also show that the majority of the gained digital capabilities only benefit the stakeholders that have initiated to implement the DTs and are not shared with other stakeholders. As a result, potentials for the creation of shared digital capabilities are missed. Therefore, to truly unlock the innovative strengths of DTs in order to create shared digital capabilities – that is, capabilities that emerge from the collaboration and exchange of the acquired resources and gained digital capabilities – stakeholders of the sports industry need to cooperate more closely.

From an academic perspective, our work provides important insights into the applications of DTs in the emerging literature stream of sports digitalization in IS research. First, we proposed a conceptual framework that helps to understand the stakeholder composition and the digital ecosystem of the sports industry. Future research should focus on an identification of new stakeholders to investigate how these stakeholders further shape the digital ecosystem of the sports industry. Such research would extend our understanding of how the entrance of new stakeholders change the ecosystem dynamics. Second, we analyzed currently scarce IS research on DTs in the sports industry, linked our findings with the existing literature stream of institutional sports, and identified desired benefits in the use of DTs. As a result, we provided a unique contribution to an upcoming literature stream that will be highly relevant in the near future. More precisely, our applied mapping can be used in future research to classify and identify further benefits of DTs in the sports industry. Third, our research model comprises relevant constructs – stakeholder levels, the institutional activity spheres of sports, the roles that DTs can play, and logical links – that explain how DTs can beneficially be leveraged to constitute shared digital capabilities. In a next step, our research model can be applied in an empirical setting (such as a case study) to generate a more in-depth understanding of the role of digital technologies in the sports industry. Likewise, our research model can be extended by specific design and engineering requirements to establish an integrated framework for the implementation of DTs within the digital ecosystem of the sports industry. Lastly, given that our investigation has showed that there is no article published in the Senior Scholars' Basket of Journals, we call for more articles focusing on the sports industry in scholarly journals.

From a practical perspective, our research allows us to determine which DTs practitioners should focus on to achieve desired benefits. Hence, our results are especially relevant to executives who navigate their organizations between both sports outcomes and business outcomes. In addition, practitioners can use the results as a guide to design their own digital ecosystem within the overall digital ecosystem of the sports industry. In this sense, our results should also be seen as a wake-up call for the stakeholders of the sports industry, inspiring them to work closer together to create shared digital capabilities. For instance, designers can implement the identified DTs and seek for collaboration partners to produce higher-level products and services. Higher-level products and services can only be achieved by comprehensively exchanging the gained digital capabilities through DTs. For instance, real-time performance data collected by sports organizations can be shared with streaming services to enhance fans' television experiences.

This study is not free from limitations: First, we restricted our analysis to articles which stem from the Senior Scholars' Basket of Journals or articles that are published at one of the AIS conferences only. Second, it is noticeable that we only performed one-step of forward and backward search doing our literature review process. Third, the dataset consists of 16 articles only. Despite the small number of analyzed articles, our sample shows a high degree of topicality as 14 of the articles were published within the last four years (2015-2018). Fourth, it is fair to say that not all stakeholders of the digital ecosystem of the sports industry are mentioned in our investigations. Fifth, as this study aims to improve the comprehension of the role of DTs in the sports industry in the IS discipline, it is reasonable to assume that not every characteristic is covered, partly due to the novelty of the topic. Studies covering new aspects are likely to be conducted in the near future. Thus, an extension of this study's scope can become relevant.

Conclusion

Based on a concept-driven literature review, we investigated the role of DTs in the sports industry in IS research and identified patterns of logical links, which describe how DTs can achieve specific benefits. For research, the described conceptual framework and the identified logical links provide a theoretical explanation of how DTs facilitate desired shared digital capabilities. Future research should concentrate on how new DTs and new stakeholders shape the digital ecosystem of the sports industry. Implications for practitioners are guidelines which DTs they should concentrate on to achieve specific benefits that are relevant for their business and sports outcomes.

References

- Babiak, K., & Wolfe, R. (2009). Determinants of Corporate Social Responsibility in Professional Sport: Internal and External Factors. *Journal of Sport Management*, 23(6), 717–742. <https://doi.org/10.1123/jsm.23.6.717>
- Bower, J. L., & Christensen, C. M. (1995). Disruptive Technologies: Catching the Wave. *Harvard Business Review*, 73(1), 43–53.
- Caya, O., & Bourdon, A. (2016). A Framework of Value Creation from Business Intelligence and Analytics in Competitive Sports. *Proceedings of the 49th Annual Hawaii International Conference on System Sciences (HICSS) 2016*. <https://doi.org/10.1109/HICSS.2016.136>

- Chiasson, M. W., & Davidson, E. (2005). Taking Industry Seriously in Information Systems Research. *MIS Quarterly*, 29(4), 591–606. <https://doi.org/10.2307/25148701>
- Cordes, V., & Olfman, L. (2016). Sports Analytics: Predicting Athletic Performance with a Genetic Algorithm. *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS) 2016*.
- Davenport, T. H. (2014). What Businesses Can Learn From Sports Analytics. *MIT Sloan Management Review*, 10–13.
- Davern, M. J., & Kauffman, R. J. (2000). Discovering Potential and Realizing Value from Information Technology Investments. *Journal of Management Information Systems*, 16(4), 121–143. <https://doi.org/10.1080/07421222.2000.11518268>
- Fohrholz, C., & Glaschke, C. (2016). Predicting travel volumes for long-distance coach services through big data analytics - a case study on German public viewing events during the UEFA EURO 2016. *Proceedings of the 49th Annual Hawaii International Conference on System Sciences (HICSS) 2016*. <https://doi.org/10.1109/HICSS.2016.154>
- Forbes. (2019). *The NFL Is A Sleeping Giant In China*. <https://www.forbes.com/sites/russellflannery/2019/02/01/the-nfl-is-a-sleeping-giant-in-china/#4cb00e995a42e>
- Hanisch, B., & Hanisch, J. (2007). Injury Management: The Development and Implementation of Innovative Software in an Elite Sporting Club. *Proceedings of the 11th Pacific Asia Conference on Information Systems (PACIS) 2007*.
- Holland, C. P. (2015). Internet and Social Media Strategy in Sports Marketing. *Proceedings of the 23rd European Conference on Information Systems (ECIS) 2015*. <https://doi.org/10.18151/7217357>
- Jankowicz, D. (2003). *The Easy Guide to Repertory Grids*. Hoboken, NJ: John Wiley & Sons Ltd.
- Lam, K. C., Sauers, E. L., & McLeod, T. C. V. (2016). The Athletic Training Practice-Based Research Network: A National Model for Point-of-Care Sports Injury and Outcomes Documentation to Improve Athlete Health and Wellness. *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS) 2016*.
- Loucopoulos, P., & Kavakli, E. (2016). Capability Modeling with Application on Large-scale Sports Events. *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS) 2016*.
- Loy, J. W. (1968). The Nature of Sport: A Definitional Effort. *Quest*, 10(1), 1–15. <https://doi.org/10.1080/00336297.1968.10519640>

- Mignerat, M., & Audebrand, L. (2010). Towards the Adoption of e-Refereeing and e-Ticketing in Elite Soccer Championships: an Institutional Perspective. *Proceedings of the 31st International Conference on Information Systems (ICIS) 2010*.
- Morgan, J., & Ravindran, S. (2017). Using Business Analytics to Target Baseball Free Agents: A Case Study. *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS) 2017*.
- Nambisan, S., Lyytinen, K., Majchrzak, A., & Song, M. (2017). Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. *MIS Quarterly*, 41(1), 223–238. <https://doi.org/10.25300/MISQ/2017/41:1.03>
- Pitts, B. G., Fielding, L. W., & Miller, L. K. (1994). Industry segmentation theory and the sport industry: Developing a sport industry segment model. *Sport Marketing Quarterly*, 3(1), 15–24.
- PricewaterhouseCoopers. (2018). *Sports industry: lost in transition?* [https://www.pwc.ch/en/publications/2018/PwC Sports Survey-2018_web.pdf](https://www.pwc.ch/en/publications/2018/PwC%20Sports%20Survey-2018_web.pdf)
- Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News*, 27(6), 31–39. <https://doi.org/10.1108/01409170410784185>
- Senyo, P. K., Liu, K., & Effah, J. (2019). Digital business ecosystem: Literature review and a framework for future research. *International Journal of Information Management*, 47, 52–64. <https://doi.org/10.1016/j.ijinfomgt.2019.01.002>
- Shah, F., Kretzer, M., & Maedche, A. (2015). Designing an Analytics Platform for Professional Sports Teams. *Proceedings of the 36th International Conference on Information Systems (ICIS) 2015*.
- Tan, F. T. C., Hedman, J., & Xiao, X. (2017). Beyond ‘Moneyball’ to Analytics Leadership in Sports: An Ecological Analysis of FC Bayern Munich’s Digital Transformation. *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS) 2017*.
- van Hillegersberg, J., Vroling, M., & Smit, F. (2017). Improving Decision Making in Ocean Race Sailing Using Sensor Data. *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS) 2017*.
- vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. *Proceedings of the 17th European Conference on Information Systems (ECIS) 2009*.
- Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26(2), xiii–xxiii. <https://doi.org/10.1.1.104.6570>

- Wilkerson, G. B., & Gupta, A. (2016). Sports Injuries and Prevention Analytics: Conceptual Framework & Research Opportunities. *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS) 2016*.
- Wright, M. B. (2009). 50 years of OR in sport. *Journal of the Operational Research Society*, 60(S1), 161–168. <https://doi.org/10.1057/jors.2008.170>
- Wulf, J., Söllner, M., Leimeister, J. M., & Brenner, W. (2015). FC Bayern München Goes Social - The Value of Social Media for Professional Sports Clubs. *Proceedings of the 23rd European Conference on Information Systems (ECIS) 2015*. <https://doi.org/10.18151/7217535>
- Xiao, X., Hedman, J., Ter, F., Tan, C., Tan, C.-W., Lim, E. T. K., Clemenson, T., Henningsson, S., Vatrappu, R., Rao Mukkamala, R., & Van Hillegersberg, J. (2017). Sports Digitalization: An Overview and A Research Agenda. *Proceedings of the 38th International Conference on Information Systems (ICIS) 2017*.
- Xu, C., & Yu, Y. (2015). Measuring NBA Players' Mood by Mining Athlete-Generated Content. *Proceedings of the 48th Annual Hawaii International Conference on System Sciences (HICISS) 2015*. <https://doi.org/10.1109/HICSS.2015.205>

Publication II

The New Window to Athlete's Soul – What Social Media Tells Us About Athletes' Performances



Author(s)

Arne Grüttner

Min Vitisvorakarn

Thiemo Wambsganss

Roman Rietsche

Andrea Back

Published in the Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS), 2020.

Abstract

Professional sports has evolved from a game to an organization that has been codified, strategized, and commercialized. One factor that is shaping the sports industry is the pervasiveness of social media. On the one hand, social media is used as a powerful medium for distributing and getting news, engaging in topical discussions, and empowering brands. On the other hand, social media has become a crucial mouthpiece for athletes to interact with peers, share opinions, thoughts, and feelings. However, millions of followers, tweets, and likes later, researchers, practitioners, and athletes alike ask whether social media has an impact on an athlete's performance. We conducted a social media usage and a sentiment analysis of 124,341 Twitter tweets extracted from 31 tennis athletes. We linked these data to 8,095 corresponding match day performances. The results show that high social media usage has a significant negative impact on athletes' performance.

Introduction

In 1975, pupillometry pioneer and psychologist Eckhard Hess conducted a study to investigate the idea that changes in attitude can be detected by measuring changes in pupil size. Hess found that pupils are sensitive indicators of the mental state. This experiment later became known as the “window to your soul experiment”, describing the fact that the best way to read someone's mind is to look into their eyes (Hess, 1975). Today, personal contact to analyze peoples' behaviors and mental states is not necessary anymore, since people are sharing their opinions, thoughts, and feelings on social media platforms, such as Facebook, Instagram, and Twitter. For instance, about 2.1 million Snaps per minute are recorded by Snapchat users and 300 million tweets are tweeted on Twitter each day (Leetaru, 2019; Statista, 2019b). As a result, what used to be the exclusive domains of friends and family members can now be accessed and analyzed by the public. 3.02 billion people, or about 40% of the world's population, will use social media platforms in 2021 (Statista, 2019a).

One industry that has been strongly shaped by the use of social media in recent years is the sports industry (Davenport, 2014; Gruettner, 2019; Xiao et al., 2017). On the one hand, clubs, fans, and sponsors are using social media as a powerful medium for distributing and getting the latest sports news, engaging in topical discussions and empowering brands. For example, the German sporting goods manufacturer Adidas already uses 90 percent of its marketing budget for digital campaigns and social media (Klingelhöfer, 2018). On the other hand, social media has also become a crucial

mouthpiece for professional athletes to interact with stakeholders and peers (e.g., fans, sponsors, and celebrities) and, in turn, to build up a personal brand beyond the actual field of play. Consequently, almost every athlete has a social media profile in these days. It is even deemed as a necessity for athletes to maintain a social media presence (Tan et al., 2017; Xiao et al., 2017). However, millions of followers, tweets, and likes later, with social media now at the heart of so many supposed sports scandals and athletes subject to vicious attacks that compromise their mental well-being, researchers, practitioners, and even athletes alike ask whether social media has an impact on athletes' match day performance. For instance, National Basketball Association (NBA) star Stephen Curry just recently challenged the impact of social media in an interview conducted by CNN: "This is a new era in terms of, you know, the spotlight that every NBA athlete and any athlete in general is under. In terms of social media, in terms of the expectations put on us every single day" (Cable News Network [CNN], 2019, 1:24).

Researchers have demonstrated in various theoretical as well as empirical contexts that social media can negatively impact user behavior (see for a literature review on social media's impact on user behavior Kapoor et al. (2018)). For instance, pervasive social media usage can result in technology dependency and excessive usage, which in turn can lead to negative outcomes for its users (Lowry et al., 2016; Turel & Qahri-Saremi, 2016). However, only a few studies (e.g., Xu and Yu (2015)) have looked at the impact of social media on athletes' performance. Therefore, several researchers and practitioners call for studies that investigate how social media impacts athletes' performance (e.g., DiMoro (2015), Xiao et al. (2017), or Xu and Yu (2015)). An understanding of the impact of social media on athletes' performance would not only be interesting for coaches, team managers, scouts, and bookmakers within the sports industry but also for researchers that investigate how information technology (IT) usage impacts human behavior. Consequently, this paper addresses the following two research questions (RQs):

RQ1: *How does Twitter usage impact athletes' match day performance?*

RQ2: *How does athletes' pre-game mood extracted by Twitter sentiment polarity scores impact match day performance?*

Our research answers these research questions with the help of two evaluations: Firstly, by investigating the impact of low and high Twitter usage on athletes' match day performance and, secondly, by analyzing the Twitter sentiment polarity, which we claim is a proxy for athletes' pre-game mood. Specifically, we evaluate the performance of 31 professional tennis athletes. Our data set comprises 124,341 tweets. These social media

data are linked to 8,095 corresponding match day performances. In future research, the proposed methodological approach consisting of our hypotheses, data set, and analytics techniques should be applied and extended to different sports disciplines.

The remainder of this paper is structured as follows: Section 2 introduces the theoretical background and related work. Our hypotheses are stated in Section 3. The methodology is described in Section 4. Section 5 presents our findings. We then discuss our findings and avenues for future research as well as limitations in Section 6. Finally, we conclude our paper in Section 7.

Theoretical Background

The Role of Social Media in Sports

Although social media has received significant attention from researchers of various fields in the last years, including Information Systems (IS), there is no common definition of social media in academia yet (Kapoor et al., 2018). We define social media as being made up of various user-driven platforms (e.g., Facebook, WhatsApp, or Twitter) that facilitate diffusion of compelling content, dialogue, creation, and communication to a broader audience. Social media provides an environment that is conducive for interactions and networking to occur at different levels (e.g., personal, professional, business, marketing, political, and societal) (Kapoor et al., 2018). Social media offers the possibility to share user-generated content (UGC). UGC describes the various forms of digital content (e.g., posts, pictures, videos, or audio) created by end-users outside a professional context and which is publicly available (Brooks, 2015). A well-known example of UGC created by an athlete – which we refer to as athlete-generated content in this paper – is when the Italian soccer athlete Francesco Totti scored a goal in the so-called Roma derby, as part of the celebration took a selfie with the emotional, flag-waving supporters in the background, and posted the photo on Facebook. However, as the seemingly unbreakable boundary between the physical world and the online world in sports is getting blurry, it can be asked whether athletes' social media activity has an impact on the performance on the field (DiMoro, 2015; Xiao et al., 2017; Xu & Yu, 2015). On this matter, NBA commissioner Adam Silver believes that social media is doing more harm than good. Silver said during the MIT Sloan Sports Analytics Conference: “I think we live a bit in the age of anxiety. ... I think part of it is a direct product of social media” (Cable News Network, 2019, 0:46).

Analytics and predictive techniques to investigate athletes' performances have commonly been used in the sports industry since the seminal work by Michael Lewis on the legendary story of Oakland Athletics, a Baseball Team that demonstrated how robust technical analysis based on performance data outperformed intuition and old-school wisdom (Davenport, 2014). Latest developments in digital technologies that led to the availability (and abundance) of statistical data have further stimulated interests in this area, which is commonly referred to as sports analytics (Xiao et al., 2017). However, the impact of athletes' social media activity on an athlete's performance has only received little attention in literature (DiMoro, 2015; Xiao et al., 2017; Xu & Yu, 2015). Although, for instance, with its unique features and its large user base, Twitter offers a rich data pool of athletes' communications, their opinions, their thoughts, and their feelings. Social media usage (e.g., the amount of Twitter tweets) and specific online athlete-generated content have great potential to become an information source for not only coaches, team managers, and scouts but also for bookmakers to discern athletes' mood status and shaky performance before matches. One of the few existing studies which investigated the impact of social media on athletes' performance is published by Xu and Yu (2015). In this study, a sentiment analysis of athlete-generated content on Twitter was used to detect NBA athletes' pre-match moods. The authors found that the mood of athletes has significant impact on driving athletes' on-court performances. Similarly, Jones et al. demonstrated in their study that late-night tweeting is associated with a decreased next-day match performance for NBA athletes (Jones et al., 2019). In our paper, we adopted and adjusted both studies to a professional tennis context. Therefore, we extended both studies by (1) investigating how low and high pre-match Twitter usage 36 hours prior to a match impacts athletes' match day performance and by (2) utilizing the Valence Aware Dictionary and Sentiment Reasoner (Vader) lexicon for our sentiment analysis, which is specifically designed for sentiments expressed in social media (Hutto & Gilbert, 2014).

Linking Social Media and Performance

A major controversial topic of social media, which is currently not only discussed in the sports industry but also in political and entrepreneurial discussions, is its specific impact on users' performances. A topical example is the debate about the potentially negative impact of social media usage on employee productivity. As a consequence of this concern, many companies have banned social media platforms like Facebook and Twitter from the workplace (Bizzi, 2017). In this paper, we define performance as the effectiveness with which people (i.e., athletes) perform activities that contribute to

predefined goals and outcomes (i.e., sporting success) (Borman & Motowidlo, 1997). Existing research on the impact of social media on users' performances is two-fold: While a couple of studies have revealed the positive side of social media (e.g., Dienlin et al. (2017) or Utz and Breuer (2017)) – for instance, the usage of LinkedIn can result in knowledge sharing and dissemination, or YouTube can be used for training and recruitment (Kapoor et al., 2018) – social media's negative impact on performance is evident in the majority of studies. These studies commonly report that pervasive social media usage can, for instance, result in technology dependency and excessive usage, which in turn can lead to negative outcomes for users' performances (Brooks, 2015; Jones et al., 2019; Mansi & Levy, 2013). For example, too much communication via social media can cause information overload and distraction, confuse user's focus, and hamper the ability to make decisions (Mansi & Levy, 2013).

In a sports context – as a practical example to illustrate how social media already impacts athletes' behaviors – Kliff Kingsbury, the head coach of the National Football League (NFL) club Arizona Cardinals, recently mandated "cell phone breaks" to browse social media during team meetings. It was created as a way to help those athletes who may lose focus during meetings itching to get back on their mobile device: "You start to see kind of hands twitching and legs shaking and you know they need to get that social media fix, so we will let them hop over there and then get back in the meeting to refocus" (York, 2019, 5:06). In this sense, the stress and pressure felt by today's athletes is arguably greater than ever. Contracts are bigger, stakes are higher. More people are watching, and, with social media, athletes are more visible than ever. Thus, not only the mere usage of social media can be a distraction for athletes. In particular, the content of others relating to the athlete may affect athletes' performances once read by the athlete.

Distraction-Conflict Theory

In order to explain social media's negative impact on users' performances, researchers commonly use distraction-conflict theory (Baron, 1986). Distraction-conflict theory explains how distraction can create an attentional conflict or a situation in which an individual feels the tendency, desire, or obligation to allocate attention to multiple exclusive inputs (Baron, 1986; Baron et al., 1978). This, in turn, can lead to performance inefficiency. The distraction-conflict model can be broken down into three steps (Baron, 1986; Baron et al., 1978): (1) distraction, (2) distraction leads to attentional conflict, and (3) attentional conflict impacts the performance and motor behavior. Brooks, for example, tested the impact of social media usage on students' task performances by creating a classroom task environment (Brooks, 2015). Brooks (2015) found that higher

amounts of social media usage led to lower performance on the task, as well as higher levels of technostress and lower happiness. Similarly, Cao and Yu (2019) investigated the effects of social media's different usage patterns on employee job performance. Their results also showed that social media usage can generate conflict between technology use and work demand. This, in turn, negatively influences job performance.

Furthermore, distraction-conflict theory assumes that in settings in which there is pressure to perform a task quickly and well, attentional conflict can be created not only by external distractions but also by internal distractions, i.e., cognitive activity that is not directly relevant to the task solution but has an impact on it (Baron, 1986). In the context of this paper, such internal distracting cognitions can include, for instance, the content of others read by an athlete on social media in advance of the match but which comes to the athlete's mind during the game. For example, a post from a user who claims that the athlete will never win a trophy because he is not cool enough in tight situations.

Hypothesis Development

Development of Hypothesis 1

Several studies have demonstrated in different contexts (e.g., Brooks, 2015 or Mansi and Levy (2013)) that increased social media usage has a negative impact on performance (see, Section 2.2.). Although the performance variable in a sports context differs from the examined contexts in these studies, we applied distraction-conflict theory to the context of our paper for the following reasons:

Firstly, studies have proven that social media usage can lead to distraction in both professional and non-professional settings (e.g., Brooks (2015), Cao and Yu (2019), or Mansi and Levy (2013)). We believe that high Twitter usage before a match leads to distraction for athletes due to two reasons: (1) The focus and time spent on posting a tweet. As a consequence, the necessary focus on training, tactical preparation for the game, or even enough sleep is missing. (2) Athletes' awareness of the Twitter users' responses to their previously posted tweets and, in particular, the content of others relating to the athlete may trigger internal distractions (see, Section 2.3.) and thus affects athletes' match day performances. Secondly, prior work (e.g., Brooks (2015) or Mansi and Levy (2013)) has utilized the distraction-conflict theory to investigate the impact of social media on cognitive tasks. However, distraction-conflict theory claims validity for both cognitive and motor tasks (Baron, 1986). In a sports context, for instance, tennis strokes (e.g., a first serve) are understood as motor tasks (Wolpert & Flanagan, 2010).

We therefore see the distraction-conflict theory as an appropriate theoretical outline for the athletic setting.

Consequently, one should expect that high social media usage, expressed in the number of pre-match tweets, is associated with decreasing match day performance. This leads to the following hypothesis:

H1: *High pre-match social media usage has a negative impact on athletes' match day performance.*

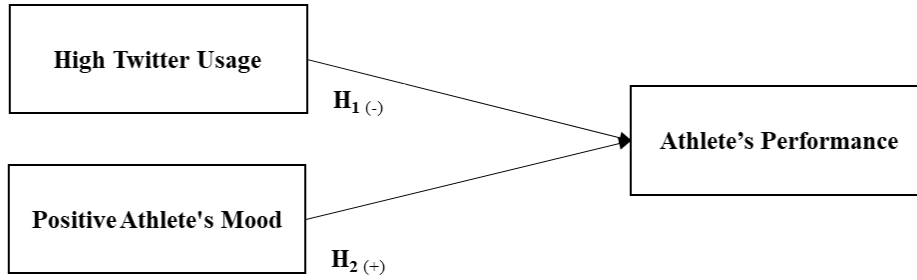
Development of Hypothesis 2

We extended our research by taking the athlete-generated content of the tweets into account. Several studies have shown that tweets can be used to extract the collective (Bollen et al., 2011) or individual mood (e.g., Golder and Macy (2011) or Xu and Yu (2015)). For instance, Golder and Macy (2011) applied sentiment analysis to Twitter data to extract emotional content from 509 million tweets of 2.4 million users in 84 different countries. Using this information, the authors were able to identify daily and seasonal individual mood variations in cultures across the world.

Extrapolating from this, athlete-generated content can be used as a proxy to determine athletes' pre-match mood which, in turn, can impact athletes' performance on the field. In this sense, Totterdell (2000) analyzed the relationship between mood and performance for cricket athletes in an offline context. The results indicated a correlation between individual performance and the collective mood of teammates. Likewise, Xu and Yu (2015) found a positive association between NBA athletes' individual mood, which has been extracted from pre-match tweets, and their match performance. As our focus also lies on the individual athlete's performance, we expect similar results in our setting. This results in the following hypothesis:

H2: *A positive pre-match mood has a positive impact on athletes' match day performance.*

Our research model is depicted in Figure 1.

Figure 1*Research Model*

Methodology

In this paper, we used social media and performance data of professional tennis athletes to investigate (1) the impact of low and high social media usage on athletes' performance and to analyze (2) whether we can use sentiment polarity scores to detect athletes' pre-match mood, which in turn has an impact on athletes' match day performance. We chose professional tennis athletes for our analysis for two reasons: Firstly, due to the availability of very high-quality performance data recorded at professional sports competitions as well as huge amounts of available social media data and, secondly, due to the nature of tennis as a single sport, which guarantees that no external team factors affect individual performance. To validate our hypotheses, we followed an adjusted knowledge discovery process in databases (KDD) as described by Fayyad et al. (1996). KDD describes the structured extraction of useful information from a large volume of data. For the purpose of our paper, KDD's iterative cycle approach was combined and formed into three steps: (1) data selection, (2) data preprocessing and transformation, and (3) data mining and evaluation, which are explained and applied in the following subsections.

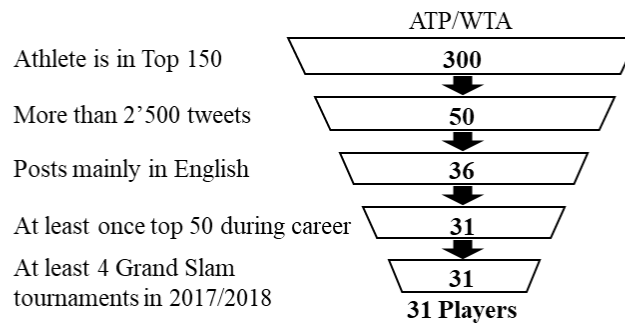
Data Selection

The first step (data selection) of our adjusted KDD included the selection of data sources. In this paper, we aimed to acquire Twitter tweets from athletes participating in the Association of Tennis Professionals (ATP) and the Women's Tennis Association (WTA). We chose Twitter for our study as athletes are able to display behavioral traits on Twitter (Xu & Yu, 2015). We used the python-based library Twint to collect the entire Twitter timeline of a given athlete (see, Figure 2 for the include/exclude criteria for the selection of athletes). To narrow down the selection of athletes, in a first step, we

started with the current Top 150 ATP and WTA athletes (as of April 19, 2019). We then included and excluded certain athletes based on chosen criteria, e.g., athletes which posted less than 2,500 tweets, resulting in 31 athletes for further analysis. In total, 124,341 tweets from 31 tennis athletes were collected, with 12 being ATP (i.e., male) athletes and 19 being WTA (i.e., female) athletes. The match performance data for each examined athlete was collected by using open source data sets from tennis-data.co.uk, tennis.wettpoint.com and a GitHub data repository (Sackmann, 2019; Tennis-data.co.uk, 2019; Wettpoint.com, 2019). Our final data set for our analysis is publicly available and can be accessed via the link below.¹

Figure 2

Include and Exclude Criteria for Athlete Selection (as of April 19, 2019)



To test the developed hypotheses H1 and H2, we generated two distinct data sets for each athlete – (1) a Twitter data set and (2) a match performance data set. The Twitter data set contained the full tweet text in UTF-8 encoding with emoticons being displayed. Additionally, metadata such as the post time and the username were included. We excluded retweets from our analysis as retweets did not contain useful information for our analysis. The match performance data set included all ATP and WTA tournament singles matches between 2012 and 2018 from the chosen athletes. For an overview of ATP/WTA tournaments, see, Tennis-data.co.uk (2019). In total, the match performance data set counted 9'340 matches before preprocessing and cleaning the data. We included metadata such as the result, the match date and time, the age as well as the rank of the athlete in our data set. As a performance variable, we chose the first serve fault. The first serve fault indicates the percentage of serves missed and consequently leads to a second serve for a given athlete. We chose the first serve fault as our performance indicator, as this metric is less dependent on the opponent's strength than aces, unforced errors, or

¹ See <https://www.kaggle.com/appliedresearcher/social-media-the-new-window-to-athletes-soul>, for more information.

break points saved. Thus, the first serve fault displays an athlete's individual primary task performance and can be theoretically linked to the distraction-conflict theory.

Data Preprocessing and Transformation

After extracting and collecting relevant data, the next step was data preparation. We removed three outliers which posted more than 50 tweets before the match as well as matches with incomplete performance data. In total, we deleted 1,245 matches from our sample. Resulting in a total of 8,095 matches. Then, the preprocessed data got transformed into a structured format that can be accessed and processed for further analysis. We tested H1 by conducting a paired-samples t-test between two groups of social media usage (low and high). To test H2, we first performed a sentiment analysis, then we clustered the tweets into two mood groups (non-positive and positive mood) in accordance to their polarity score as this study focuses only on the impact of a positive mood of athletes on their performance. This step was followed by a paired-samples t-test as well. Thus, the data was preprocessed and transformed to meet the purposes of these two analyses. Preprocessing and data analytics techniques were performed using the programming language Python 2.73, since it is widely known, easy to use, and supports major libraries for Natural Language Processing (NLP) tasks (Bird et al., 2009).

First, the extracted 124,341 tweets were imported into a Pandas data frame. Then, the text of our tweets was transformed into lower case. Regular expressions were used to filter punctuation (e.g., “?” - except “!”) and numbers, since they were not relevant for the sentiment of a tweet. Furthermore, hyperlinks, mentions of other Twitter users and hashtags were removed and substituted with an acronym (e.g., “www.twitter.com/...” was substituted with “URL”, or “@UserXY” was substituted with “AT_USER”). However, after applying further libraries we decided to not tokenize and stem our tweets, since the applied sentiment library performed with more accurate results with the full strings.

Data Mining and Evaluation

Evaluation of Hypothesis 1

The next step of our adjusted KDD included the data mining and evaluation phase. To test our hypothesis, we linked the Twitter data set to the generated performance data set. We defined a time span of 36 hours prior to a match as the pre-match time span in accordance with (Xu & Yu, 2015), who tested the causality relation between different

time spans and performance. Therefore, we used the time stamp of the Twitter tweets as well as the concrete start time of the match. Differences due to time shifts were taken into account. Finally, 13,954 tweets out of the totally extracted 124,341 tweets were linked to 8,095 corresponding match day performances. We believe that a higher number of pre-match tweets within this 36-hour time span represents a longer time spent on Twitter. This, in turn, is expected to distract athletes in accordance with the distraction-conflict theory (see, Section 3.1.). If an athlete started her/his career or signed up for Twitter after 2012, the aggregation process started at the earliest date possible. The same methodology was also applied for the sentiment analysis to test H2.

Finally, the impact of high Twitter usage on athletes' performance (H1) was tested by carrying out a two-tailed paired-samples t-test between two groups of social media usage (i.e., low and high). In detail, each tennis athlete's mean performance was compared between matches with low Twitter usage and high Twitter usage. We labeled a data point as low Twitter usage if the athlete posted between zero and four tweets before a match day and as high Twitter usage if the pre-match tweets exceeded four tweets. This classification resulted from calculating the average pre-match Twitter usage and adding one standard deviation to the value in order to define the threshold for high Twitter usage (see, Table 2.).

Evaluation of Hypothesis 2

To test H2, we aimed to analyze the Twitter sentiment polarity, which we claim is a proxy for the athletes' pre-game mood. Therefore, we used a dictionary-based method from Natural Language Toolkit (NLTK), that is, the Vader lexicon (Hutto & Gilbert, 2014), to extract sentiments from our tweets. Based on this sentiment analysis, polarity scores were retrieved that correspond to athletes' pre-game mood. Scores were labeled between -1 and 1 according to a "positive", a "negative", and a "neutral" mood. A tweet which had a polarity score above zero was classified as "positive". Tweets which showed a score below zero were classified as "negative". All other tweets that had a score that equaled exactly zero were labelled with "neutral". For an example of Vader's polarity score and the resulting pre-game mood, see, Table 1.

We defined a 36-hour time span prior to a match day as our investigation period to investigate the impact of athletes' mood. The length of this time span ensured that the pre-match mood of a specific match did not conflict with the mood of the precedent match as tennis athletes usually play in a two- to three-day rhythm. In contrast to the evaluation for H1, we deleted all matches in which an athlete did not post a tweet from

our data set for further analysis, as these data points did not contain any information on athletes' mood. As a result, the number of matches to test H2 was reduced from 8,095 matches to 4,686 matches. In a last step, we aggregated all polarity scores occurring in the defined time span for the corresponding match day and calculated the mean polarity score. Afterwards, we merged all pre-match modes which were labeled as "negative" or "neutral" according to the results of the sentiment analysis into one group, that is, the "non-positive" pre-match mood group. In a last step, to test whether a positive pre-match mood has a positive impact on athletes' match day performance, we applied a two-tailed paired-samples t-test.

Table 1

Sample Tweets Showing Vader Mood Classification

Tweet Text	Tennis Athlete	Polarity Score	Mood/Label
i don't believe in luck... but if i did mine would be categorized as bad and terrible.	Serena Williams	-.8737	Negative
just woke up and preparing my breakfast now...	Julia Goerges	0	Neutral
happy, happy, happy, happy .. i think i'm happy qualified for the maindraw.. damn feels good thanks for your support	Kirsten Flipkens	.9758	Positive

Results

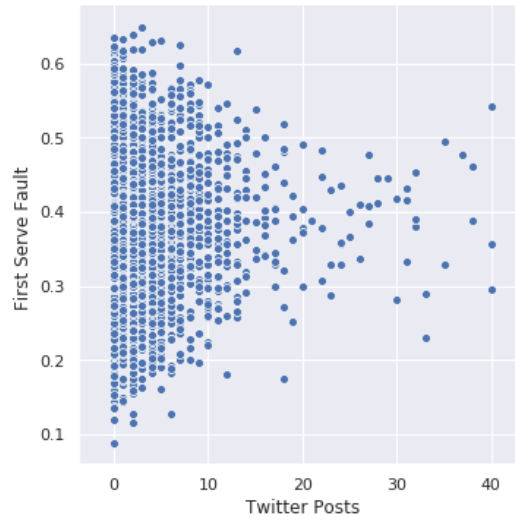
Descriptive Statistical Results

The final data set to test H1 comprised 13,682 tweets, which were linked to 8,095 corresponding match day performances for the predefined cut-off time of 36 hours. To test H2, these 8,095 match day performances were reduced to 4,686 data points due to matches in which athletes did not post a tweet as described in Section 4. For an overview of the descriptive statistics of our data sample, see, Table 2. For an overview of the Twitter usage with the corresponding match performance, see, Figure 3. The relation between athletes' mood and performance is depicted in Figure 4.

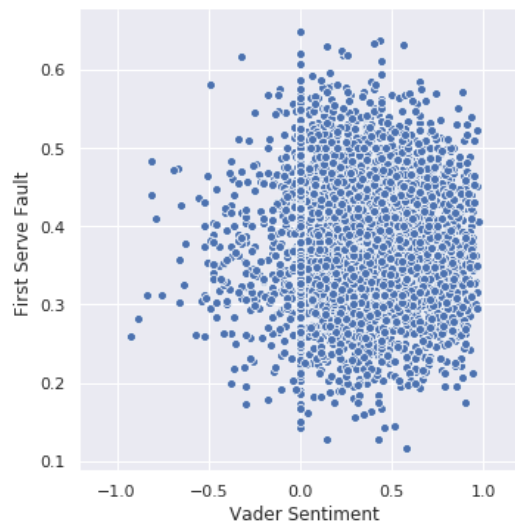
Table 2*Descriptive Statistics for Social Media Usage, Sentiment Analysis, and Performance*

Label	Social Media Usage		Sentiment Analysis		Performance	
	Low Usage	High Usage	Non-Positive Mood	Positive Mood	First Serve Fault	
Number of Matches	7,363	732	1,332	3,354	8,095	
Total Matches	8,095		4,686		8,095	
Number of Tweets	13,682*		13,954		-	
Demographics	Mean	SD	Mean	SD	Mean	SD
Total	1.690	3.117	.296	.306	.379	.078
Gender Female (n = 19)	1.511	2.477	.294	.309	.384	.081
Male (n = 12)	1.893	3.702	.298	.302	.373	.075
Age Experienced	1.610	3.276	.284	.307	.378	.078
Inexperienced	1.914	2.613	.322	.301	.386	.079
*As described in Section 4.2., for the social media usage analysis the number of tweets is reduced due to the removal of outliers.						
SD = Standard deviation						

On average, the identified athletes posted 1.69 tweets per match ($\sigma = 3.12$). Male athletes posted 1.89 pre-match tweets per match on average ($\sigma = 3.70$), whereas female athletes posted 1.51 tweets per match on average ($\sigma = 2.48$). The highest number of pre-match tweets posted by a male athlete was 40 (two posts) tweeted by Stanislas Wawrinka. Likewise, the highest number of pre-match tweets posted by a female athlete, that is Serena Williams, was 40. Taking different age groups into account (i.e., athlete's age on match day) and defining experienced athletes as any athlete over 24 years of age, we identified a slight difference between experienced and inexperienced athletes. On average, experienced athletes posted 1.61 pre-match tweets per match ($\sigma = 3.28$), while inexperienced athletes posted 1.91 pre-match tweets per match ($\sigma = 2.61$). Looking at the performance data, the average for the first serve fault for male and female athletes was 0.38 ($\sigma = 0.08$). On average, the first serve fault for male athletes was 0.37 ($\sigma = 0.08$), while female athletes' first serve failed 0.38 on average ($\sigma = 0.08$).

Figure 3*Twitter Usage and Performance*

As described above, the data set to test H2, including our sentiment analysis, comprised 4'686 matches. The average polarity score for the identified athletes counted 0.30 ($\sigma = 0.31$), which indicated that most tweets were associated with a positive athletes' mood. Splitting the data set into male and female athletes, male athletes average a polarity score of 0.30 ($\sigma = 0.30$), whereas female athletes average a polarity score of 0.29 ($\sigma = 0.31$). Comparing experienced and inexperienced athletes in the same way as described above, experienced athletes expressed a polarity score of 0.28 ($\sigma = 0.31$) and inexperienced athletes scored 0.32 on average ($\sigma = 0.30$), indicating that inexperienced athletes tweeted more positively than experienced athletes.

Figure 4*Pre-Game Mood and Performance*

Impact of High Social Media Usage on Performance

H1 states that high social media usage (i.e., an athlete posted five or more tweets before a match day) has a negative impact on athletes' match day performance. The performed two-tailed paired-samples t-test between the low Twitter usage group and high Twitter usage group showed that the latter had a negative impact on athletes' match day performance in contrast to low Twitter usage. This difference was significant at $p < 0.05$. On average, tennis athletes performed 1.011 percentage points worse in terms of delivering the first serve when having high Twitter usage before a match. Consequently, as a result of our analysis, H1 can be validated. Table 3 summarizes the results from the two-tailed paired-samples t-test.

Table 3

The Impact of High Social Media Usage on Match Performance

Variable	Low Twitter Usage	High Twitter Usage	Difference	t
First Serve Fault	.3812	.3922	-.0101	-2.270*
Significance with * = $p < .05$ / ** = $p < .01$ / *** = $p < .001$				

Impact of Athletes' Mood on Performance

To test whether a positive pre-match mood has a positive impact on athletes' match day performance (H2), we also applied a two-tailed paired-samples t-test between the non-positive mood group and the positive athlete mood group. The results of our analysis showed no significant difference ($p > 0.05$). We repeated the same analysis for different polarity score thresholds, e.g., classifying polarity scores above 0.25 as positive mood, but no significant results were found. We believe that state-of-the-art sentiment lexica such as Vader still need to be trained and refined in order to deliver strong results for detecting the individual mood of athletes. Therefore, H2 has to be rejected. Table 4 indicates the results from the two-tailed paired-samples t-test.

Table 4

The Impact of Athletes' Moods on Match Performance

Variable	Non-Positive Mood	Positive Mood	Difference	t
First Serve Fault	.3813	.3849	-.0036	-1.166
Significance with * = $p < .05$ / ** = $p < .01$ / *** = $p < .001$				

Discussion, Future Research, and Limitations

To investigate how social media usage (RQ1) and athletes' pre-game mood (RQ2) impact athletes' match day performance, we conducted a social media usage as well as a sentiment analysis of 31 tennis athletes and extracted 124,341 tweets. We linked these data to 8,095 corresponding match day performances. To test our hypothesis that high social media usage has a negative impact on athletes' performance (H1), we counted the number of tweets that were tweeted 36 hours before match day by an individual athlete and formed two groups of Twitter usage (low and high). The results of the two-tailed paired-samples t-test confirmed our hypothesis and demonstrated that high social media usage has a significant ($p < 0.05$) negative impact on athletes' performance. Hence, our results support the concerns and fears of practitioners, researchers, and athletes alike that social media usage can have a negative impact on athletes' performances. By conducting a sentiment analysis, we identified polarity scores for each tweet, which were used as indicators for athletes' pre-match mood to test our hypothesis that a positive athletes' pre-match mood has a positive impact on athletes' performance (H2). Although there are studies which have proven a positive relationship between athletes' pre-match mood and their on-field performance in offline and online contexts (e.g., Totterdell (2000) and Xu and Yu (2015)), our analysis showed no significant results ($p = 0.253$) that an athlete with a positive pre-match mood performed better on the field than an athlete who did not show a positive pre-match mood in her/his tweets. Therefore, we must reject H2.

From an academic perspective, our paper makes several contributions to the existing body of literature. By answering our first research question, we contribute to IS research by addressing the call made by several researchers for the investigation of how social media usage impacts athletes' performance (e.g., DiMoro (2015), Xiao et al. (2017), and Xu and Yu (2015)). While our results demonstrate that high social media usage can have a significant negative impact on athletes' performances in the context of this study, our research model should be applied to different contexts. It can be expected that depending on, for instance, the type of sports (i.e., individual vs. team sports) results will differ. Likewise, it should also be investigated whether there are individual athletes for whom social media usage has a positive impact. In addition, we contribute to the upcoming literature stream of sports analytics by providing an explanation of how distraction-conflict theory can be applied to explain the impact of social media usage on athletes' performance. Moreover, we contribute to the body of knowledge by applying distraction-conflict theory in a new setting, that is, athletes' performances. Most existing studies on the impact of social media usage on performance investigate how social

media usage distracts users' cognitive tasks at the concrete moment the task is actually performed (e.g., Brooks (2015) or Mansi and Levy (2013)). We enrich existing findings by investigating how internal distraction impacts performance, not only at the actual moment of activity but also over a longer time span (i.e., 36-hour time span) for motor tasks. We further contribute to research as well as practice by making our data set publicly available.

To answer our second research question, we conducted a sentiment analysis using the Vader lexicon. We think that the Vader lexicon is inadequate for the analysis of athlete-generated content, although it was developed specifically for the analysis of social media data, since athletes use a very specific language in their tweets. Social media offers a rich data pool of athletes' communications, their opinions, their thoughts, and their feelings. Hence, we believe it has great potential to become a relevant information source for future research. Therefore, our research contributes to the literature stream of sports analytics by representing one of the early efforts from the IS field to bring together athlete-generated content and analytics techniques to investigate athletes' pre-match mood. Our paper analyzes the impact of social media in a sports context. We argue that the sports industry provides many opportunities for comprehending IT-driven phenomena that might display interesting dynamics due to the uniqueness of the context at a theoretical level (Chiasson & Davidson, 2005; Gruettner, 2019; Xiao et al., 2017). Thus, dismissing the sports industry as just another empirical context will translate to missed opportunities. Therefore, this paper also contributes to the IS literature in general by examining how social media impacts human behavior.

From a practical perspective, our results provide important insights into the impact of social media on athletes' match day performance. Thus, they are especially relevant for the stakeholders of the sports industry. For instance, coaches can identify athletes who use social media excessively. In turn, they can adjust the line-up on match day accordingly. Similarly, team managers and scouts can use our results to determine which athlete they want to hire based on their social media usage and athlete-generated content. Likewise, bookmakers can adjust their calculation of odds based on the social media activity of athletes. Regarding this, the development of an interactive live dashboard that displays various key information to stakeholders could be interesting in the future. Finally, our results should also be seen as a wake-up call for athletes themselves, inspiring them to reduce their social media activities and to focus on their performance on match days.

This paper provides numerous opportunities for future research: For instance, researchers can investigate whether different types of social media activities (e.g.,

listening, posting, reading, or watching) have a different impact on athletes' performances. Similarly, it would be interesting to explore how various communications via social media (e.g., communications with peers vs. fans) impact athletes' match day performances. To extend our research, an analytical model can be built to predict athletes' performances. Furthermore, our research model can be extended to various other social media platforms such as Instagram using, for instance, image recognition. Future research can also focus on the development of a more specific and comprehensive lexicon which can be applied to the context of professional sports athletes.

This paper is not free from limitations: Firstly, our performance variable (i.e., first serve fault) is not completely independent from other factors, such as age, current score, match importance, or opponent. Secondly, we believe that the lack of control variables (such as post length) is a limitation. Thirdly, we only investigated the negative impact of social media usage. It should also be investigated whether there are athletes on whom social media has a positive effect. Fourthly, it is reasonable to assume that the distilled moods from athletes' tweets are not free of bias, which could be attributed to a number of factors. For instance, professional athletes often use agencies to maintain their social media profiles in these days. Lastly, our data set consists of 31 athletes only. Thus, an extension of this study's scope can become relevant.

Conclusion

Based on a social media usage and a sentiment analysis, we investigated the impact of (1) low and high social media usage and of (2) athletes' pre-match mood on professional athletes' match day performance. For research, the conducted analysis provides theoretical evidence that high social media usage has a significant negative impact on athletes' performance. Future research should focus on different types of social media activities (e.g., posting or reading) or on various types of communications (e.g., communications with peers vs. fans) and their specific impact on athletes' performance. Implications for practitioners are insights into the impact of social media on athletes' performances. Therefore, supporting them in specific decision-making processes (e.g., line-up adjustments on match days).

References

- Baron, R. S. (1986). Distraction-Conflict Theory: Progress and Problems. *Advances in Experimental Social Psychology*, 19, 1–40. [https://doi.org/10.1016/S0065-2601\(08\)60211-7](https://doi.org/10.1016/S0065-2601(08)60211-7)

- Baron, R. S., Moore, D., & Sanders, G. S. (1978). Distraction as a source of drive in social facilitation research. *Journal of Personality and Social Psychology*, 36(8), 816–824. <https://doi.org/10.1037//0022-3514.36.8.816>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc. <https://doi.org/10.1097/00004770-200204000-00018>
- Bizzi, L. (2017). Should HR managers allow employees to use social media at work? Behavioral and motivational outcomes of employee blogging. *International Journal of Human Resource Management*. <https://doi.org/10.1080/09585192.2017.1402359>
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Borman, W. C., & Motowidlo, S. J. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research. *Human Performance*, 10(2), 99–109. https://doi.org/10.1207/s15327043hup1002_3
- Brooks, S. (2015). Does personal social media usage affect efficiency and well-being? *Computers in Human Behavior*, 46, 26–37. <https://doi.org/10.1016/j.chb.2014.12.053>
- Cable News Network. (2019). *Is social media harming sports stars?* <https://edition.cnn.com/videos/sports/2019/04/23/nba-nfl-social-media-tiffany-jones-adam-silver-steph-curry-spt-intl.cnn/video/playlists/international-sport-playlist-general-videos/>
- Cao, X., & Yu, L. (2019). Exploring the influence of excessive social media use at work: A three-dimension usage perspective. *International Journal of Information Management*, 46, 83–92. <https://doi.org/10.1016/j.ijinfomgt.2018.11.019>
- Chiasson, M. W., & Davidson, E. (2005). Taking Industry Seriously in Information Systems Research. *MIS Quarterly*, 29(4), 591–606. <https://doi.org/10.2307/25148701>
- Davenport, T. H. (2014). What Businesses Can Learn From Sports Analytics. *MIT Sloan Management Review*, 10–13.
- Dienlin, T., Masur, P. K., & Trepte, S. (2017). Reinforcement or Displacement? The Reciprocity of FtF, IM, and SNS Communication and Their Effects on Loneliness and Life Satisfaction. *Journal of Computer-Mediated Communication*, 22(2), 71–87. <https://doi.org/10.1111/jcc4.12183>
- DiMoro, A. (2015). *The Growing Impact of Social Media On Today's Sports Culture*. <https://www.forbes.com/sites/anthonydimoro/2015/07/02/the-growing-impact-of-social-media-on-todays-sports-culture/#69a6b80f31a2>

- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceedings of the 2nd International Conference on Knowledge Discovery And Data Mining (KDD) 1996*, 82–88.
- Golder, S. A., & Macy, M. W. (2011). Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, 333(6051), 1878–1881. <https://doi.org/10.1126/science.1202775>
- Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. *Proceedings of the 25th Americas Conference on Information Systems (AMCIS) 2019*.
- Hess, E. H. (1975). The Role of Pupil Size in Communication. *Scientific American*, 233(5), 110–119. <http://www.jstor.org/stable/24949943>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM) 2014*.
- Jones, J. J., Kirschen, G. W., Kancharla, S., & Hale, L. (2019). Association between late-night tweeting and next-day game performance among professional basketball players. *Sleep Health*, 5(1), 68–71. <https://doi.org/10.1016/j.sleh.2018.09.005>
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in Social Media Research: Past, Present and Future. *Information Systems Frontiers*, 20(3), 531–558. <https://doi.org/10.1007/s10796-017-9810-y>
- Klingelhöfer, C. (2018). *Adidas Goes After Nike with Digital Strategy*. <https://www.ispo.com/en/companies/adidas-goes-after-nike-digital-strategy>
- Leetaru, K. (2019). *A Fading Twitter Changes Its User Metrics Once Again*. <https://www.forbes.com/sites/kalevleetaru/2019/04/23/a-fading-twitter-changes-its-user-metrics-once-again/#316b64267a31>
- Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, 27(4), 962–986. <https://doi.org/10.1287/isre.2016.0671>
- Mansi, G., & Levy, Y. (2013). Do instant messaging interruptions help or hinder knowledge workers' task performance? *International Journal of Information Management*, 33(3), 591–596. <https://doi.org/10.1016/j.ijinfomgt.2013.01.011>
- Sackmann, J. (2019). *Tennis ATP*. GitHub Repository. <https://github.com/JeffSackmann/>

- Statista. (2019a). *Number of social media users worldwide 2010-2021*. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- Statista. (2019b). *User-generated internet content per minute 2019*. <https://www.statista.com/statistics/195140/new-user-generated-content-uploaded-by-users-per-minute/>
- Tan, F. T. C., Hedman, J., & Xiao, X. (2017). Beyond ‘Moneyball’ to Analytics Leadership in Sports: An Ecological Analysis of FC Bayern Munich’s Digital Transformation. *Proceedings of the 23rd Americas Conference on Information Systems (AMCIS) 2017*.
- Tennis-data.co.uk. (2019). *Tennis-data.co.uk*. <http://www.tennis-data.co.uk/alldata.php>
- Totterdell, P. (2000). Catching Moods and Hitting Runs: Mood Linkage and Subjective Performance in Professional Sport Teams. *Journal of Applied Psychology*, 85(6), 848–859. <https://doi.org/10.1037/0021-9010.85.6.848>
- Turel, O., & Qahri-Saremi, H. (2016). Problematic Use of Social Networking Sites: Antecedents and Consequence from a Dual-System Theory Perspective. *Journal of Management Information Systems*, 33(4), 1087–1116. <https://doi.org/10.1080/07421222.2016.1267529>
- Utz, S., & Breuer, J. (2017). The Relationship Between Use of Social Network Sites, Online Social Support, and Well-Being: Results From a Six-Wave Longitudinal Study. *Journal of Media Psychology*, 29(3), 115–125. <https://doi.org/10.1027/1864-1105/a000222>
- Wettpoint.com. (2019). *Tennis Statistics Wettpoint*. <https://tennis.wettpoint.com/en/>
- Wolpert, D. M., & Flanagan, J. R. (2010). Motor learning. *Current Biology*, 20(11), 467–472. <https://doi.org/10.1016/j.cub.2010.04.035>
- Xiao, X., Hedman, J., Ter, F., Tan, C., Tan, C.-W., Lim, E. T. K., Clemenson, T., Henningsson, S., Vatrappu, R., Rao Mukkamala, R., & Van Hillegersberg, J. (2017). Sports Digitalization: An Overview and A Research Agenda. *Proceedings of the 38th International Conference on Information Systems (ICIS) 2017*.
- Xu, C., & Yu, Y. (2015). Measuring NBA Players’ Mood by Mining Athlete-Generated Content. *Proceedings of the 48th Annual Hawaii International Conference on System Sciences (HICISS) 2015*. <https://doi.org/10.1109/HICSS.2015.205>
- York, J. (2019). *Is social media making this generation of sports stars the “loneliest” ever?* <https://edition.cnn.com/2019/04/23/sport/nba-social-media-tiffany-jones-spt-intl/index.html>

Publication III

From Data to Dollar –

Using the Wisdom of an Online Tipster Community to Improve Sports Betting Returns



Author(s)

Arne Grüttner

Thiemo Wambsganss

Andrea Back

In press for publication in the European Journal of International Management (EJIM) –
Special Issue on: "International Sports Management".

Abstract

With thousands of (online) bookmakers accepting wagers on sporting events, sports betting has become a billion-dollar business worldwide. Therefore, researchers and practitioners have gathered interest in investigating the “wisdom-of-crowds” effect in online tipster communities to predict the outcomes of sports events. We extracted 1,534,041 tips of 3,484 tipsters from Blogabet.com and used this user-generated content to investigate whether there is wisdom in online tipster communities that can be used to improve betting returns. We applied state-of-the-art data mining and natural language processing techniques and tested our hypotheses using quantitative research methods. Our results demonstrate that there is indeed wisdom in such online tipster communities that can improve sports betting returns. Tipsters won 3.29% more tips than the implied win probability set by bookmakers and produced averaged yields of 3.97%. We further identified four characteristics that are significant indicators for smarter sub-crowds within the overall crowd of an online tipster community.

Keywords: data mining; natural language processing; online communities; sports betting; sports innovation; user-generated content; wisdom-of-crowds

Introduction

Sports betting has become a billion-dollar business around the globe. In Germany, for example, each day people place bets worth EUR 25 million. An increase of 21 percent over the previous year (Deutscher Sportwettenverband, 2020). Online tipster communities, such as Betadvisor.com, Blogabet.com, or Oddsportal.com, offer semi-professional sports bettors, so-called tipsters, the opportunity to publish, share and explain their carefully elaborated tips over the internet. Community members, on the other hand, can comment and discuss those publicly available tip recommendations. Online tipster communities can be seen as a new type of sports-based entrepreneurship (see, Ratten (2011)) driven by the emergence and rise of innovative digital technologies in the sports industry that cover a wide range of sports and have a lively community (Gruettner, 2019; Ratten, 2017). Considering the forecasting power (e.g., in terms of diverse knowledge and expertise), recent academic studies published in scholarly journals have shown interest in investigating the underlying dynamics of online tipster communities (e.g., Brown and Reade (2019)). In this vein, the user-generated content of tipsters – which we will refer to as tipster-generated content (TGC) in this study – offers the potential to become a revealing data source to improve sports betting returns.

TGC has proven to be valuable in predicting the outcomes of sports events, as it not only contains concrete predictions of match results but also (often) background information about the tipsters or even detailed textual match analyses. Existing studies have been published on the “wisdom-of-crowds” effect of Surowiecki (2004) (e.g., Brown and Reade (2019), O’Leary (2017), Peeters (2018), or Schumaker et al. (2016)). The wisdom-of-crowds effect operates on the premise that an averaging of forecasts eliminates individual prediction errors and thus leads to greater accuracy. In other words, large groups of individuals are better at making predictions than individuals are. The effect has tremendous practical implications: First, it suggests that decisions made by the majority rule (or by averaging opinions) will outperform decisions made by single experts. Second, it suggests that decisions made by the majority rule will often be accurate in an absolute sense – an implication that partially accounts for the rapidly increasing use of information markets to predict events (Simmons et al., 2011).

TGC can be seen as a valuable source to extract the wisdom of an online tipster community to predict the outcomes of sports events. However, existing studies on TGC and the wisdom-of-crowds effect come with several shortcomings: First, existing studies – except Brown and Reade (2019) – have not tested the wisdom-of-crowds effect in a realistic setting using data extracted from a real-world online tipster community. Second, the data sets with which existing studies have performed their analyses are not very comprehensive. As a consequence, they have not included rich information on the characteristics of the crowd to identify, for instance, smarter sub-crowds, although relevant literature on the wisdom-of-crowds effect have concluded that not only can crowdsourcing outperform experts but additionally the characteristics of the crowd are likely to influence the prediction results (O’Leary, 2017). Third, current evidence from related literature streams, such as financial studies, innovation management and entrepreneurship, suggests that textual user-generated content provides a rich crowd-based pool on information that can be easily collected and analysed to extract the wisdom of crowds (Chen et al., 2014; Otterbacher, 2009; Rhyn & Blohm, 2019). However, existing studies have just taken the tipsters’ concrete results predictions as well as a few tipster-related characteristics, such as tipsters’ prior tip experience, into account. Therefore, they have not examined the detailed textual match analysis attached to a published tip.

To address the abovementioned shortcomings, we extracted data from Blogabet.com, an online tipster community founded in 2006 that covers a wide range of sports and has a

lively community with around 291,156 members.¹ Our final data set comprises 1,534,041 verified tips (which we will refer to as picks in the rest of this study) extracted from 3,484 tipsters. For each of these picks, our data set includes (1) contributor-/(tipster)-related variables such as prior experience, past performance and location, (2) textual-content-/(match-analysis)-related variables such as its length, its readability and its specificity as well as (3) community-related variables that measure the feedback and reaction of the community to specific tipsters in terms of the number of followers. We aim to leverage our collected data set to investigate whether (1) there is wisdom in a crowd of an online tipster community that can be used to improve betting returns and whether (2) we can identify specific characteristics that are indicators for smarter sub-crowds within the overcall crowd of an online tipster community.

To do so, we conducted two evaluations: First, we averaged and compared the implied win probability of the odds set by the bookmakers with the actual win percentage of the picks proposed by tipsters on Blogabet.com. Second, we conducted a mixed-effects logistic regression model (MELR) that identifies the characteristics of significantly smarter sub-crowds. The contributions of our study are as follows: For researchers, we prove that there is indeed wisdom in online tipster communities. Moreover, we propose a set of variables that explain specific characteristics that are indicators for smarter sub-crowds. For practitioners, our results provide important insights into TGC and the wisdom of online tipster communities that are especially relevant for bookmakers as well as tipsters, which either protect them from losses or improve their (betting) returns.

The remainder of this study is structured as follows: Section 2 introduces the theoretical background and related work. Our hypotheses are stated in Section 3. The methodology is described in Section 4. Section 5 presents our results as well as the discussion. We then propose the implication, avenues for future research and the limitations in Section 6. Finally, we conclude our study in Section 7.

Using Tipster-Generated Content to Extract the Wisdom of an Online Tipster Community to Predict Sports Outcomes

In recent years, user-generated content has raised the interest of researchers and practitioners alike as it allows to leverage publicly available data that often contains valuable information (e.g., customer needs or opinions) (Krumm et al., 2008). Therefore, many organisations are currently racing towards extracting insights from

¹ The numbers are based on Blogabet.com's website and are as of 13 December 2019.

user-generated content and leveraging them based on new business models (Byrum & Bingham, 2016). In the sports industry, user-generated sports content becomes a revealing data source as it offers (freely) available information such as discussions and experiences of fans or even thoughts and feelings of professional athletes (Gruettner et al., 2020). In the context of this study, TGC can be seen as user-generated tip recommendations that provide a valuable source to extract the wisdom of an online tipster community to predict the outcomes of sports events.

The wisdom-of-crowds effect operates on the premise that the independent judgement of a crowd of individuals (as measured by any form of central tendency) will be relatively accurate, even when most of the individuals in the crowd are ignorant and error-prone (Surowiecki, 2004). The effect has been studied and discussed in many research fields and scholarly articles in recent years, for instance, in financial studies to predict future stock returns. In this vein, Chen et al. (2014) proved that a textual analysis of users' posts on Seekingalpha.com, a popular opinion forum for stock market investors, has predictive power for future stock returns. In (product) innovation management, Hoornaert et al. (2017) as well as Beretta (2018) demonstrated that adding crowd-related information such as feedback (e.g., in the form of online comments) to the idea selection process helps to identify ideas that are more likely to be successful. Similarly, in the entrepreneurship literature, Mollick and Nanda (2016) showed evidence that support on the crowdfunding website Kickstarter.com is a better predictor of the success of theatre productions than evaluations by a designated expert panel. However, although many of the conducted studies support the wisdom-of-crowds effect, there have also been studies which challenge the accuracy and fundamental premises of crowd prediction: For example, Haan et al. (2005) concluded that experts are less sensitive to the emergence of new information than crowds. Thus, experts are likely to act less impulsively. Critics also observed that some crowd members might simply select crowd favourites rather than evaluate the data independently. This reinforcing behaviour could, for instance, lead to over-valuing the crowd favourite and can have a negative impact on the prediction accuracy (Budescu & Chen, 2015; Peeters, 2018). As a consequence of these ongoing discussions, we believe that in a sports outcome prediction context a fair test of the wisdom-of-crowds effect requires an investigation of a crowd in a realistic real-world market setting.

Several research articles based on the wisdom-of-crowds effect have been published in recent years to investigate whether crowd wisdom can be used to predict sports outcomes. One of the most prominent examples includes Twitter, which has been studied as a predictor of soccer games. For instance, Schumaker et al. (2016)

investigated whether the sentiment contained in tweets can serve as a meaningful proxy to predict match outcomes. The authors found that crowdsourced sentiment can be a better predictor of match outcomes than odds. Likewise, Peeters (2018) concluded that information extracted from Transfermarkt.de evaluations – where online users submit transfer valuations of soccer players – could be used to generate sizeable betting returns. In detail, the author showed that forecasts of international soccer results based on the crowd's evaluations are more accurate than those based on standard predictors (e.g., FIFA ranking). O'Leary (2017) compared the performance of a Yahoo crowd to experts in predicting the outcomes of matches in the FIFA World Cup 2014. The analysis found that the crowd was statistically significantly better at predicting outcomes of matches than experts and very similar in performance to established betting odds. However, none of the existing studies dealt with a specific online tipster community context. There is just one study – to the best of the knowledge of the authors – which evaluated the wisdom-of-crowds effect in a realistic real-world online tipster community setting. Brown and Reade (2019) extracted data from Oddsportal.com and investigated the accuracy of crowd forecasts. The authors found that the crowd outperforms bookmakers in specific cases, leading to the conclusion that tip recommendations (i.e., TGC) in online tipster communities contain information that is not in betting prices.

Existing studies come with several weaknesses in the way that they did not meet the four conditions for crowd wisdom set out by Simmons et al. (2011). According to the authors, a crowd is smart when the members in the crowd are (1) knowledgeable, (2) motivated to be accurate, (3) diverse and (4) independent. The majority of the existing studies used either social media platforms such as Twitter (e.g., Schumaker et al. (2016)) or data extracted from online websites such as Yahoo (e.g., O'Leary (2017)) for their analyses. Social media platforms or online websites have not been set up with the aim of eliciting crowd wisdom. Thus, they usually generate collective judgements (Peeters, 2018). Furthermore, they typically do not provide users with any explicit incentives to induce accurate reporting. For example, the incentives on social media to provide accurate information for forecasting may arguably be weak. Unlike online tipster communities, accurate social media forecasts may enhance an individual's reputation but are not directly profitable. Even worse, there are many instances of misinformation on social media (Antretter et al., 2019; Chen et al., 2014). In the same vein, social media platforms and online websites often make little attempt to reach a diverse user population. Finally, they usually allow (and indeed stimulate) communications between users, which may limit the independence of users' opinions (Peeters, 2018).

We believe that our online tipster community setting from Blogabet.com meets the four conditions set up by Simmons et al. (2011) for the following reasons: First, online tipsters are knowledgeable as they publish sports picks regularly, mostly for only a specific selection of sports types or games. Second, we assume that they are motivated not only to publish accurate picks because they (probably) want to place their own money on a particular pick but also to build up a strong track record within the tipster community (e.g., a favourable ranking between all tipsters of the online tipster community) that allows them, for example, to offer access to their picks on a paid basis. Third, tipsters in an online tipster community are embedded in a broad and diverse network of people over the internet. Accordingly, they have different backgrounds and are even interested in different types of sports or sports clubs. Hence, the crowd is sufficiently diverse, decentralised through the reach of the internet, able to be summarised and rapidly independent. Consequently, and fourth, tipsters' evaluations of specific picks (mostly) rely on their own information and are not influenced by other members of the online tipster community.

Extracting TGC and the wisdom of an online tipster community poses several challenges: As described, contributions in online tipster communities are submitted by a diverse network of people with different backgrounds and degrees in expertise. As a result, the quality of the tipsters as well as of the published picks vary drastically from excellent to noise and ambiguity, such as abuse and spam. In addition, besides structured data such as predefined match-related metadata (e.g., the kick-off time), there is a wide array of unstructured data such as TGC. Figure 1 shows an overview of the tipsters ordered by the number of followers of an online tipster community. Figure 2 illustrates an example of a published pick of a tipster. As a consequence of the abovementioned challenges, the process of manually reviewing and filtering the large amount of tipsters and TGC to identify valuable picks and separating them from low-quality contributions that should not be used to extract the wisdom of the crowds is a latent challenge. Text mining and Natural Language Processing (NLP) techniques represent promising solutions to cope with the vast amount of contributions in an online tipster community. Thus, they provide the means to discover patterns and extract useful information from textual data in a fast, automatic, scalable and repeatable way (Rhyn & Blohm, 2019).

Research from prior literature on the wisdom-of-crowds effect as well as from related literature streams applied such techniques to extract the wisdom of vast amounts of contributions. The findings commonly reported that highly valuable contributions are marked by specific characteristics. In this vein, Hoornaert et al. (2017) proposed a model that can be adapted to the context of this study and helps to analyse TGC to extract the

wisdom of an online tipster community. The authors identified three sources of information (the “3Cs”) available in online communities: (1) Characteristics about the contributor, (2) the textual content of the contribution and (3) the community's feedback and reaction. The “contributor” category refers to TGC that contains information about the tipster who published a pick. For instance, previous studies have proven evidence that crowds improve their forecasting performance over time as they become more experienced and skilled (e.g., Budescu and Chen (2015), Goldstein et al. (2014), or Lamberson and Page (2012)). Therefore, it is reasonable to assume that experienced and skilled tipsters also form smarter sub-crowds in an online tipster community. The “content” category refers to the content-related textual features of a textual contribution, that is, the detailed textual match analysis attached to a published pick, which is usually expressed in unstructured, written text. Content-related textual features, such as the length (e.g., Riedl et al. (2013) or Wang and Strong (1996)), the readability (e.g., Flesch (1943) or Otterbacher (2009)) or the specificity (e.g., Otterbacher (2009) or Weimer and Gurevych (2007)), can be used to examine how carefully a tipster has elaborated their picks. This, in turn, can be used to assess the quality of a pick and, hence, helps to identify patterns of smarter sub-crowds. The “community” category refers to the feedback and reactions of the community. For instance, studies highlight that positive feedback from the community in the form of, for example, comments or likes can be seen as a proxy for the communities’ satisfaction with a specific tipster (e.g., Antretter et al. (2019), Beretta (2018), and Hoornaert et al. (2017)). Thus, it can be used as a means to identify smarter sub-crowds of tipsters.

In this study, we build upon the principles of the “3Cs” to derive our hypotheses that are explained in the following section.

Figure 1

Overview of the Tipsters Ordered by the Number of Followers of an Online Tipster Community (Here Blogabet.com)

TIPSTERS

FILTER BY LANGUAGE

Show all

FILTER BY PICK TYPE

Show all

FILTER BY SPORTS

All Sports

FILTER BY SPORTS PERCENT

All

FILTER BY LEAGUES

All Leagues

FILTER BY NUMBER OF PICKS

All

FILTER BY ACTIVITY

Active last 12 months

FILTER BY BOOKIES USED

Nothing selected

FILTER BY BOOKIE PERCENT

All

ORDER BY

Number of followers



bobic

sabobic.blogabet.com



2012

Since

1103

Picks

+724

Profit

+18%

Yield

98%

Verified

13981

Followers



FOLLOW



dedi22

dedi22.blogabet.com



2015

Since

2234

Picks

+3797

Profit

+37%

Yield

100%

Verified

10486

Followers



FOLLOW

55€ / MONTH





Vivaspain

vivaspain.blogabet.com



2012

Since

1172

Picks

+425

Profit

+18%

Yield

100%

Verified

7413

Followers



FOLLOW



Aussie1126

aussie1126.blogabet.com



2014

Since

1321

Picks

+984

Profit

+21%

Yield

100%

Verified

6764

Followers



FOLLOW



cheser

cheser.blogabet.com



2014

Since

2323

Picks

+1174

Profit

+14%

Yield

100%

Verified

6606

Followers



FOLLOW

40€ / MONTH



Figure 2

Illustration of a Published Pick of a Tipster in an Online Tipster Community (Here Blogabet.com)

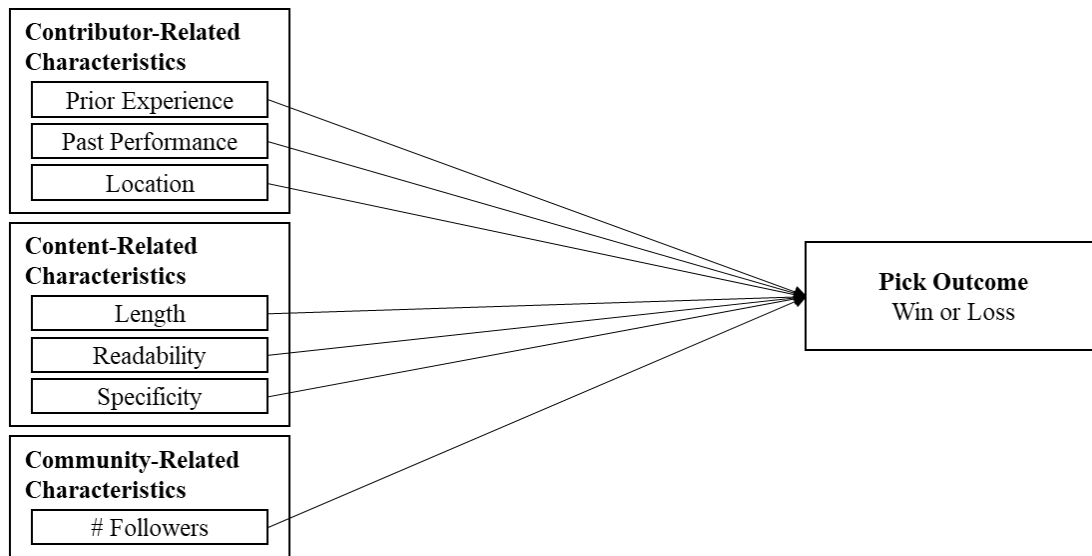
The screenshot shows a tipster's pick for a football match. At the top, it says "bobic published a new pick:" with a timestamp of "Sat, Aug 17th, 2019, 12:55". The match is "Royal Pari FC - Nacional Potosi". Below the match name, it shows "Full Event Home -0.50 (AH) @ 2.02 3/10 PINNACLE" with a red arrow pointing down and "-3.00 x 0:2 (0:0)". It also indicates "Football / Bolivia / Kick off: 17 Aug 2019, 23:30". A blue button says "Bet on PINNACLE via BET-IBC".

The tipster's analysis follows: "Roya Pari are strong home side. They have good squad and coach in face of Mosquera. At home they are always dominating against every team in Bolivia. On the other hand Nacional Potosi are weaker than last year. Their best players, stikers Reina 20/7 , Royon 12/6 and Vladimir Castellon 25/12 left the team. Thats a huge blow in their attack. Nacional Potosi are very poor in their visitors. Their record in last 5 visits is just 1-0-4. Losing vs teams like Sport Boys and Destroyers talks about their poor performance away from home. Royal Pari is better team and will have big home advantage of hot weather. Big value on Royal Pari. Fair odds Royal Pari -0.5 @1.60. GL!"

Below the analysis, there is a "Like" button and "Likes (8) Comments (2)". Two comments are visible: one from "b3tting" at 13:01 saying "2,05 on 365" and another from "bobic" at 13:24 saying "33 celsius expected today big advantage for Royal Pari,home team also sign national team player Rodrigo Vargas." At the bottom, there is a "Write a comment..." field and a "Post Comment" button.

Hypotheses Development

For the development of our research model (see, Figure 3), the following hypotheses draw from prior literature on the wisdom-of-crowds effect as well as from related literature streams (e.g., crowdsourcing, entrepreneurship, financial studies, innovation and idea generation, and data quality and computational filtering approaches) to provide a discussion of what can be measured about each of the “3Cs” proposed by Hoornaert et al. (2017) to extract the wisdom of an online tipster community.

Figure 3*Research Model of this Study*

Development of Contributor-Related Crowd Characteristics (Hypothesis 1)

Contributor (i.e., tipster-related) characteristics, in our study, refer to (1) the prior experience, (2) the past performance and (3) the location of the crowd in an online tipster community. Previous studies have noted that the prior experience and the past performance of a crowd are positively related to the accuracy of predictions and can, therefore, outperform larger crowds as well as experts. For instance, Goldstein et al. (2014) identified smaller sub-crowds that beat the wisdom of bigger crowds in a Fantasy Football Player selection. Their findings were that both the prior experience, measured as average years the crowd has played, and the past performance, measured as ranking within the community, have a significant positive impact. These findings are in line with Lamberson and Page (2012), who investigated the optimal group composition for accurate forecasts, as well as with Budescu and Chen (2015), who improved the quality of aggregate forecasts by eliminating poorly performing individuals from the crowd. Extrapolating from these studies, prior tip experience and a history of successful picks in the past may indicate a sub-crowd in an online tipster community that has expertise and knowledge, which in turn suggests that they are indicators of smarter sub-crowds within the overall crowd of an online tipster community.

Evidence for the wisdom-of-crowds effect has been shown with large, diverse samples in many different contexts. However, several previous findings concluded that it matters far less whether the crowd is homogenous or diverse along demographic dimensions, such as age, sex or the location of crowd members (e.g., de Oliveira and Nisbett (2018)

or van Dijk et al. (2012)). In contrast to these findings, we assume that demographic characteristics of the crowd could be indicators of smarter sub-crowds in the context of sports betting for the following reasons: Recently published studies commonly reported that the sports betting markets reveal inefficiency and information asymmetries. For example, Elaad et al. (2019) found that individual bookmakers are not efficient. Their own odds do not appear to fully use the information contained in their competitors' odds. Furthermore, Brown and Reade (2019) demonstrated, as described, that online tipster communities contain information that is not in betting prices. We expect that crowds of tipsters that bet on picks played in their home country will have more detailed information and expertise about the pick than foreign tipsters and bookmakers. This is especially true for marginal sports like table tennis or lower country-specific sports leagues in which usually less public information is available (Peurala, 2013). Therefore, the location of a crowd is added to our research model.

Consequently, one should expect that the prior experience, the past performance and the location of a crowd of tipsters are indicators of smarter sub-crowds.

H1a: *The characteristics of the crowd in terms of its prior experience is positively related to correctly predicting the outcome of a pick.*

H1b: *The characteristics of the crowd in terms of its past performance is positively related to correctly predicting the outcome of a pick.*

H1c: *The characteristics of the crowd in terms of its location is positively related to correctly predicting the outcome of a pick.*

Development of Content-Related Crowd Characteristics (Hypothesis 2)

The analysis of content-related textual features is commonly used in practice to extract, for instance, the creativity, expertise or workforce of a given crowd (Rhyn & Blohm, 2019). Hence, several recent studies provide supportive evidence of the value of this type of crowd wisdom. For instance, Klein and Garcia (2015) presented an approach called the “bag of lemons”, which enables crowds to filter ideas based on textual features with accuracy superior to conventional approaches. Chen et al. (2014) showed, as described, that a textual analysis of users' posts on Seekingalpha.com has predictive power for future stock returns. In this study, we focus on three content-related textual features for picks that include a detailed textual match analysis that are likely to be indicators of smarter sub-crowds: That is (1) the length, (2) the readability and (3) the specificity of a published pick.

In an online tipster community, tipsters can use text to explain why they have chosen to bet on a specific pick. Providing sufficient and detailed information about the pick facilitates the evaluation process and demonstrates that the tipster has spent time to elaborate it. In turn, this increases the likelihood that a pick has a positive outcome. The amount of information in a textual contribution (i.e., its length) has frequently been discussed as one of its most important features by related literature (Wang & Strong, 1996). Longer contributions contain more information that could potentially be relevant for the outcome of the pick than shorter ones. On the other hand, researchers emphasised that contributions that are short and less elaborated tend to deliver less information (e.g., Riedl et al. (2013)). In the same vein, the readability of a pick can be used to analyse the syntactic and semantic complexity of a published pick, which we claim is also a proxy of how carefully a tipster has elaborated their pick (Flesch, 1943). Higher readability of a pick often indicates a better-evaluated pick and thus makes it easier to extract relevant cues or information. Past research has shown that a better readability score is likely to enhance the interpretability or clarity of a textual contribution and may enable the acquisition of the embedded information (Otterbacher, 2009). Lastly, related literature emphasises the need to consider the specificity and relevance of the information in a textual contribution. For instance, Weimer and Gurevych (2007) used similarity features to measure the relatedness of an online post to a forum topic. Likewise, Otterbacher (2009) quantified the extent to which a product review contains terms that are statistically important across other reviews as a significant indicator of the helpfulness of a review. In the context of this study, we believe that smarter sub-crowds base their decisions to publish a pick on clear and specific criteria, such as information that is based on quantitative pre-match analyses (e.g., win/loss ratio of home and away games), or other specific information, such as missing or injured players.

Following the argumentations above, we assume that the length, the readability and the specificity of a pick are used by smarter sub-crowds. Consequently, our content-related hypotheses to extract the wisdom of an online tipster community are as follows:

H2a: *The content-related characteristics of the crowd's picks in terms of its length is positively related to correctly predicting the outcome of a pick.*

H2b: *The content-related characteristics of the crowd's picks in terms of its readability is positively related to correctly predicting the outcome of a pick.*

H2c: *The content-related characteristics of the crowd's picks in terms of its specificity is positively related to correctly predicting the outcome of a pick.*

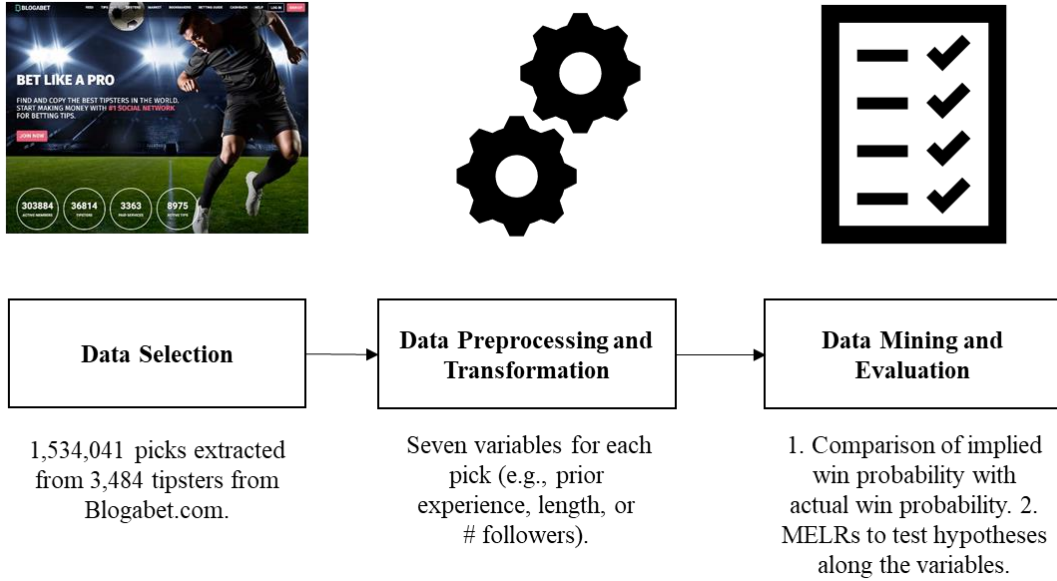
Development of Community-Related Crowd Characteristics (Hypothesis 3)

While H1 and H2 focus on the characteristics of the crowd in terms of its contributors and content, this study finds that it is also desirable to incorporate the feedback and the reaction of the online tipster community into the research model to identify specific characteristics that are indicators for smarter sub-crowds in an online tipster community. Studies on crowdsourcing (e.g., Hoornaert et al. (2017)) and idea selection (e.g., Antretter et al. (2019)) highlighted that positive feedback from the community indicates the popularity and approval of a member within the online community. Thus, it can represent a way of evaluating a particular crowd of tipsters. For example, Antretter et al. (2019) showed that the number of followers on Twitter is among the most important predictors for new venture survival. Likewise, Beretta (2018) noted that one should not only attract a large number of contributors to extract the wisdom of crowds but also considers types of participants to engage with to enable access to diverse knowledge and expertise. Therefore, we pose the following hypothesis:

H3: *The feedback and the reaction of the community in terms of the number of followers are positively related to correctly predicting the outcome of a pick.*

Methodology

To validate our hypotheses, we followed an adjusted knowledge discovery process in databases (KDD) as described by Fayyad et al. (1996). KDD describes the structured extraction of useful information from a large volume of data. For the purpose of this study, KDD's iterative cycle approach was combined and formed into three steps: (1) data selection, (2) data pre-processing and transformation and (3) data mining and evaluation (see, Figure 4), which are explained in the following subsections.

Figure 4*Adjusted KDD and the Corresponding Results/Approach*

Data Selection

The first step of our adjusted KDD included the selection of data sources. As described, we chose Blogabet.com as a representative of an online tipster community. We chose Blogabet.com for two major reasons: First, all historical data about tipster-related characteristics (i.e., prior experience, past performance and location), content-related characteristics (i.e., the length, readability and specificity of the textual match analyses), community-related characteristics (i.e., number of followers) as well as the pick outcome are publicly available. Second, Blogabet.com has implemented the “verified odds” concept. This concept guarantees that the odds displayed are accurate and available at the time of publishing the pick. To achieve that, Blogabet.com uses multiple direct bookmakers’ feeds where tipsters can verify the right picks for their selections in real time. Furthermore, Blogabet.com has implemented a review system in which tipsters correct and verify themselves. Consequently, a high reliability and accuracy of the picks included in our data set are guaranteed.

We added the entire pick archive of a given tipster (see, Figure 5 for the inclusion/exclusion criteria for the selection of tipsters and picks) to our data set for further analysis: Blogabet.com recorded 11,378 active tipsters (i.e., the tipster published at least one pick in the last twelve months) who had published 4,005,176 picks at the time of our analysis (the numbers are as of 28 September 2019). To narrow down the selection of tipsters, in a first step, we deleted all tipsters that had less than 100 picks or

more than 2,000 picks from our data set. It is generally accepted in the sports betting market that tipsters that have less than 100 picks are not reliable, and their past performance may be due to luck or coincidence. Likewise, a closer look at the tipsters that had more than 2,000 picks made the impression that multiple tipsters or commercial ventures maintained these accounts. In the next step, we excluded all tipsters and picks that were not 100% verified as described above, as well as tipsters that were banned, paused or reset their pick history during the data collection phase. We then excluded all picks whose match analysis was not written in English in order to obtain reliable results, since some of our applied data analytics techniques are designed for the English language. In total, 1,534,041 picks from 3,484 tipsters were included in our final data set for further analysis. This data set is publicly available and can be accessed via the following reference: Gruettner (2020).

Figure 5

Inclusion and Exclusion Criteria for Tipsters and Picks

	# Tipsters	# Picks
Overview Blogabet.com as of September 28 th , 2019	11,387	4,005,176
Tipsters with at least 100 picks	4,721	3,831,027
Tipster with not more than 2,000 picks	4,307	2,101,116
Tipsters with 100% verified picks	3,544	1,694,022
Unmatched tipster or pick (banned, paused, or reset)	3,484	1,580,211
Textual match analysis in English	3,484	1,534,041

Data Pre-Processing and Transformation

After extracting and collecting relevant data, the next step was data preparation and transformation. Therefore, we transformed the collected data into a structured format that can be accessed and processed for further analysis. To validate our developed hypotheses (i.e., H1 to H3), we generated several variables that are based on the extracted data. Three example picks of the used variables can be found in Table 1 and are explained as follows: To examine if there is wisdom in a crowd of an online tipster community that can be used to improve betting returns, we averaged and compared the implied win probability of the odds set by bookmakers with the actual win percentage of the picks proposed by tipsters. The implied win probability was calculated as:

$$\text{Implied Win Probability} = 1 / \text{Odds}$$

The actual win percentage was calculated as:

$$\text{Actual Win Percentage} = \text{Number of won picks} / \text{Number of all picks}$$

To identify specific characteristics that are indicators for smarter sub-crowds within the overall crowd of an online tipster community, we first pre-processed and transformed all data related to the contributor-related crowd characteristics (H1). The prior experience variable displayed the number of all picks that a tipster had published. The past performance variable contained the so-called yield of a tipster. The yield is commonly used in sports betting to compare the performance of tipsters. To test whether a pick was published in the home country of a tipster (H1c), we created a location dummy variable in which 1 stood for a pick that was placed in the same location and 0 for a pick that was not.

To test the content-related crowd characteristics (H2), different data analytics techniques using the programming language Python 3.7 were applied, since it is widely known, easy to use and supports major libraries for NLP tasks: First, all picks that included a textual match analysis with less than four words were set to 0 because they did not contain any relevant information or included noise for further analysis. Afterwards, we calculated the length of each match analysis by splitting the picks' strings into tokens. We deleted all English stop words based on the Natural Language Toolkit (NLTK) stop word list for the English language (Bird et al., 2009). Finally, we counted the number of filtered tokens of each pick. The token length of a pick was then used to test our hypothesis H2a. Next, we aimed to measure the readability of each match analysis to test if it correlates to the outcome of a pick. To measure the syntactic readability of texts, several measures have been used in research (Khawaja et al., 2010). We selected the Flesch-Reading-Ease (FRE) to capture the readability of the picks since this score combines language complexity measurements, such as the average sentence lengths and the average syllables per word, into one number (Flesch, 1943). The score has been widely used before to determine the readability of a message in computer-mediated communication (e.g., Walther (2007)) or for measuring the readability of posts in online forums (e.g., Wambsganss and Fromm (2019)). We used the following formula:

$$\text{Flesch Reading Ease} = 206.835 - (1.015 * \text{asl}) - (84.6 * \text{asw})$$

asl: average sentence length of a response, asw: average syllables per word

The scores of our answers ranged from 0 to 120.20. The higher the FRE score was, the better the readability of the match analysis. The FRE was used to test our hypothesis H2b. Moreover, we aimed to test the specificity of information in a given pick. Therefore, we retrieved a dictionary of domain-related vocabularies from “word net” (Princeton University, 2019), representing related words for sports injuries. We controlled if the word stem of any token in a pick matches the word stem of any word in our dictionary. For stemming the tokens of our picks and the dictionary entries, we used the English Porter Stemmer provided by NLTK by Bird et al. (2009).

For the validation of H3, we used the number of followers as a variable that measures the community-related crowd characteristics extracted from Blogabet.com.

To measure the success of a published pick, we chose the pick outcome in terms of whether a tipster won or lost the pick. We chose the outcome of the pick as our dependent variable, as it is the performance measure used by tipsters in practice and, therefore, can be applied as a proxy for the quality of the pick and in turn helps to extract the wisdom of the crowd.

Table 1

Three Examples of a Published Pick and their Corresponding Variables

Textual Match Analysis	Pick Outcome	Implied Win Probability	Prior Experience	Past Performance	Location	Length	Readability	Specificity	# Followers
United has got a lot of injured player, Arsenal have got a high quality attacker players...	Win	47.62%	108	2%	0	8	55.97	1	14,464
I think Leganes wants this game much more. They are strong in their home. Even the games they lost in their home they played pretty well. Leganes has the momentum and the passion to win this game. Possible scores 1-0 / 2-1	Loss	27.78%	435	7%	0	22	98.55	0	133
-	Win	55.56%	1,769	-4%	1	0	0	0	5

Data Mining and Evaluation

The last step of our adjusted KDD included the data mining and evaluation phase. As described, to examine if there is wisdom in a crowd of an online tipster community that can be used to improve betting returns, we averaged and compared the implied win probability of the odds set by the bookmakers with the actual win percentage of the picks proposed by tipsters. The difference between both variables indicates whether there is

wisdom in a crowd of an online tipster community that can be used to improve betting returns (see, Brown and Reade (2019)). To identify specific characteristics that are indicators for smarter sub-crowds within the overall crowd of an online tipster community, we measured the impact of the abovementioned variables on the outcome of the pick. In our data set, a particular tipster produces multiple data points (i.e., picks) over time. Therefore, traditional statistical approaches (e.g., linear/logistic regressions or analysis of variance) are of limited use because of restrictive assumptions concerning the variance-covariance structure of the repeated measures in longitudinal data sets (Hedeker & Gibbons, 2006; Laird & Ware, 1982). For instance, (1) error terms that are correlating with each other or (2) variances that lead to different sources of heterogeneity (e.g., between tipsters or within a tipster themselves) (Fitzmaurice et al., 2012). To deal with these challenges and to predict dichotomous outcome variables (i.e., binary outcome variables), researchers commonly use MELRs when observations are correlated. MELRs have shown to be sensitive and statistically powerful while dealing with longitudinal data sets as well as missing values in various theoretical as well as practical studies (e.g., Vermunt (2005)). Therefore, we believe that an MELR guarantees robustness and reliability to test H1 to H3 in our study. For our analysis, we scaled the values of our data set to meet the requirements of the lme4 MELR R package version 1.7 as proposed by Bates et al. (2014).

Results and Discussion

This study set out with the aim of assessing whether (1) there is wisdom in the crowd of an online tipster community that can be used to improve betting returns and whether (2) we can identify specific characteristics that are indicators for smarter sub-crowds of an online tipster community. To validate our hypotheses, as mentioned, a comparison of the implied win probability with the actual win percentage of the picks and an MELR was conducted. The detailed results, including a discussion, are presented in the subsections below.

Descriptive Statistical Results

Our final data set comprised 3,484 tipsters, who published a total of 1,534,041 picks (see, Table 2 for an overview of the descriptive statistics). Overall, the results showed that 797,769 or 52.00% out of the total 1,534,041 picks had a positive pick outcome (i.e., the tipster won the pick). In contrast, the averaged implied win probability set by the bookmakers of all picks was only 48.71%. The difference is 3.29%. Extrapolating

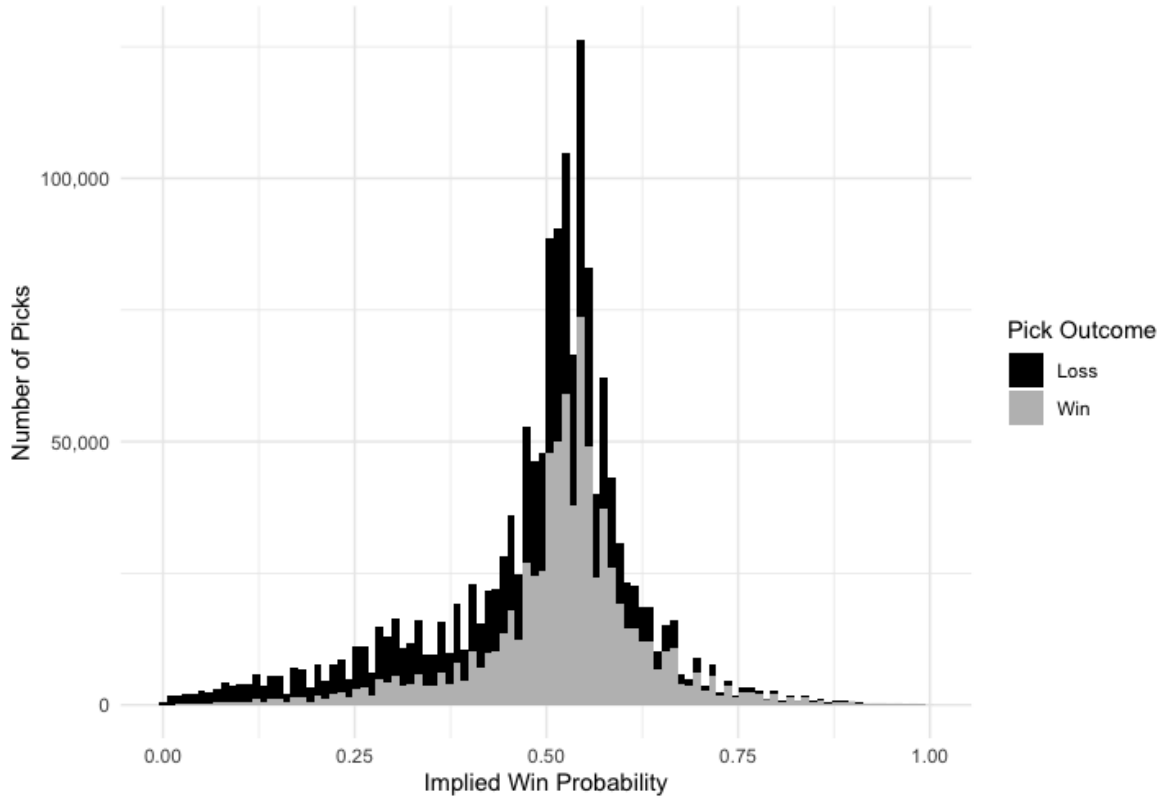
from this, we can prove that there is indeed wisdom in the crowd of an online tipster community that can be used to improve betting returns. In this vein, Figure 6 shows a detailed overview of the relationship between the implied win probability, the number of picks and the pick outcome. In 67,024 picks, or to put it differently, in 4.44% of the overall analysed picks, tipsters included a textual match analysis. These picks showed an actual win percentage of 54.90%. In contrast to the averaged implied win probability of 50.54%. The difference, that is 4.36%, indicates that the subset of picks which included a textual match analysis achieved even better results than the overall dataset including all picks. Out of the total 1,534,041 picks, 205,039 picks (or 13.37%) were played in the local home country of the tipster who published the pick. 55.54% of these picks had a positive pick outcome, in contrast to an average implied win probability of 50.55%. The difference is the largest (4.99%), indicating that smarter crowds publish picks in their home country.

Tipsters' prior experience showed a mean of 478 published picks. They were registered on Blogabet.com for a mean of 2.83 years. Furthermore, tipsters had average yields of 3.97%, indicating a positive past performance of the majority of tipsters in our data set. The length of picks that included a textual match analysis showed a mean of 19.95. While the readability score of these picks presented a mean of 82.89, the specificity score showed a mean of 0.07. The community's feedback, in the form of the number of followers for each tipster included in our data set, showed a mean of 66.

Table 2

Descriptive Statistics of the Final Data Set

Data Set Overview	Overall		Including Textual Match Analysis	Location
Number of tipsters = 3,484				
Number of Picks	1,534,041		67,024 (4.44%)	205,039 (13.37%)
Actual Loss Percentage	48.00% (736,272)		45.10% (30,225)	44.46% (91,168)
Actual Win Percentage	52.00% (797,769)		54.90% (36,799)	55.54% (113,871)
Av. Implied Win Probability	48.71%		50.54%	50.55%
Difference*	3.29%		4.36%	4.99%
Contributor Characteristics	Mean	SD	Min	Max
Prior Experience	477.63	417.44	100	1,997
Past Performance	3.97%	11.03%	-96%	110%
Content Characteristics				
Length	19.95	22.73	0	418
Readability	82.89	16.60	0	120.20
Specificity	0.07	0.31	0	9
Community Characteristics				
Number of followers	65.82	254.99	0	6,288
* The difference is calculated as: Actual Win Percentage – Av. Implied Win Probability SD = Standard deviation				

Figure 6*Relationship Between Implied Win Probability, Number of Picks, and Pick Outcome*

Evaluating Contributor-Related Crowd Characteristics

The first hypothesis proposed in this study was that the characteristics of the crowd in terms of its prior experience (H1a), its past performance (H1b) and its location (H1c) are positively related to correctly predicting the pick outcome. As Figure 7 and Table 3 show, the results of our performed MELR for the prior experience variable were significant at the $p < 0.01$ level and for the past performance variable highly significant at the $p < 0.001$ level. These findings confirm that prior experience (H1a) and past performance (H1b) are significant indicators of a smarter sub-crowd in an online tipster community that possesses expertise and knowledge and, thus, is likely to propose additional successful future picks.

These results stand in contrast to the only study by Brown and Reade (2019) that also investigated the wisdom of a crowd effect in a realistic real-world online tipster community setting. In their study, Brown and Reade (2019) concluded that sub-crowds that are evaluated based on prior experience (more past tips) or past performance (higher historical returns on their tips) did not achieve better accuracy than the overall crowd in an online tipster community. Brown and Reade (2019) defined their crowd of

experienced tipsters as “those who have previously lodged more tips than the median tipster who lodged a tip on the same event” (Brown & Reade, 2019, p. 3). Likewise, they defined a skilled crowd as “those who have, at the time, a higher hypothetical return on their tips than the median tipster who lodged a tip on the same event” (Brown & Reade, 2019). We believe that classifying tipsters on an event basis involves several risks: First, depending on which specific event is analysed, the threshold for being classified as an experienced/skilled or as an inexperienced/unskilled tipster can vary drastically. Second, therefore, it is likely that the same tipster is classified once in the crowd of experienced/skilled tipsters and once in the crowd of inexperienced/unskilled tipsters, depending on which other tipsters it is compared with. As a consequence, the data set will have dependencies in its observations. Thus, we believe that a Micer-Zarnowitz regression, which was applied in Brown and Reade’s study, is not appropriate for such an analysis. To handle the abovementioned risks, we applied an MELR and implemented specific inclusion and exclusion criteria (see, Subsection 4.1) to our final data set. Therefore, we think that our results are reliable as they stand in line with previous findings on the wisdom-of-crowds effects (e.g., Budescu and Chen (2015), Goldstein et al. (2014), or Lamberson and Page (2012)). In this vein, we assume that although sports betting is often associated with gambling tipsters improve their betting performance over time as they become more experienced and skilled (Levitt et al., 2012).

It was also hypothesised that a crowd of tipsters could be smarter based on its location (H1c). In detail, we expected that tipsters that bet on games played in their local home country would have more detailed information and expertise about the game and hence are more likely to propose additional information for successful future picks. The findings of our analysis showed significance at the $p < 0.001$ level (see, Table 3). Thus, H1c can also be confirmed. Prior studies have commonly reported that they did not find any evidence that crowds are smarter based on demographics (e.g., de Oliveira and Nisbett (2018) or van Dijk et al. (2012)). In our opinion, one can assume that the sports betting market differs from the previously conducted studies for two reasons: First, in an online tipster community, many tipsters specialise themselves, for example, on one specific type of sport or a specific sports league within a specific country. In turn, these specialised tipsters have a higher likelihood to have information that is not included in bookmakers’ odds and consequently are more likely to propose additional successful future picks. Second, the sports betting market is characterised by information asymmetries (Brown & Reade, 2019; Elaad et al., 2019). These asymmetries are often triggered by insider information as well as game manipulations, which can even be

multiplied in specific contexts, such as marginal sports, individual sports types and lower country-specific sports leagues (Peurala, 2013). Our findings, therefore, are consistent with previous studies that have shown that the wisdom-of-crowds effect is even stronger when less public information is available in the market (e.g., Da & Huang, 2019).

Evaluating Content-Related Crowd Characteristics

The second hypothesis in this study hypothesised that the characteristics of the crowd's textual match analyses in terms of its length (H2a), its readability (H2b) and its specificity (H2c) are positively related to correctly predicting the pick outcome. Content-related textual features are commonly discussed in related literature streams to extract the wisdom out of a crowd and have proved evidence of the high value of this type of crowd wisdom (e.g., Chen et al. (2014), Klein and Garcia (2015), or Rhyn and Blohm (2019)). Thus, we expected similar results in our study. However, the findings of our study did not support previous research (see, Table 3). Neither the length nor the readability and specificity variable could achieve any significant results. Therefore, we must reject H2a, H2b and H2c. Several factors could explain this observation: First, on Blogabet.com it is possible to only allow access to picks on a paid basis. When collecting the data set, we noticed that some of the tipsters who offered their picks on a paid basis deleted their match analysis after the game of a particular pick was played. A reason for that could be that they did not want to give insights into their betting strategies to the whole community as historical picks are publicly available to all members. Second, it is reasonable to assume that the picks that did not include a detailed textual match analysis are also elaborated carefully as tipsters in an online tipster community are intrinsically motivated enough to induce accurate picks (see, Section 2). We, therefore, assume that our online tipster community setting differs from other settings investigated so far in that, for example, in contrast to an idea selection context, no detailed textual match analysis needs to be included in a pick. However, although the results of this study were not significant for H2, we are still convinced that tipsters' textual match analyses provide a valuable information source to extract the wisdom of an online tipster community. Therefore, more research on this topic needs to be conducted.

Evaluating Community-Related Crowd Characteristics

The last hypothesis (H3) stated that the feedback and the reaction of the community in terms of the number of followers for a specific tipster is positively related to correctly

predicting the pick outcome. We assumed that the number of followers can be seen as a proxy for the communities' satisfaction with a specific tipster. Members of an online tipster community look for tipsters that publish picks that will help them to improve their betting returns. Therefore, they start following certain tipsters to monitor their published picks. This, in turn, leads to even more members being attracted, as other members also trust the tipsters with the most followers. In this sense, the tipsters in our study on Blogabet.com are ordered by default according to the number of followers. The results showed evidence for this hypothesis at the $p < 0.01$ level (see, Figure 7 and Table 3). This finding broadly supports the work of other studies from, for instance, product innovation in which positive feedback measured as the number of likes or the number of positive ratings have shown evidence to be a significant indicator for idea generation (e.g., Hoornaert et al. (2017)). While this finding is supported by previous studies in different contexts, it would be worthwhile to investigate whether the feedback and reaction of the community can also be used to improve pick forecasts. For example, it would be interesting to investigate if comments on a particular pick that are discussed on Blogabet.com contain information that is not included in the previously published pick, which could, in turn, improve the picks' forecasts.

Figure 7

Research Model Including Statistical Results of Conducted MELR

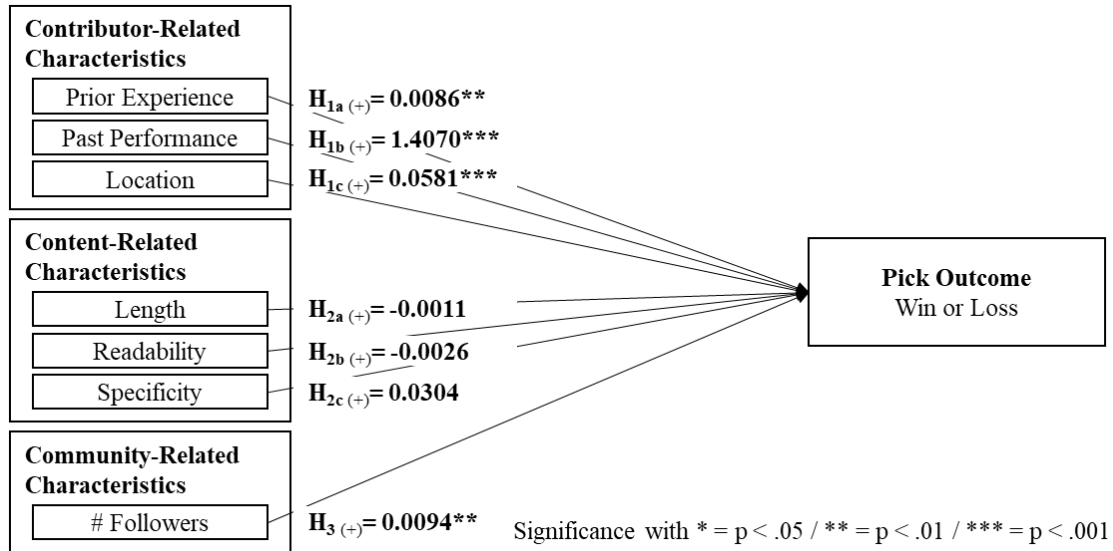


Table 3*Results of the MELRs*

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Intercept	-1.999*** (0.0088)	-1.995*** (0.0089)	-2.074*** (0.0081)	-2.005*** (0.0087)	-2.078*** (0.0081)	-1.999*** (0.0087)	-1.999*** (0.0087)	-1.999*** (0.0087)	-1.999*** (0.0088)	-1.995*** (0.0087)	-2.077*** (0.0081)
Implied Win Probability	4.246*** (0.0163)	4.247*** (0.0162)	4.290*** (0.0155)	4.243*** (0.0162)	4.289*** (0.0153)	4.246*** (0.0162)	4.246*** (0.0162)	4.246*** (0.0162)	4.246*** (0.0163)	4.248*** (0.0161)	4.289*** (0.0153)
Prior Experience		0.0085* (0.0041)			0.0105*** (0.0027)						0.0086** (0.0027)
Past Performance			1.430*** (0.0255)		1.426*** (0.0252)						1.4070*** (0.0258)
Location				0.0553*** (0.0060)	0.0581*** (0.0056)						0.0581*** (0.0056)
Length						0.0002 (0.0019)			-0.0009 (0.0024)		-0.0011 (0.0023)
Readability							0.0006 (0.0019)		0.0008 (0.0023)		-0.0026 (0.0022)
Specificity								0.0259 (0.0260)	0.0281 (0.0273)		0.0304 (0.0272)
# Followers										0.0590*** (0.0044)	0.0094** (0.0031)
Random Effect Tipster Intercept Variance	0.0257	0.0257	0.0079	0.0251	0.0073	0.0258	0.0258	0.0258	0.0258	0.0236	0.0073
AIC	2,008,580	2,008,578	2,006,294	2,008,498	2,006,174	2,008,582	2,008,582	2,008,582	2,008,585	2,008,409	2,006,169

Significance with * = $p < .05$ / ** = $p < .01$ / *** = $p < .001$

Logit scale regression coefficients followed by their standard errors in brackets.

An MELR was conducted on various models for this analysis, considering the pick outcome as the binary dependent variable. In models (2)-(4), (6)-(8) & (10) each of our explanatory variables was tested individually, whereas model (5) & (9) show the combined contributor and content variables, respectively. Finally, in model (11) an MELR was conducted with all our independent variables. To explain their relative explanatory power and model fit, we added the variance of the random effect intercept (tipsters) and the Akaike Information Criterion (AIC).

Implications, Future Research and Limitations

To investigate whether (1) there is wisdom in a crowd of an online tipster community that can be used to improve betting returns and whether (2) we can identify specific characteristics that are indicators for smarter sub-crowds within the overall crowd of an online tipster community, we developed a realistic real-world setting and extracted 1,534,041 picks which stem from 3,484 tipsters on Blogabet.com. The results showed that 797,769 or 52.00% out of the total of 1,534,041 picks had a positive pick outcome (i.e., the tipster won the pick). In contrast, the averaged implied win probability set by the bookmakers of all picks was only 48.71%. The difference is 3.29%. Similarly, tipsters had average yields of 3.97%, indicating a positive past performance of the majority of tipsters in our data set. Extrapolating from both of these findings, we can confirm that there is indeed wisdom in the crowd of an online tipster community that can be used to improve betting returns. To identify specific characteristics that are indicators for smarter sub-crowds of an online tipster community, we developed three hypotheses in accordance with the three sources of information (the “3Cs”) available in online communities as proposed by Hoornaert et al. (2017). The results of our MELR confirmed H1 that contributor-related crowd characteristics (i.e., prior experience, past performance and the location) are significantly positively related to correctly predicting the pick outcome. For H2, we assumed that content-related characteristics (i.e., a detailed textual match analysis) of the crowd’s picks in terms of its length, its readability and its specificity are positively related to correctly predicting the outcome of a pick. Content-related textual features have proven to be valuable in extracting the wisdom of online crowds in various contexts (e.g., Klein and Garcia (2015) or Ma et al. (2019)). However, we were not able to achieve any significant results for any content-related crowd characteristics in our study. Thus, we must reject H2. The last hypothesis (H3) stated that the feedback and the reaction of the community in terms of the number of followers for certain tipsters is positively related to the correct prediction of the pick outcome. This study showed significant results to confirm this hypothesis.

From an academic perspective, the contribution of this study is twofold: First, this study contributes to the literature stream of technological innovations in sports-based entrepreneurship and especially to online sports communities (see, Ratten (2011)). We proposed the second study, which applied the wisdom-of-crowds effect in a realistic real-world online tipster community setting and demonstrated that there is indeed wisdom in online tipster communities by analysing TGC. In doing so, we identified four characteristics, that is, prior experience, past performance, the location as well as the

number of followers, that are significant indicators for smarter sub-crowds within a community of online tipsters. These findings stand in contrast to the only other study conducted by Brown and Reade (2019), which also used data from an online tipster community. In future research, a closer look at the proposed characteristics could become relevant in identifying whether more successful crowds focus on, for instance, specific types of sports such as marginal sport types, lower country-specific sports leagues or individual vs. team sports. Likewise, although we were not able to achieve any significant results for any content-related crowd characteristics, we are still convinced that tipster-generated detailed textual match analyses provide a valuable information source to extract the wisdom of an online tipster community. Therefore, more research on this topic needs to be undertaken. Second, our results provide important insights into the value of user-generated content and the dynamics and the wisdom of online communities in general and, therefore, go beyond the sports literature. In this vein, this study proposes a set of variables to academia that explains specific characteristics that are indicators for smarter sub-crowds in online communities that should be studied in different contexts in future research. Especially the demographic variable, that is, the location of a crowd member, should be used by researchers in future studies since prior studies did not find any evidence for demographic characteristics within smarter sub-crowds (e.g., de Oliveira and Nisbett (2018) and van Dijk et al. (2012)). For example, in a financial market context. We further contribute to research as well as practice by making a novel, comprehensive, reliable and high-quality data set publicly available that provides many possibilities for future research.

From a practical perspective, the contributions are especially relevant for bookmakers and tipsters. On the one hand, we empirically demonstrated how digital technologies such as data analytics solutions can be beneficially implemented to extract the wisdom of online sports communities (Gruettner, 2019; Ratten, 2017). Thus, we provide a concrete example that can be used by bookmakers and tipsters to either protect themselves from losses or to improve their betting returns. In this vein, bookmakers can understand our results as a wake-up call to have a closer look at online tipster communities. For instance, bookmakers can use our findings to identify those crowds of tipsters that are most successful or influential in online tipster communities. In turn, they can adjust their odds directly after specific crowds of tipsters have published a pick. This provides them with an early protection system against losses. To do so, they can either build a dashboard that monitors online tipster communities for specific crowds of tipsters or they can build a prediction model based on machine learning algorithms and our published data set that classifies individual pick outcomes. On the other hand, this

study shows that tipsters can improve their betting returns using the wisdom of an online tipster community. Tipsters can, for example, adjust their betting strategies using our identified four characteristics of smarter sub-crowds to identify crowds of tipsters that publish the most valuable picks in online tipster communities. To generalise our study, organisations of any type that are currently racing towards extracting insights from user-generated content should be encouraged to use our findings to leverage user-generated content in their businesses (Byrum & Bingham, 2016).

This study is not free from limitations: First, online tipster communities such as Blogabet.com enable tipsters to observe the picks of others. Tipsters, thus, may be influenced which would lead in some cases to correlated forecast errors and an inferior crowd forecast. Similarly, this presents an issue for researchers as we cannot disentangle individuals' beliefs from the crowd's beliefs. Second, we believe that the lack of further control variables could be a limitation. Third, we used a specific set of Python libraries and pre-trained techniques (such as language detection). The accuracy of the techniques is limited to a certain degree; however, we believe that our results display the overall notion of the data. Last, we only investigated the positive impact of crowd characteristics on the outcome of a pick. It should also be investigated whether there are characteristics of crowds that indicate negative performances in terms of betting returns.

Conclusion

By investigating a real-world online tipster community from Blogabet.com, we analysed whether (1) crowd prediction can be used to improve sports betting returns and whether (2) there are specific characteristics that are indicators for smarter sub-crowds within such communities of sports tipsters. For research and practice, the conducted analyses showed evidence that such communities indeed contain wisdom which can be used to improve betting returns (tipsters won 3.29% more picks than the implied win probability set by bookmakers and produced average yields of 3.97%). We further identified four characteristics, that is, prior experience, past performance, the location as well as the number of followers that are significant positive indicators for correctly predicting the pick outcome and, thus, are characteristics which are typical for smarter sub-crowds. Our results provide important insights into user-generated content and the dynamic and the wisdom of online communities in general and, therefore, go beyond the sports literature. Future research should either try to identify further characteristics of successful sub-crowds or should concentrate on the proposed set of variables of this study and apply it to different research settings. Similarly, future research should dive

deeper into whether tipster-generated textual match analyses provide a valuable information source to extract wisdom, which did not show any significant results in this study. Our results are especially relevant to bookmakers and tipsters, who either want to protect themselves from losses or want to improve their betting returns.

References

- Antretter, T., Blohm, I., Grichnik, D., & Wincent, J. (2019). Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy. *Journal of Business Venturing Insights*, 11(e00109). <https://doi.org/10.1016/j.jbvi.2018.e00109>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. In *R package version 1.7*.
- Beretta, M. (2018). Idea Selection in Web-Enabled Ideation Systems. *Journal of Product Innovation Management*, 36(1), 5–23. <https://doi.org/10.1111/jpim.12439>
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media, Inc. <https://doi.org/10.1097/00004770-200204000-00018>
- Brown, A., & Reade, J. J. (2019). The wisdom of amateur crowds: Evidence from an online community of sports tipsters. *European Journal of Operational Research*, 272(3), 1073–1081. <https://doi.org/10.1016/j.ejor.2018.07.015>
- Budescu, D. V., & Chen, E. (2015). Identifying Expertise to Extract the Wisdom of Crowds. *Management Science*, 61(2), 267–280. <https://doi.org/10.1287/mnsc.2014.1909>
- Byrum, J., & Bingham, A. (2016). Improving Analytics Capabilities Through Crowdsourcing. *MIT Sloan Management Review*, 57(4), 43–48. <https://doi.org/10.7551/mitpress/11633.003.0020>
- Chen, H., De, P., Hu, Y., & Hwang, B.-H. (2014). Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media. *The Review of Financial Studies*, 27(5), 1367–1403. <https://doi.org/10.1093/rfs/hhu001>
- Da, Z., & Huang, X. (2019). Harnessing the Wisdom of Crowds. *Management Science*, 66(5), 1783–2290. <https://doi.org/10.1287/mnsc.2019.3294>
- de Oliveira, S., & Nisbett, R. E. (2018). Demographically diverse crowds are typically not much wiser than homogeneous crowds. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9). <https://doi.org/10.7302/Z2D50K5H>

- Deutscher Sportwettenverband. (2020). *Sportwettenmarkt in Deutschland bricht Umsatzrekorde*. <https://dswv.de/sportwettenmarkt-in-deutschland-bricht-umsatzrekorde/>
- Elaad, G., Reade, J. J., & Singleton, C. (2019). Information, prices and efficiency in an online betting market. *Finance Research Letters*. <https://doi.org/10.2139/ssrn.3378257>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceedings of the 2nd International Conference on Knowledge Discovery And Data Mining (KDD) 1996*, 82–88.
- Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied Longitudinal Analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Flesch, R. (1943). Marks of readable style; a study in adult education. *Teachers College Contributions to Education*, 897, ix + 69.
- Goldstein, D. G., McAfee, R. P., & Suri, S. (2014). The Wisdom of Smaller, Smarter Crowds. *Proceedings of the 15th ACM Conference on Economics and Computation (EC) 2014*. <https://doi.org/10.1145/2600057.2602886>
- Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. *Proceedings of the 25th Americas Conference on Information Systems (AMCIS) 2019*.
- Gruettner, A. (2020). *Data Set - The Wisdom of an Online Tipster Community*. <https://www.alexandria.unisg.ch/259455/>
- Gruettner, A., Vitisvorakarn, M., Wambsganss, T., Rietsche, R., & Back, A. (2020). The New Window to Athletes' Soul-What Social Media Tells Us About Athletes' Performances. *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS) 2020*.
- Haan, M. A., Dijkstra, S. G., & Dijkstra, P. T. (2005). Expert Judgment Versus Public Opinion - Evidence from the Eurovision Song Contest. *Journal of Cultural Economics*, 29(1), 59–78. <https://doi.org/10.1007/s10824-005-6830-0>
- Hedeker, D. R., & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Hoboken, NJ: John Wiley & Sons, Inc.
- Hoornaert, S., Ballings, M., Malthouse, E. C., & Van den Poel, D. (2017). Identifying New Product Ideas: Waiting for the Wisdom of the Crowd or Screening Ideas in Real Time. *Journal of Product Innovation Management*, 34(5), 580–597. <https://doi.org/10.1111/jpim.12396>

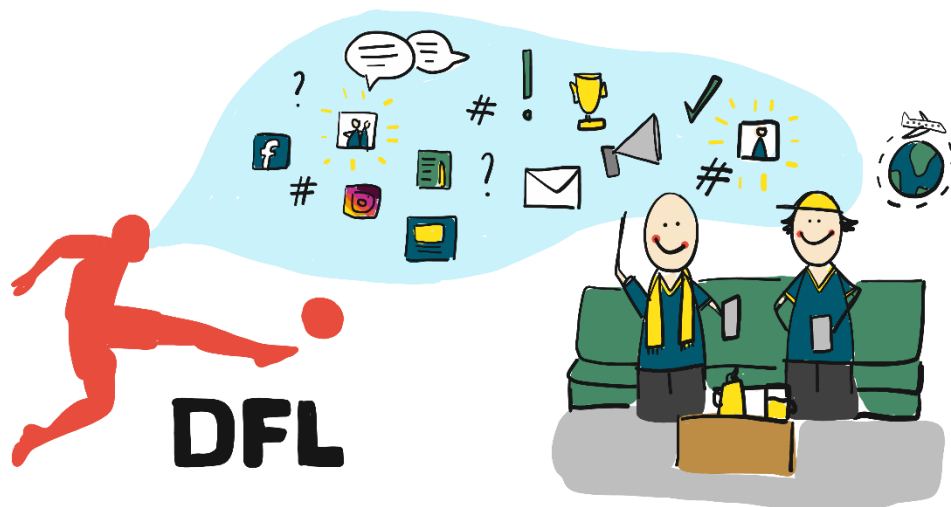
- Khawaja, M. A., Chen, F., & Marcus, N. (2010). Using language complexity to measure cognitive load for adaptive interaction design. *Proceedings of the 15th International Conference on Intelligent User Interfaces, Proceedings (IUI) 2010*. <https://doi.org/10.1145/1719970.1720024>
- Klein, M., & Garcia, A. C. B. (2015). High-Speed Idea Filtering with the Bag of Lemons. *Decision Support Systems*, 78, 39–50. <https://doi.org/10.1016/j.dss.2015.06.005>
- Krumm, J., Davies, N., & Narayanaswami, C. (2008). User-Generated Content. *IEEE Pervasive Computing*, 7(4), 10–11. <https://doi.org/10.1109/MPRV.2008.85>
- Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4), 963–974. <https://doi.org/10.2307/2529876>
- Lamberson, P. J., & Page, S. E. (2012). Optimal Forecasting Groups. *Management Science*, 58(4), 805–810. <https://doi.org/10.1287/mnsc.1110.1441>
- Levitt, S. D., Miles, T. J., & Rosenfield, A. M. (2012). Is Texas Hold 'Em a Game of Chance? A Legal and Economic Analysis. *The Georgetown Law Journal*, 101(581), 581–636.
- Ma, J., Lu, Y., & Gupta, S. (2019). User innovation evaluation: Empirical evidence from an online game community. *Decision Support Systems*, 117, 113–123. <https://doi.org/10.1016/j.dss.2018.11.003>
- Mollick, E., & Nanda, R. (2016). Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. *Management Science*, 62(6), 1533–1553. <https://doi.org/10.1287/mnsc.2015.2207>
- O'Leary, D. E. (2017). Crowd Performance in Prediction of the World Cup 2014. *European Journal of Operational Research*, 260(2), 715–724. <https://doi.org/10.1016/j.ejor.2016.12.043>
- Otterbacher, J. (2009). “Helpfulness” in Online Communities: A Measure of Message Quality. *Proceedings of the 27th Conference on Human Factors in Computing Systems (CHI) 2009*. <https://doi.org/10.1145/1518701.1518848>
- Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17–29. <https://doi.org/10.1016/j.ijforecast.2017.08.002>
- Peurala, J. (2013). Match-manipulation in football - The challenges faced in Finland. *International Sports Law Journal*, 13, 268–286. <https://doi.org/10.1007/s40318-013-0027-z>

- Princeton University. (2019). *WordNet*.
<http://wordnetweb.princeton.edu/perl/webwn?o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&s=hurt>
- Ratten, V. (2017). *Sports Innovation Management* (1st ed.). New York, New York: Routledge.
- Ratten, V. (2011). Sport-based entrepreneurship: Towards a new theory of entrepreneurship and sport management. *International Entrepreneurship and Management Journal*, 7, 57–69. <https://doi.org/10.1007/s11365-010-0138-z>
- Rhyn, M., & Blohm, I. (2019). Patterns of Data-Driven Decision-Making: How Decision-Makers Leverage Crowdsourced Data. *Proceedings of the 40th International Conference on Information Systems (ICIS) 2019*.
- Riedl, C., Blohm, I., Leimeister, J. M., & Krcmar, H. (2013). The Effect of Rating Scales on Decision Quality and User Attitudes in Online Innovation Communities. *International Journal of Electronic Commerce*, 17(3), 7–36. <https://doi.org/10.2753/jec1086-4415170301>
- Schumaker, R. P., Jarmoszko, A. T., & Labedz, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems*, 88, 76–84. <https://doi.org/10.1016/j.dss.2016.05.010>
- Simmons, J. P., Nelson, L. D., Galak, J., & Frederick, S. (2011). Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds. *Journal of Consumer Research*, 38(1), 1–15. <https://doi.org/10.1086/658070>
- Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York, New York: Anchor.
- van Dijk, H., van Engen, M. L., & van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, 119(1), 38–53. <https://doi.org/10.1016/j.obhdp.2012.06.003>
- Vermunt, J. K. (2005). Mixed-Effects Logistic Regression Models for Indirectly Observed Discrete Outcome Variables. *Multivariate Behavioral Research*, 40(3), 281–301. https://doi.org/10.1207/s15327906mbr4003_1
- Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, 23, 2538–2557. <https://doi.org/10.1016/j.chb.2006.05.002>

- Wambsganss, T., & Fromm, H. (2019). Mining User-Generated Repair Instructions from Automotive Web Communities. *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS) 2019*. <https://doi.org/10.24251/hicss.2019.144>
- Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5–34. <https://doi.org/10.1080/07421222.1996.11518099>
- Weimer, M., & Gurevych, I. (2007). Predicting the Perceived Quality of Web Forum Posts. *Proceedings of the 6th Conference on Recent Advances in Natural Language Processing (RANLP) 2007*.

Publication IV

Going Global: Enhancing International Social Media Fan Engagement – Evidence from the German Bundesliga



Author(s)

Arne Grüttner

Laurenz O. H. Haferbeck

Roman Rietsche

Andrea Back

Submitted to the Proceedings of the 29th European Conference on Information
Systems (ECIS), 2021.

Abstract

The unbreakable boundary between the physical and online world in sports is getting blurry. In turn, the glamour of sports moves to a global digital spotlight. Sports organizations try to seize this opportunity to digitally attract international fans by utilizing social media. However, they currently lack an understanding of the sports content drivers to enhance fan engagement and how to interact with international fans due to cross-cultural effects. We build upon consumer engagement theory and link sports organizations' brand attributes with the concept of visual emotional and informational appeals. Our experiment demonstrated: it is crucial for sports organizations to understand how to adjust visual appeals to maximize content effectiveness, as our analyses yielded evidence that emotional appeals have a significant positive effect on fan engagement. Furthermore, the results showed that the relationship between sports content, visual appeals, and fan engagement is significantly moderated by cross-cultural and media type effects.

Keywords: cross-cultural effect, fan engagement, media type, social media engagement, social media usage, sports content, sports digitization, visual appeal

Introduction

Professional sports has become a billion-dollar business that attracts millions of sports consumers, so-called fans, around the globe. For instance, in the 2017/2018 season of the German football Bundesliga around 550,000 fans a week attended the match of their favorite club live in the stadium (Deutscher Fußball-Bund [DFB], 2018). However, in recent years, the seemingly unbreakable boundary between the physical world and the online world in sports has been getting blurry. In turn, the glamour of sports moves more and more towards a global digital spotlight. For instance, as noted by Nasser Al-Khelaifi (President of Paris Saint-Germain Football Club), "The world today is a direct, digital world. It has taken 50 years to make Real Madrid a world club. Now you can do it in five years" (11Freunde, 2020, p.4). Sports organizations, such as associations and clubs, try to seize this opportunity to digitally attract new international fans beyond their traditional local fan base by utilizing social media. Social media removes geographical barriers of traditional media outlets, so they transcend both temporal and geographical constraints to enable sports organizations to directly engage with fans on a worldwide basis outside of a traditional home market. As a result, social media has enabled novel marketing strategies that offer innovative experiences and interactions to build and

deepen unique relationships between sports organizations and international fans (Xiao et al., 2017).

Social media offers sports organizations the freedom to design and communicate a broad spectrum of sports content (e.g., fan stories, player portraits, or sports highlights, and news) in a variety of different media types (e.g., texts, photos, or videos). Previous research from related literature streams has shown that such multimedia content can be instrumental on social media because content is an important driver of consumer engagement (Ordenes et al., 2019). In a sports context, it was demonstrated that high fan engagement on social media impacts key organizational performance outcomes such as ticket and merchandise sales and fan loyalty (Vale & Fernandes, 2018). However, while the freedom to use multimedia content on social media is favorable to enhance fan engagement in general, there is limited understanding of the sports content drivers that can enhance fan engagement. For instance, does a video that shows a goal scored by a world-class striker lead to higher fan engagement than a photo that shows the flag-waving fans of a club on social media? On top of that, although many sports organizations face the challenge of globalizing their businesses and researchers call for such investigations, sports organizations currently lack an understanding of how to interact with new international fans due to cross-cultural effects such as a geographical heterogeneity and cross-cultural behaviors (e.g., attitudes and values) (e.g., Achen et al. (2018), Romney and Johnson (2020), or Vale & Fernandes (2018)). Yet, research on social media content drivers and their impact on a global level is still scarce and needs further exploration – in research in general (Akpınar & Berger, 2017; Rietveld et al., 2020) and in the sports context in particular. Therefore, we tackle the following two research questions (RQs) in this paper:

RQ1: *What are the sports content drivers to enhance fan engagement on social media?*

RQ2: *How do cross-cultural effects moderate the relationship between sports content drivers and fan engagement?*

We conducted a joint research project with the DFL Digital Sports GmbH (DFL DS). DFL DS is responsible for the entire social media content of the German Bundesliga, including the Bundesliga and Bundesliga 2 leagues. We build upon consumer engagement theory (CET) as proposed by Pansari and Kumar (2017) and link sports organizations' brand attributes with the concept of visual emotional and informational appeals and tested our research model in an experiment using a between-subjects design with four treatments (N = 167). For researchers, we propose a unique research model to

academia that should be extended and adjusted to different contexts. For social media practitioners, our study provides guidelines on how to create influential social media content to maximize content effectiveness. The remainder of this paper is structured as follows: Section 2 introduces the theoretical background and related work. Our research model and hypotheses as well as the method are described in Section 3. Section 4 presents our findings as well as the discussion. We then propose the implications, avenues for future research, and the limitations in Section 5. Finally, we conclude our paper in Section 6.

Theoretical Background & Related Work

Fan Engagement on Social Media

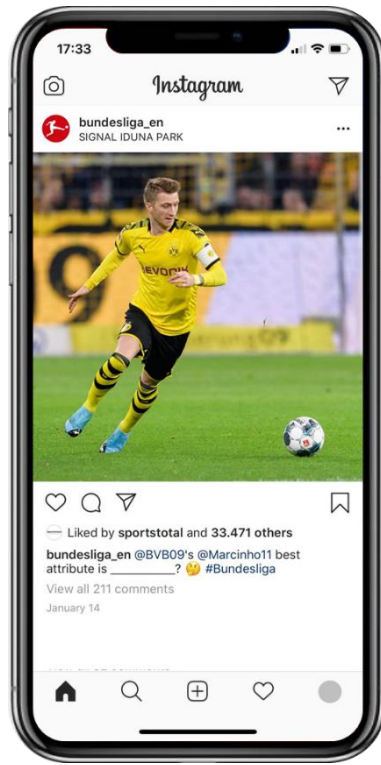
The advent of social media has had a profound impact on how people consume sports, as it offers an interactive channel between fans and the sports itself. We define social media as being made up of various user-driven platforms, so-called social networking sites (SNSs) (e.g., Facebook, Instagram, or TikTok), that facilitate a diffusion of compelling content, dialog, creation, and communication to a broader audience (Kapoor et al., 2018). Social media offers sports organizations the possibility to design and communicate a broad spectrum of sports content in a variety of different media types. Previous research has shown that such multimedia content can be instrumental on social media because content is an important driver of consumer engagement (Ordenes et al., 2019). Engagement on social media has been studied extensively by practitioners as well as in various academic disciplines including Information Systems (IS) research in recent years (e.g., Liu et al. (2020)). However, there is no common consensus on how engagement should be defined in academia yet. As a result, there are variations in the nomenclature as well as in its conceptualization of engagement. In this paper, we draw on CET, which was proposed by Pansari and Kumar (2017) to provide a holistic definition of consumer engagement. The theory suggests that consumers' brand- or firm-related experiences influence their emotional or affective states, which then influence the nature of their direct and indirect engagement with firms (Pansari & Kumar, 2017). For example, the viewing experience of advertising can lead to consumers' positive attitude towards a product, which, in turn, leads to direct engagement that can occur in the form of purchases or to indirect engagement in the form of social media conversations consumers have about the brand.

Consumer engagement has proven to be beneficial for organizations in different practical and theoretical endeavors. For instance, a positive impact on firm performance has been reported in terms of financial returns and shareholder value (e.g., Rietveld et al. (2020)). In a sports context, it was demonstrated that high fan engagement on social media impacts organizational performance outcomes such as brand awareness, ticket and merchandise sales, and fan loyalty (Vale & Fernandes, 2018). Accordingly, it is important for sports organizations to understand what kind of social media sports content enhances fan engagement. Researchers have investigated both (1) “what” content sports organizations post on social media (e.g., Aichner (2019)) and (2) “why” – the motives – fans consume sports content on social media (e.g., Vale and Fernandes (2018)). For a detailed literature review on social media in the field of sports, see, Filo et al. (2015). However, although many sports organizations currently face the challenge of globalizing their businesses and researchers call for such investigations, little research has elaborated on the social media sports content drivers that lead to high fan engagement and in particular in an international sports context in which fans differ due to cross-cultural effects such as a geographical heterogeneity and cross-cultural behaviors (e.g., attitudes and values) (e.g., Achen et al. (2018), Romney and Johnson (2020), or Vale and Fernandes (2018)).

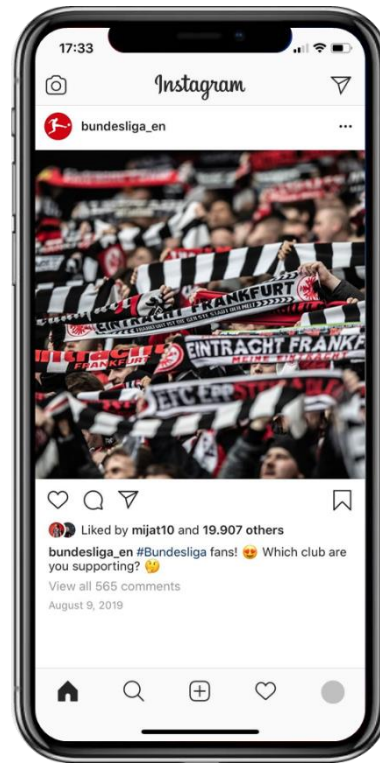
To close this gap, we build upon Parganas et al. (2015), who developed a sports organization-branding model that is based on Keller's (1993) well-known hierarchy of brand associations (i.e., attributes, attitudes, and benefits). Parganas et al. (2015) argued that sports organizations contain specific brand attributes that can be controlled and used in order to influence fan engagement. Their model is constituted of two categories that are adjusted to the context of this study as follows: (1) product-related attributes – that is, those attributes that originate from the actual game on-site the field of play, for example, player skills or match highlights (see, Figure 1); (2) non-product-related attributes that originate from off-site the actual field of play and are relevant for the consumption of sports, such as figures and statistics or fan cultures (see, Figure 2). In this paper, we link this distinction between product-related and non-product-related attributes to social media sports content categories to identify social media sports content drivers that can enhance fan engagement.

Figure 1

Player Post From the Product-Related Sports Content Category

**Figure 2**

Fan Culture Post From the Non-Product-Related Sports Content Category



Visual Appeals as Drivers for Social Media Fan Engagement

The past few years have witnessed a shift in social media from text-centric to visually oriented content (i.e., photos and videos). According to the extensive visual advertising literature, visual components can affect cognitive (e.g., attention, attitude, or preference) and behavioral (e.g., clicks, purchase intention, or sales) processes (Li & Xie, 2020). To examine and to understand what kind of visual content enhances consumer engagement on social media, scholarly research from the marketing discipline has adopted and utilized emotional and informational visual appeals included in social media content (e.g., Akpınar and Berger (2017), Li and Xie (2020), Rietveld et al. (2020), or Yoo and MacInnis (2005)). Visual appeals operate under the assumption that message content in terms of its emotional and informational appeals emerges as a key driver of engagement behavior. We adopt the emotional versus informational appeals distinction and define emotional appeals as visual post content designed to invoke consumers' emotions. Emotional appeals are designed to make the consumer feel good about the brand or product and can lead to positive reactions (Goldberg & Gorn, 1987). Therefore, emotions can motivate and persuade consumers and often guide attitude formation and consumer behavior (Bagozzi et al., 1999). Informational appeals, on the other hand, are

defined as visual post content designed to appeal to a consumers' rationality by providing information on the benefits and attributes of brands and products (Yoo & MacInnis, 2005). Informational appeals are designed to invoke cognitive processing by using objective information describing a brand's benefits (MacInnis et al., 2002). Informational appeals aim to change a consumer's brand beliefs based on arguments conveyed in the message content (MacInnis et al., 2002; Rietveld et al., 2020). The distinction between emotional and informational appeals has been proven to offer a meaningful lens for studying the effect of message content on engagement behavior in different offline and online applications. For instance, Akpinar and Berger (2017) found that online ads with strong emotional appeals are more likely to be shared, while online ads with informational appeals drive brand evaluations and purchase. Similarly, Lee et al. (2018) found that adjusting emotional and informational appeals in content on Facebook leads to different brand-related outcomes. Two example photos with different emotional and informational visual appeals in a sports context are depicted in Figure 3 and Figure 4.

Given that visual appeals can be controlled and adjusted by sports organizations on social media, it is worth studying them to understand how sports content influences fan engagement. In this vein, Romney and Johnson (2020) performed a content analysis of photos of sports networks (e.g., ESPN and FOX Sports) on Instagram. The authors found that photos that contained emotional messages resulted in great interest and engagement by fans. However, our understanding of emotional and informational visual appeals in social media sports content is still limited. Therefore, we link social media sports content as described in the aforementioned section with the concept of emotional and informational visual appeals to identify sports content drivers that enhance fan engagement.

Figure 3*Sports Content with Emotional Appeals***Figure 4***Sports Content with Informational Appeals*

MATCHDAY 14 

RANK	TEAM	PLD	GD	PTS
1	 BORUSSIA M'GLADBACH	14	14	31
2	 RB LEIPZIG	14	23	30
3	 BORUSSIA DORTMUND	14	14	26
4	 FC SCHALKE 04	14	7	25
5	 SPORT-CLUB FREIBURG	14	7	25
6	 BAYER 04 LEVERKUSEN	14	4	25
7	 FC BAYERN MÜNCHEN	14	15	24
8	 TSG HOFFENHEIM	14	-4	21
9	 VFL WOLFSBURG	14	1	20
10	 1. FC UNION BERLIN	14	-1	19
11	 EINTRACHT FRANKFURT	14	-2	18
12	 FC AUGSBURG	14	-6	17
13	 1. FSV MAINZ 05	14	-14	15
14	 SV WERDER BREMEN	14	-7	14
15	 HERTHA BERLIN	14	-9	12
16	 FORTUNA DÜSSELDORF	14	-13	12
17	 SC PADERBORN 07	14	-15	8
18	 1. FC KÖLN	14	-18	8

Cross-Cultural and Media Type Effects in Social Media Fan Engagement

Professional sports has become a global phenomenon. Hence, acknowledging the potential of cross-cultural effects is crucial for understanding the sports content drivers of fan engagement. The concept behind culture is not easy to define, although research on culture is abundant in academia today. The most widely used definition of culture in a national context stems from Geert Hofstede. Hofstede defines national culture as “its essence is collective mental programming: it is that part of our conditioning that we share with other members of our nation, region, or group but not with members of other nations, regions, or groups” (Hofstede, 1983, p. 76). Today, the concept of culture has been elaborated on in relation to dozens of nations and in dozens of different offline and online contexts. As a result, research commonly supports the view that online cultures mirror the offline cultures of which they are a product (Jackson & Wang, 2013). Fans can also be segmented culturally, socially, and demographically in terms of their engagement behaviors. This view is shared by Bouzdine-Chameeva et al. (2015), who stated that a population of fans is heterogeneous, with fans of different backgrounds or preferences differing fundamentally in the way they consume sports. For example, fans are different and distinct in their practices and cultures. Fans may exhibit distinct fandom levels, ranging from simply watching games to assuming the successes and losses of their club as their own (often referred to as self-identity). Thus, the degree of passion and emotional significance distinguishes fans (Fillis & Mackay, 2014). In this vein, Hu and Cole (2016) expressed that fans vary in the intensity with which they

support their favorite team, with deeply loyal and ardent fans usually seeking very particular platforms to converse with. With these results in mind, it can be said that fans exhibit different levels of social and cultural attachment to sports. It can be assumed that these differences are even multiplied in an international fan context as the internet growth continues to shape fandom.

Similar to the cultural effects on social media fan engagement, previous research has shown that the media type in which multimedia content is presented on social media has a significant effect on engagement outcomes such as clicks, purchase intention, or sales. For example, Li and Xie (2020) found a significant and robust positive mere presence effect of photo content on user engagement on Twitter. The authors also found that high-quality and professionally shot pictures consistently lead to higher engagement. In a social media sports context, these findings raise questions, such as ‘Does a video that shows a world-class goal scored by a striker (i.e., sports content from the product-related sports content category) lead to a higher fan engagement than a photo that shows the flag-waving fans of a club (i.e., non-product-related sports content category)?’. To answer such questions, we included both cultural and media type effects in our research model, which is described in detail in the upcoming section.

In sum, CET, as described by Pansari and Kumar (2017), serves as an overarching theoretical perspective to ground our conceptual framework in. We furthermore build on related research into sports organizations’ brand attributes (Parganas et al., 2015) and the concept of emotional and informational visual appeals (Yoo & MacInnis, 2005). From a CET perspective, we argue that (1) sports organizations publish sports content on social media; (2) fans’ experiences with these content will be associated with fans’ affective states; (3) fans’ affective states will, in turn, influence the nature of their digital engagement with the content; (4) by managing visual emotional and informational appeals in their social media sports content, sports organizations can influence fans’ digital engagement; (5) this relationship between visual appeals and fan engagement is moderated by cross-cultural and media type effects. In other words, through their social media activities, sports organizations can send specific stimuli that lead to higher fan engagement taking into account that fans’ engagement behavior on social media is infused with their cultural values and influenced by media type effects.

Research Model and Hypotheses

Hypotheses Development

For the development of our research model (see, Figure 5), the following hypotheses are based on the concepts described in Section 2 as well as on related literature streams and are explained as follows.

According to Parganas et al. (2015), sports organizations contain specific brand attributes that can be controlled and used in order to influence fan engagement. In this paper, we follow the authors and classify social media sports content according to these attributes. In detail, we expect that product-related and non-product-related sports content triggers different levels of fans' emotional and informational appeals. We assume that sports content that originates from the actual game on-site the field of play (i.e., product-related sports content) and shows, for example, skills or match highlights has a strong positive effect on the emotional appeals of fans, as this kind of sports content invokes fans' emotions. Hence, we state the following hypothesis:

H1: *Product-related sports content has a stronger positive effect on emotional appeals than non-product-related sports content.*

Similar to the abovementioned discussion about the emotional appeals, we assume that non-product-related sports content, that is, content that originates off-site the field of play and includes sports content related to figures and statistics or information about fan cultures, has a strong positive effect on the informational appeals of fans. Reasons for our hypothesis are that such a kind of sports content appeals to fans' rationality and thus invokes cognitive processing. Consequently, we pose the following hypothesis for the non-product-related sports content:

H2: *Non-product-related sports content has a stronger positive effect on informational appeals than product-related sports content.*

In our study, emotional appeals are defined as visual post content designed to invoke fans' emotions. Emotional appeals have been proven to affect consumer behavior and enhance engagement in different offline and online contexts. For instance, Goldberg and Gorn (1987) demonstrated that emotional appeals can lead to engagement in the form of positive reactions by showing emotional TV commercials. Similarly, in Yoo and MacInnis (2005) it was demonstrated that ads with an emotional ad format strengthen positive feelings and thereby enhance the brand attitude format process positively. In an online social media context, Wang et al. (2019) found that emotional social media

content is a strong motivator for followers to repost content. Similarly, Rietveld et al. (2020) demonstrated that visual emotional and informational appeals encoded in brand-generated content influence consumer engagement in terms of likes and comments on Instagram. Following these previous findings, visual sports content that is composed of affective emotional appeals evokes arousal and thus is likely to affect fans' digital engagement. That should be especially true in a sports context in which fans are mainly driven by emotions such as passion and social values rather than by rational evaluations (Gruettner, 2019). Therefore, one should expect that sports content with high emotional appeals is positively affecting fan engagement:

H3: *Visual sports content with high emotional appeals has a positive effect on fan engagement.*

Informational appeals are defined as visual post content designed to appeal to a consumers' rationality. On the one hand, evidence for a positive effect of informational appeals on engagement behavior has been shown with large, diverse samples (e.g., Akpınar and Berger (2017), Li and Xie (2020), or Yoo and MacInnis (2005)). These studies commonly report that content with high informational appeals can provide relevant information about a product, its use, and its benefits, thereby helping consumers to better evaluate it and thus leading to direct engagements in the form of, for instance, consumer purchases. On the other hand, previous findings contrast that a high informational effect can also have a negative effect on consumer engagement (e.g., Rietveld et al. (2020)). In this vein, Muntinga et al. (2011) argued that informational appeals may signal a persuasion attempt, which is incongruent with motivations to engage with brands on social media. In the context of this study, we are not dealing with consumer purchase decisions. In contrast, the sports content posted by sports organizations on social media aims to raise emotions, general interest, or to educate and inform fans (e.g., about performance data, standings, or transfers and injuries). As such data often changes on a daily/weekly basis, this information becomes highly relevant for fans. In this vein, various studies have shown that information gathering is one key motivation for fans to follow sports accounts on social media (e.g., Vale and Fernandes (2018)). Following the argumentation above, we assume that sports content with high informational appeals positively affects fan engagement:

H4: *Visual sports content with high informational appeals has a positive effect on fan engagement.*

Fans can be segmented culturally, socially, and demographically in terms of their engagement behaviors (see, Section 2.3). Therefore, we hypothesize that the relationship

between visual emotional/informational appeals and fan engagement is moderated by cross-cultural effects. Our reasoning is as follows: Consumer segments and cultural effects in online contexts have been investigated. Researchers commonly report that online cultures mirror offline cultures (Jackson & Wang, 2013). In a social media sports context, Billings et al. (2019) found that fans from China and the United States use social media for starkly contrasting motives. While fans from China used social media for camaraderie, entertainment, and maintaining relationships, fans from the United States used social media to express their fandom and to gain information. These results are in line with other studies such as Jackson and Wang (2013). In the context of this study, we assume similar results, that is, that cultural effects on the motives (e.g., entertainment, information, integration, and social interaction, or personal identity motives (see, Vale and Fernandes (2018)) to consume sports and the accompanied different levels of fandom will moderate how strongly fans perceive emotional and informational appeals in social media sports content and, in turn, have an effect on fan engagement. Accordingly, we pose the following hypotheses:

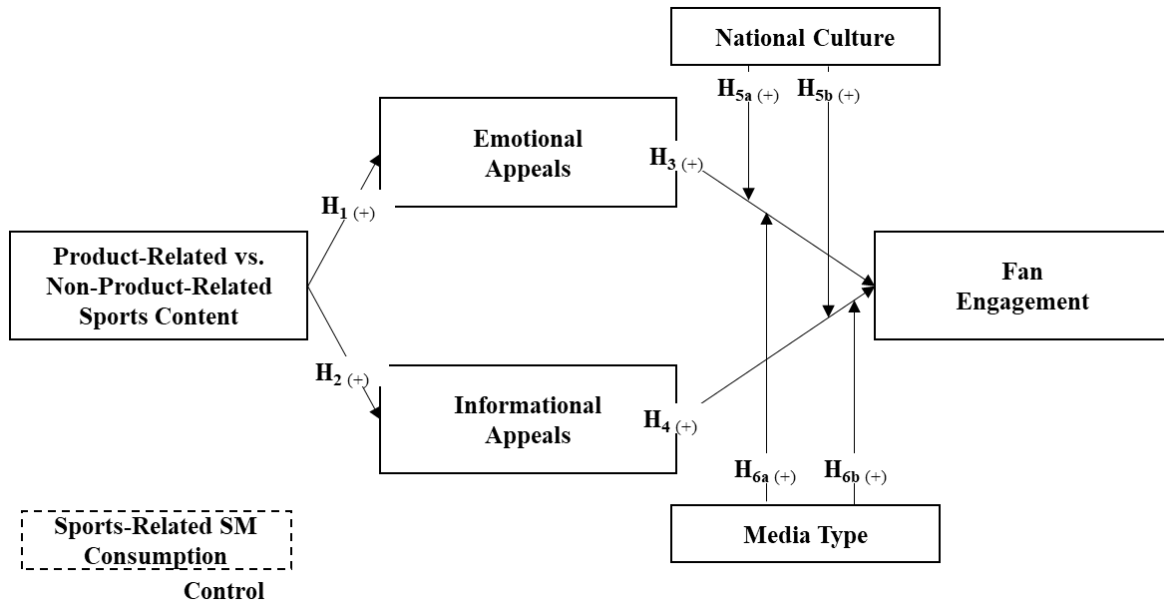
H5a: *The relationship between sports content, visual emotional appeals, and fan engagement is moderated by cross-cultural effects.*

H5b: *The relationship between sports content, visual informational appeals, and fan engagement is moderated by cross-cultural effects.*

Previous research has shown that multimedia content unleashes its emotional and informational appeals differently depending on the type of media (i.e., texts, photos, or videos) in which the content is presented. Therefore, we assume that media types moderate the relationship between sports content, emotional and information appeals, and fan engagement. Evidence for this assumption can be found, for instance, in Li and Xie (2020) who demonstrated that different media types have a significant effect on cognitive (e.g., attention, attitude, or preference) and behavioral engagement outcomes (e.g., clicks, purchase intention, or sales). Consequently, we include the media type in which the sports content is presented in our research model that moderates the relationship between sports content, emotional and informational appeals, and fan engagement:

H6a: *The relationship between sports content, visual emotional appeals, and fan engagement is moderated by media type.*

H6b: *The relationship between sports content, visual informational appeals, and fan engagement is moderated by media type.*

Figure 5*Research Model*

Method

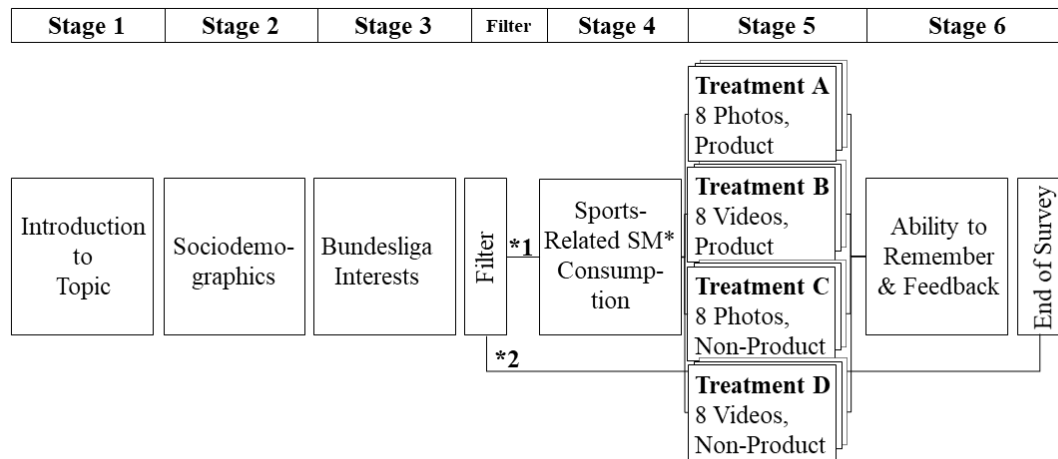
Experimental Design and Manipulations

This study is conducted in a joint research project with DFL DS. DFL DS is a subsidiary of the German Football League (DFL), the German national football federation. DFL DS is responsible for and operates the entire social media content of the German Bundesliga, including the Bundesliga and Bundesliga 2 leagues. Therefore, we chose the German Bundesliga as the research context of our study. The defined research model and hypotheses are tested in an experiment using a between-subjects design with four treatments. Subjects to collect data for validating or rejecting the hypotheses (H_1 to H_6) were collected through the official German Bundesliga channels (i.e., “www.Bundesliga.com”, “Bundesliga Official” on Facebook, and “bundesliga_en” on Instagram). In 2019, the official Bundesliga website (“www.Bundesliga.com”) had an active monthly user base of 1.74 million on average. “Bundesliga Official” recorded 7.7 million followers and “bundesliga_en” had 5.3 million followers at the time of our analysis (the numbers are as of 26 April 2020). To target subjects for our experiment, a banner/post targeting American and Indian citizens was shared on the abovementioned channels between 24 April 2020 and 29 April 2020 in which subjects were asked to take part in a survey to help the German Bundesliga improve their social media content. We chose India and the United States for two reasons. First, English is the official language

in both countries. Therefore, it is guaranteed that we can use the same sports content for all subjects of the experiment. Second, both countries are international emerging markets with respect to their interest in football and are therefore interesting for sports organizations such as the DFL DS. We refrained from recruiting subjects, for example, through a lottery, because we wanted to have intrinsically motivated participants who could deliver reliable results. Subjects were then redirected to an external website to take part in the experiment. In the experiment, subjects were asked to evaluate social media sports content that had been previously posted on the official German Bundesliga social media channels. We only presented the concrete photos and videos of the previously published content and excluded captions and comments of other users. This guaranteed that we could measure the specific effects represented in our research model. The experiment consisted of six stages, which are described as follows and depicted in Figure 6.

Figure 6

Experimental Design



*1: I consumed sports content on social media in the last month.

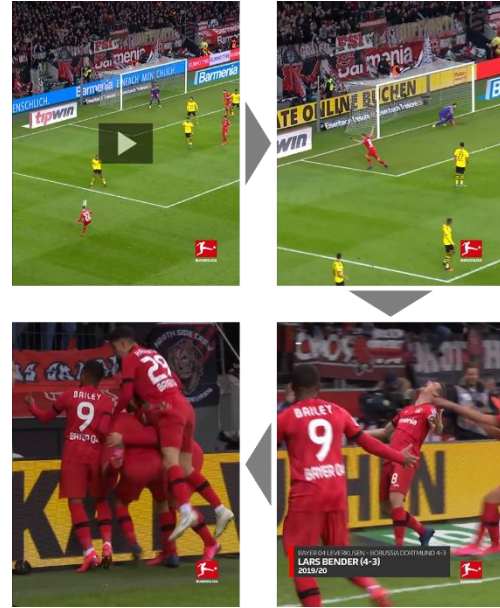
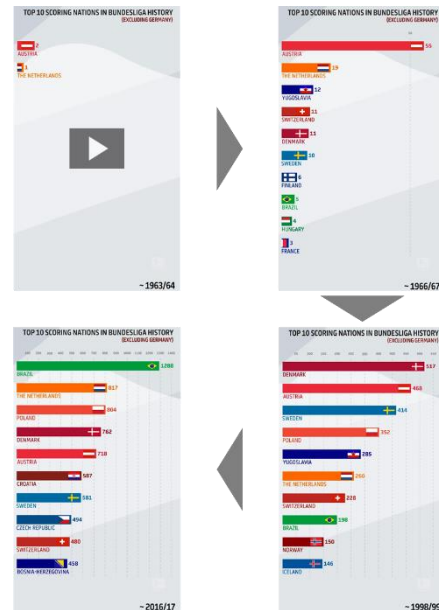
*2: I did not consume sports content on social media in the last month. *SM = Social media

In the first stage, subjects are introduced to the topic of the experiment as follows: “You are confronted with a variety of social media sports content (e.g., fan stories, sports highlights, sports news, or sports statistics) on social media every day. In our project, we want to investigate what kind of sports content fans are most interested in consuming via social media”. The second stage of the experiment asked for the sociodemographic characteristics of the subjects. In the third stage, subjects answered initial questions to measure their interest in the German Bundesliga. Then, we implemented a filter that asked whether subjects had consumed sports content on social media in the last month. This filter guaranteed that only subjects were included that consume sports content on

social media. In the fourth stage, subjects were asked to answer questions to measure their sports-related social media consumption. In the fifth stage, subjects were randomly assigned to one out of four treatments. In all of the four treatments, subjects were, firstly, introduced to the concept of emotional and informational visual appeals as described in Akpinar and Berger (2017). In a second step, subjects were asked to imagine the following situation: “You are browsing through your social media networks. While doing so, you are seeing different content from your friends and family but also from the accounts you are following and commercial ads. In the following, you will be shown eight photos (Treatment A and C) or eight videos (Treatment B and D)¹ that show up on your screen. Please carefully observe all the eight photos/videos and then answer the questions”. In Treatment A and B, subjects were exposed to either eight photos from the product-related sports content category (Treatment A; see, Figure 7) or to eight videos from the product-related sports content category (Treatment B; see, Figure 8). Similarly, in Treatment C and D, subjects were exposed to either eight photos from the non-product-related sports content category (Treatment C; see, Figure 9) or to eight videos from the non-product-related sports content category (Treatment D; see, Figure 10). To test H1 and H2, subjects were then asked to answer questions regarding the emotional and informational visual appeals contained in the photo/video. A similar approach is used and suggested in scholarly science (see, Akpinar and Berger (2017), Rietveld et al. (2020), or Yoo and MacInnis (2005)). Afterwards, to test H3 and H4, subjects had to indicate whether they would “like” the presented photo/video on social media. The last stage of the experiment asked subjects which of the photos/videos they remembered and to give general feedback on the content shared by the German Bundesliga.

In order to assign photos and videos to the product-related or to the non-product-related sports content category, two senior researchers and a practitioner from DFL DS’s social media content team coded photos and videos independently and afterwards discussed their affiliation to the attributes developed by Parganas et al. (2015). Since we targeted subjects whose nationality is either American or Indian, the language of the experiment is English. Furthermore, we paid special attention to ensure that all pictures and videos shown in the experiment were written in English or at least had English subtitles.

¹ The concrete photos and videos used in our experiment can be found here: Gruettner and Haferbeck (2020).

Figure 7*Photo Content in Treatment A***Figure 8***Video Content in Treatment B***Figure 9***Photo Content in Treatment C***Figure 10***Video Content in Treatment D*

Measurement Model

The experiment items and validated scales used to measure the research variables are based on constructs from scholarly literature (see, Table 1). The measurement items have been adjusted to the experiment context and are described as follows: Items for the informational and emotional appeals construct were adopted from Yoo and MacInnis (2005) and were measured on a 7-point Likert scale with anchors of “Strongly Disagree

(1)” and “Strongly Agree (7)”. Subjects were asked to indicate (1) to what extent the individual photo/video appeals to their emotions and (2) whether the photo/video provides a lot of information. The measurement for social media fan engagement is adopted from Schivinski et al. (2016). The authors distinguish between three dimensions of consumer engagement with brands on social media: Consumption, contribution, and creation. We only included items from the contribution dimension to the experiment, as the other dimensions of social media engagement did not fit the context of our study. We asked whether subjects would “like” the presented photo/video. Items for the social media fan engagement construct were measured on a 7-point Likert scale anchored by “not very often (1)” and “very often (7)”. The subjects could also select the option “not at all” (coded as 0). We used the nationality of the subjects as a proxy to operationalize the moderating national culture construct as described by Hofstede (1983). Therefore, a dummy variable was created. A similar approach was used in scholarly research such as Hudson et al. (2016). Likewise, for the social media sports content categories (see, Parganas et al. (2015)) as well as for the media type moderator we created dummy variables. In addition to the research variables, we added items for our control variable to our experiment, that is, sports-related social media consumption, as described by Schivinski et al. (2016). All scales and items used in our experiment were carefully worded to fit the specific situation of our experimental sports context rather than being unspecific to fit into any social media context. This ensures that the items correspond to the concept of social media engagement in a sports context and thus guarantees construct validity.

Table 1*Research Variables, Definitions, and Sources Used in our Research Model*

Construct	Definition	Source
Research Variables		
Product-related Sports Content	Social media sports content that originates from the actual game on-site the field of play. For example, player skills or match highlights.	Parganas et al. (2015)
Non-Product-related Sports Content	Social media sports content that originates from off-site the actual field of play. For example, figures and statistics or information about fan cultures.	Parganas et al. (2015)
Emotional Appeals	Emotional appeals are defined as visual post content designed to invoke fans' emotions.	Yoo and MacInnis (2005)
Informational Appeals	Informational appeals are defined as visual post content designed to appeal to a fans' rationality.	Yoo and MacInnis (2005)
Fan Engagement	Subject's attitude towards whether it would "like" the presented photo/video on social media.	Schivinski et al. (2016)
National Culture	Dummy variable that indicates the subject's nationality. Indian citizens coded as (1) and citizens from the United States as (0).	Hofstede (1983)
Media Type	Dummy variable that indicates the media type in which the sports content is presented. Photo coded as (1) and video coded as (0).	-
Control Variables		
Sports-related Social Media Consumption	Refers to the subject's general sports content consumption on social media. For example, how often a subject consumes sports content on social media.	Schivinski et al. (2016)

Data Analysis

A total of 2,814 subjects clicked on the banner/post which was sent out on the German Bundesliga channels. 342 subjects (12.15% of the subjects who saw the starting page of the experiment) participated in the experiment and completed it. The usability of received responses for further analysis is judged on several include/exclude criteria: First, 89 answers were excluded because they did not match the nationality requirements (i.e., American or Indian). Second, 32 subjects were removed due to the filter that was embedded in the experiment. Hence, only subjects who stated that they have consumed sports content on social media in the last month were considered. Third, to prevent distortion of experimental results, subjects who selected the same response to different statements/items for all eight photos or videos in a row (so-called straight-lining; e.g., a subject always responds with a seven, irrespective of the questions) were removed. As a result, 14 subjects were excluded due to suspicious data patterns that indicate a monotonous click-through without the variability of responses. Fourth, a focus was laid on the subjects' answers on emotional and informational appeals. If the emotional and informational appeals were rated six or more times with the same value for the eight presented photos or videos, the subjects were removed. This resulted in an exclusion of further 40 responses. A total of 167 (American: 76; Indian: 91) responses remained in our data sample that met the quality requirements and can thus be used for further

analysis (36 subjects of Treatment A; 29 subjects of Treatment B; 69 subjects of Treatment C; 33 subjects of Treatment D). To investigate our proposed hypotheses H1 to H6, we applied a multiple ordinary least squares (OLS) regression. For the emotional and informational visual appeals constructs as well as for the social media fan engagement construct, we calculated the mean of the eight presented photos/videos presented for each subject. For the moderation of the effect, we used the programming language Python 3.4 and model 17 in PROCESS which is proposed by Hayes (2012).

Results and Discussion

Descriptive Statistical Results

In total, 167 subjects are included in our data sample for further analyses after applying the include/exclude criteria as described in Section 3.2. An overview of the descriptive statistics and the number of subjects of each treatment can be found in Table 2. Subjects who are included in the final sample needed 16.27 minutes on average to complete the experiment. The majority of the subjects are male (female = 5.95%; male = 92.86%; one subject preferred not to answer this question). 46.43% of the subjects are between 18 and 24 years old. Subjects in Treatment A and B who are exposed to product-related sports content responded whether the presented content appeals to their emotion at a mean of 4.67. In contrast, the answers of subjects in Treatment C and D (i.e., non-product-related sports content) have a mean of 4.18. Subjects of Treatment A and B answered with a mean of 3.56 that the displayed content provides a lot of information (Treatment C and D 3.86). The answers to the question whether the subjects would like a photo/video scored a mean of 4.73 in the product-related sports content category (Treatment A and B) and a mean value of 4.56 in the non-product-related sports content category (Treatment C and D).

Table 2*Descriptive Statistics of the Data Sample*

Variables	Product-Related				Non-Product-Related			
	Treatment A Photo		Treatment B Video		Treatment C Photo		Treatment D Video	
Demographics N = 167	n = 36		n = 29		n = 69		n = 33	
Nationality India (n = 91)	19 (11.38%)		14 (8.38%)		39 (23.35%)		19 (11.38%)	
USA (n = 76)	17 (10.18%)		15 (8.98%)		30 (17.96%)		14 (8.38%)	
Research Variables	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Emotional Appeals Ø	4.53	1.11	4.84	1.09	4.06	1.01	4.44	0.94
Informational Appeals Ø	3.43	1.27	3.72	1.17	3.66	0.82	4.28	0.98
Fan Engagement Ø	4.56	1.29	4.95	1.27	4.42	1.41	4.84	1.34

SD = Standard deviation

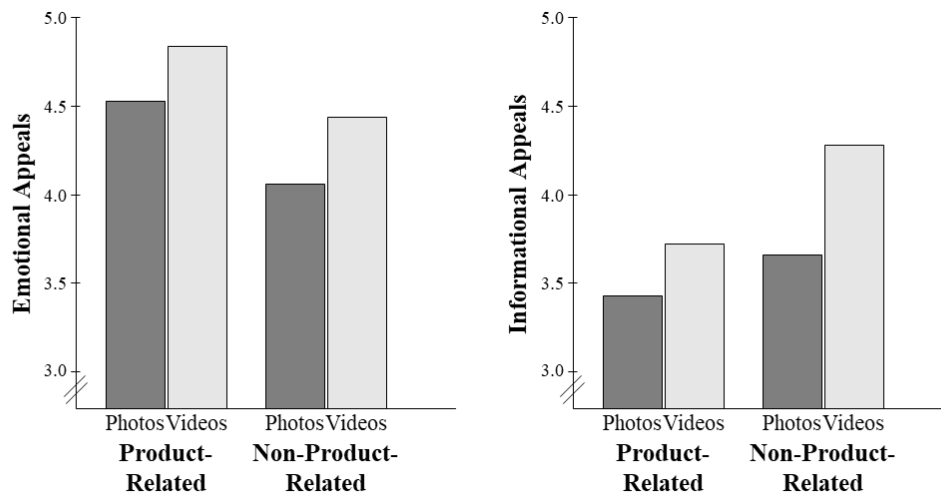
Evaluating the Effect of Social Media Sports Content on Visual Appeals (H1 – H2)

The statistically significant results of the performed analyses to test our hypotheses are depicted in Figure 12. In H1, we hypothesized that product-related sports content has a stronger positive effect on emotional appeals than non-product-related sports content. Our results confirm this hypothesis ($\beta = .50$, $t(167) = 3.01$, $p = .003$). We assume that skills and match highlights that originate from the actual field of play are of primary interest to every fan. Therefore, these contents are decisive for whether emotions are triggered in the fans or not. In contrast, content such as figures and statistics or fan cultures is more likely to be consumed as additional information to provide background information and satisfy interests needs. Therefore, such content appeals stronger to fans' rationality than to fans' emotions. This stands in line with previous studies (e.g., Vale and Fernandes (2018)), which demonstrated that fans primarily consume sports content on social media due to entertainment motives and that fans are mainly driven by emotions such as passion and social values rather than by rational evaluations (Gruettner, 2019). Along with this finding, as depicted in Figure 11 (left), a conducted t test between subjects who were exposed to photos from the product-related sports content category and subjects who were exposed to photos of the non-product-related sports content category revealed that product-related content achieved significantly higher emotional appeals ($M = 4.53$ vs. 4.06 ; $t(169) = 2.14$, $p = 0.04$). The results for H2 did not achieve significant results. Thus, we have to reject H2. However, the analysis revealed a significant positive effect of our control variable, that is, sports-related social media consumption, on informational appeals ($\beta = .19$, $t(167) = 2.63$, $p = .009$). We interpret this finding as follows: Fans who consume a lot of sports content on social media are interested in sports content that goes beyond product-related content that originates from the actual field of play. They like to consume content such as statistics

and figures. Therefore, sports organizations should consider delivering different content to subjects with a high sports-related social media consumption.

Figure 11

Subjects' Perceived Emotional and Informational Appeals to Product and Non-Product-Related Sports Content in Both Media Types



Evaluating the Effect of Emotional and Informational Appeals on Fan Engagement (H3 – H4)

In H3, we hypothesized that visual sports content with high emotional appeals has a positive effect on fan engagement. The OLS regression confirmed that emotional appeals have a significant effect on fan engagement ($\beta = .54$, $t(167) = 2.78$, $p = .006$). In H4, we hypothesized that visual sports content with high informational appeals has a positive effect on fan engagement. The results did not confirm this hypothesis although both emotional as well as informational appeals have been proven to have an effect on engagement behavior in different consumer offline and online applications. Therefore, we have to reject H4. Extrapolating from these findings, we believe that our sports context differs from these previously conducted studies for two reasons: (1) sports organizations usually do not post content that should affect consumer purchase decisions. Instead, the content posted by sports organizations on social media aims to raise emotions, general interest, or to educate and inform fans. Thus, sports organizations face the challenge of having to understand fans' reactions on social media posts with no specific call to action. In this context, Schwarz (2000) argued that different social media strategies should be considered when posts do not include specific calls to actions. (2) On top of that, we are dealing with an international sports context, as our

data sample only consists of subjects who are American and Indian only. As stated in Section 3.2, we consider both countries as international emerging markets with respect to their interest in football. We believe that fans who are new to sports content on social media will be guided more by emotional appeals than by informational appeals, as they first need to get used to the content. In a nutshell, we can confirm that sports content that triggers high emotional appeals is of relevance for sports organizations, as this content can lead to social media fan engagement. Accordingly, sports organizations should understand how to include more/different or less emotional and informational cues in the visual content of social media posts.

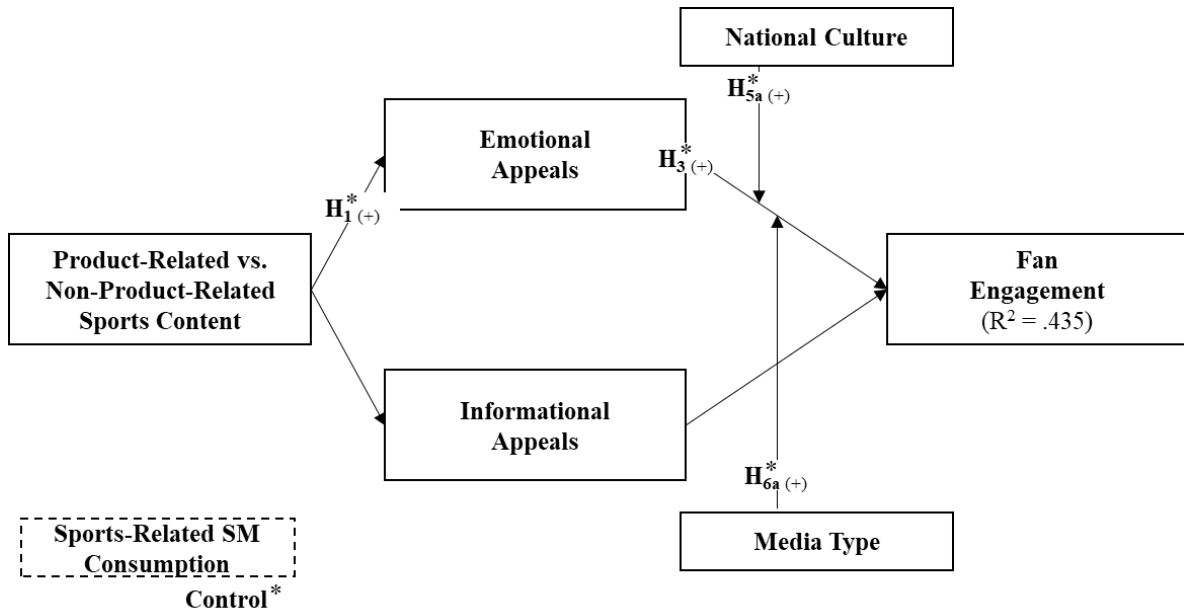
Evaluating the Moderating Effects of National Culture and Media Type on the Relationship between Sports Content, Visual Appeals, and Fan Engagement (H5 – H6)

To test our assumptions of the moderating cross-cultural and media type effects on the relationship between sports content, visual appeals, and fan engagement (H5 – H6; see, Table 3), we have only included subjects in the study that are either American or Indian. We found a statistically significant difference in the indirect conditional effect of sports content categories on fan engagement. In detail, we found evidence that the mediational effect between emotional appeals and fan engagement is significantly moderated by the national culture and by the media type (see, Figure 12). The effect is in favor of the product-related sports content category. The indirect conditional effect between the sports content categories and fan engagement is lowest when the moderators USA and video are selected ($b=.275$, $SE=.163$, $LLCI=.042$, $ULCI=.715$). The indirect conditional effect between the sports content categories and fan engagement is highest when the moderators India and photo are selected ($b=.388$, $SE=.175$, $LLCI=.13$, $ULCI=.824$). The difference between these two effects is mainly caused by national culture. In other words, product-related sports content has an indirect conditional effect on fan engagement when moderated by India rather than America. We interpret these findings as follows: Traditionally, India is a larger football nation (or at least closer to football) than the United States in which, for example, American football and basketball are more popular among fans. For the mediator informational appeals, we could not find a statistically significant effect in terms of moderation through culture or media type. In sum, our results indicated that cultural values also exist among fans. Therefore, they are in line with previously conducted research such as Jackson and Wang (2013). Accordingly, social media sports practitioners need to consider cross-cultural and media type effects in their social media strategies.

Table 3*Conditional Indirect and Moderating Effect(s) for H5 and H6*

Visual Appeals	National Culture	Media Type	Effect	SE	BootLLCI	BootULCI
Emotional	USA	Video	0.2748	0.1629	0.0418	0.7145
Emotional	USA	Photo	0.2884	0.1621	0.0364	0.6646
Emotional	India	Video	0.3748	0.1666	0.1086	0.7635
Emotional	India	Photo	0.3883	0.1745	0.1303	0.8239
Informational	USA	Video	-0.0994	0.0934	-0.3907	0.0083
Informational	USA	Photo	-0.0986	0.0914	-0.4143	0.0001
Informational	India	Video	0.0387	0.0643	-0.0522	0.2188
Informational	India	Photo	0.0395	0.0652	-0.0426	0.2325

SE = Standard error; LLCI = Lower level of confidence interval; ULCI = Upper level of confidence interval

Figure 12*Research Model Including Statistically Significant Results*

Implications, Future Research, and Limitations

To identify (1) the sports content drivers to enhance fan engagement on social media and (2) to investigate how cross-cultural effects moderate the relationship between sports content and fan engagement, we conducted an experiment using a between-subjects design with four treatments in a joint research project with DFL DS. The overarching theoretical perspective on which we ground our research model to test our hypotheses (H1 – H6) builds upon CET as described by Pansari and Kumar (2017). We furthermore build on related research streams on sports organizations' brand attributes (Parganas et al., 2015) and on the concept of visual emotional and informational appeals (Yoo & MacInnis, 2005). In total, 167 subjects were analyzed in our research model that

were collected through the official German Bundesliga channels (i.e., “www.Bundesliga.com”, "Bundesliga Official" on Facebook, and "bundesliga_en" on Instagram). The results showed that product-related sports content had a significant effect on emotional visual appeals (H1). We could not find evidence for our hypothesis that non-product-related sports content has a positive effect on informational appeals. Thus, we have to reject hypothesis H2. Emotional appeals had a significant effect on social media fan engagement (H3), while informational appeals did not show a significant effect (H4). To test our assumption of the moderating cross-cultural effects on the relationship between sports content, visual appeals, and fan engagement, we have only included subjects in the study that are either American or Indian. The results confirmed that the relationship between sports content, visual emotional appeals, and fan engagement is moderated by cross-cultural effects (H5a). However, we did not find significant results for hypothesis H5b. Lastly, we hypothesized that the relationship between sports content, visual appeals, and fan engagement is also moderated by the media type in which social media sports content is embedded. The results confirmed hypothesis H6a. However, we did not find evidence for our hypothesis H6b.

From an academic perspective, the contribution of this study is twofold: Firstly, this study contributes to the upcoming literature stream of sports digitalization in IS literature (see, Gruettner (2019) or Xiao et al. (2017)). In this vein, we answered the calls of several researchers (e.g., Achen et al. (2018), Romney and Johnson (2020), or Vale and Fernandes (2018)) to investigate the sports content drivers that trigger social media fan engagement in an international context and proposed a unique research model to academia. In detail, our research model links sports organizations' brand attributes – that is, product-related and non-product-related attributes – with social media sports content categories and the concept of visual emotional and informational appeals to explain how specific sports content categories constitute fan engagement on social media. On top of that, our research model sheds light on how the relationship between sports content, visual appeals, and fan engagement is moderated by cross-cultural and media type effects. Therefore, we conducted the first study, which investigated content drivers, visual appeals, and cross-cultural and media type effects in social media international fan engagement. In future research, our research model can be extended and adjusted to different sports content categories and national cultures. Secondly, we believe that the sports industry, due to its unique characteristics (see, Gruettner (2019) or Xiao et al. (2017)), provides us with insightful opportunities to study digitalization-related phenomena that cannot be easily observed in generic business contexts. This stands in line with Xiao et al. (2017), who proposed a research agenda for the IS

community that suggests why and how we should study sports digitalization in the IS discipline. Hence, we believe that our findings go beyond the sports literature and provide important insights into social media consumer engagement in general. In this sense, the developed research model can be adopted and adjusted to different online engagement contexts, for instance, to test the moderating effect of cross-cultural effects in different industries. For social media practitioners, our study offers important managerial implications: First, our findings provide guidelines on how to create influential posts and, more importantly, how to adjust emotional and informational appeals to maximize content effectiveness. In designing posts on social media, practitioners can, therefore, choose to include more/different or less emotional and informational cues in the visual content of posts. Our results go beyond the application of sports and give practitioners many clues about what type of social media content increases the liking and sharing of a post in general. Second, we proved that there are cross-cultural as well as media type effects between social media content, visual appeals, and online engagement. Hence, the results of this study should be seen as a wake-up call for practitioners to consider cross-cultural and media type effects in their social media strategies.

This study is not free from limitations: First, the photos and videos presented in our experiment have previously been published on the social media channels of the German Bundesliga. Therefore, on the one hand, we can guarantee a realistic real-world setting. On the other hand, this can lead in some cases to side effects (e.g., the subjects have seen the photos/video before). Second, we only investigated visual content, that is, photos and videos, in this study. Social media posts usually also consist of textual content that is often included in the form of photo or video captions. Informational and emotional appeals can be embedded in both textual as well as visual content. Thus, studying textual content and their interaction with visual content can become relevant. Third, we used the subjects' nationalities as a proxy for their national culture. Although this approach is common in research (e.g., Hudson et al. (2016)), one can argue that culture has to be measured along previously defined valid dimensions/scales. Fourth, we only included subjects whose nationality is either American or Indian in our analyses. An extension of the scope of this study to include a variety of nationalities may become relevant. Fifth, because we worked with an online population and used observed behavior, it is likely that our sample suffers from self-selection bias. Sixth, we believe that the lack of further control variables could be a limitation. For instance, the influence created by friends in SNSs. Lastly, in times of COVID-19, nobody can exactly

say how the subjects in our experiment were influenced by it. Particularly because at the time of our data collection, almost all sports events worldwide were canceled.

Conclusion

By building upon CET, we linked sports organizations' brand attributes, the concept of emotional and informational visual appeals, and fan engagement. Therefore, we proposed a unique research model to academia and analyzed (1) social media sports content drivers for fan engagement and (2) cross-cultural effects that moderate the relationship between sports content categories, visual appeals, and fan engagement. The results of the conducted experiment with four treatments demonstrated that product-related sports content – that is, content that originates from the actual game on-site the field of play (e.g., player skills or match highlights) – had a significant effect on emotional visual appeals. It is crucial that sports organizations understand how to adjust visual appeals to maximize content effectiveness, as our analyses yielded significant evidence that emotional visual appeals have a positive effect on fan engagement. Furthermore, the study showed significant results that the relationship between sports content, emotional visual appeals, and fan engagement is moderated by cross-cultural and media types effects. The results of this study, therefore, provide guidance for social media practitioners on how to create influential posts and offer important insights into consumer online engagement in an international context and in general. Future research should adopt and adjust our research model to different contexts.

References

- 11Freunde. (2020). *Der Zirkus ist in der Stadt*. <https://apps.11freunde.de/artikel/der-zirkus-ist-in-der-stadt/534881>
- Achen, R. M., Kaczorowski, J., Horsmann, T., & Ketzler, A. (2018). Exploring Off-Season Content and Interaction on Facebook: A Comparison of U.S. Professional Sport Leagues. *International Journal of Sport Communication*, 11(3), 389–413. <https://doi.org/10.1123/ijsc.2018-0013>
- Aichner, T. (2019). Football clubs' social media use and user engagement. *Marketing Intelligence and Planning*, 37(3), 242–257. <https://doi.org/10.1108/MIP-05-2018-0155>
- Akpınar, E., & Berger, J. (2017). Valuable Virality. *Journal of Marketing Research*, 54(2), 318–330. <https://doi.org/10.1509/jmr.13.0350>

- Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The Role of Emotions in Marketing. *Journal of the Academy of Marketing Science*, 27(2), 184–206. <https://doi.org/10.1177/0092070399272005>
- Billings, A. C., Broussard, R. M., Xu, Q., & Xu, M. (2018). Untangling International Sport Social Media Use: Contrasting U.S. and Chinese Uses and Gratifications Across Four Platforms. *Communication & Sport*, 7(5), 1–23. <https://doi.org/10.1177/2167479518790014>
- Bouzdine-Chameeva, T., Ferrand, A., Valette-Florence, P., & Chanavat, N. (2015). Measurement and segmentation of sport fans using brand association networks: Application to Union of European Football Associations (UEFA) Champions League (UCL). *Sport Management Review*, 18(3), 407–420. <https://doi.org/10.1016/j.smr.2014.11.002>
- Deutscher Fußball-Bund. (2018). *Bundesliga Zuschauerzahlen*. <https://www.dfb.de/bundesliga/statistik/zuschauerzahlen/>
- Fillis, I., & Mackay, C. (2014). Moving beyond fan typologies: The impact of social integration on team loyalty in football. *Journal of Marketing Management*, 30(3–4), 334–363. <https://doi.org/10.1080/0267257X.2013.813575>
- Filo, K., Lock, D., & Karg, A. (2015). Sport and social media research: A review. *Sport Management Review*, 18(2), 166–181. <https://doi.org/10.1016/j.smr.2014.11.001>
- Goldberg, M. E., & Gorn, G. J. (1987). Happy and Sad TV Programs: How They Affect Reactions to Commercials. *Journal of Consumer Research*, 14(3), 387–403. <https://doi.org/10.1086/209122>
- Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. *Proceedings of the 25th Americas Conference on Information Systems (AMCIS) 2019*.
- Gruettner, A., & Haferbeck, L. O. H. (2020). *Data Set - Going Global: Enhancing International Social Media Fan Engagement – Evidence from the German Bundesliga*. <https://www.alexandria.unisg.ch/260252/>
- Hayes, A. F. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling*. <http://www.afhayes.com/>
- Hofstede, G. (1983). The Cultural Relativity of Organizational Practices and Theories. *Journal of International Business Studies*, 14(2), 75–89. <https://doi.org/10.1057/palgrave.jibs.8490867>
- Hu, C., & Cole, S. (2016). Comparing Expenditure Patterns and Travel Characteristics among NFL fans with Different Levels of Team Identification. *Tourism Review International*, 20(1), 3–10. <https://doi.org/10.3727/154427216X14581596798906>

- Hudson, S., Huang, L., Roth, M. S., & Madden, T. J. (2016). The influence of social media interactions on consumer-brand relationships: A three-country study of brand perceptions and marketing behaviors. *International Journal of Research in Marketing*, 33(1), 27–41. <https://doi.org/10.1016/j.ijresmar.2015.06.004>
- Jackson, L. A., & Wang, J. L. (2013). Cultural differences in social networking site use: A comparative study of China and the United States. *Computers in Human Behavior*, 29(3), 910–921. <https://doi.org/10.1016/j.chb.2012.11.024>
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in Social Media Research: Past, Present and Future. *Information Systems Frontiers*, 20(3), 531–558. <https://doi.org/10.1007/s10796-017-9810-y>
- Keller, K. L. (1993). Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *Journal of Marketing*, 57(1), 1–22. <https://doi.org/10.1177/002224299305700101>
- Lee, D., Hosanagar, K., & Nair, H. S. (2018). Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook. *Management Science*, 64(11), 5105–5131. <https://doi.org/10.1287/mnsc.2017.2902>
- Li, Y., & Xie, Y. (2020). Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*, 57(1), 1–19. <https://doi.org/10.1177/0022243719881113>
- Liu, X., Zhang, B., Susarla, A., & Padman, R. (2020). Go to YouTube and See Me Tomorrow: Social Media and Self-Care of Chronic Conditions. *MIS Quarterly*, 44(1b). <https://doi.org/10.2139/ssrn.3061149>
- MacInnis, D. J., Rao, A. G., & Weiss, A. M. (2002). Assessing When Increased Media Weight of Real-World Advertisements Helps Sales. *Journal of Marketing Research*, 39(4), 391–407. <https://doi.org/10.1509/jmkr.39.4.391.19118>
- Muntinga, D. G., Moorman, M., & Smit, E. G. (2011). Introducing COBRAs: Exploring motivations for brand-related social media use. *International Journal of Advertising*, 30(1), 13–46. <https://doi.org/10.2501/IJA-30-1-013-046>
- Ordenes, F. V., Grewal, D., Ludwig, S., Ruyter, K. De, Mahr, D., & Wetzels, M. (2019). Cutting through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages. *Journal of Consumer Research*, 45(5), 988–1012. <https://doi.org/10.1093/jcr/ucy032>
- Pansari, A., & Kumar, V. (2017). Customer engagement: the construct, antecedents, and consequences. *Journal of the Academy of Marketing Science*, 45(3), 294–311. <https://doi.org/10.1007/s11747-016-0485-6>

- Parganas, P., Anagnostopoulos, C., & Chadwick, S. (2015). “You’ll never tweet alone”: Managing sports brands through social media. *Journal of Brand Management*, 22(7), 551–568. <https://doi.org/10.1057/bm.2015.32>
- Rietveld, R., van Dolen, W., Mazloom, M., & Worrying, M. (2020). What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram. *Journal of Interactive Marketing*, 49, 20–53. <https://doi.org/10.1016/j.intmar.2019.06.003>
- Romney, M., & Johnson, R. G. (2020). Show me a story: narrative, image, and audience engagement on sports network Instagram accounts. *Information, Communication & Society*, 23(1), 94–109. <https://doi.org/10.1080/1369118X.2018.1486868>
- Schivinski, B., Christodoulides, G., & Dabrowski, D. (2016). Measuring Consumers’ Engagement With Brand-Related Social-Media Content: Development and Validation of a Scale That Identifies Levels of Social-Media Engagement with Brands. *Journal of Advertising Research*, 56(1), 64–80. <https://doi.org/10.2501/JAR-2016-004>
- Schwarz, N. (2000). Emotion, cognition, and decision making. In *Cognition and Emotion* (Vol. 14, Issue 4, pp. 433–440). Psychology Press Ltd. <https://doi.org/10.1080/026999300402745>
- Vale, L., & Fernandes, T. (2018). Social media and sports: driving fan engagement with football clubs on Facebook. *Journal of Strategic Marketing*, 26(1), 37–55. <https://doi.org/10.1080/0965254X.2017.1359655>
- Wang, Q., Miao, F., Tayi, G. K., & Xie, E. (2019). What makes online content viral? The contingent effects of hub users versus non-hub users on social media platforms. *Journal of the Academy of Marketing Science*, 47(6), 1005–1026. <https://doi.org/10.1007/s11747-019-00678-2>
- Xiao, X., Hedman, J., Ter, F., Tan, C., Tan, C.-W., Lim, E. T. K., Clemenson, T., Henningsson, S., Vatrappu, R., Rao Mukkamala, R., & Van Hilleberg, J. (2017). Sports Digitalization: An Overview and A Research Agenda. *Proceedings of the 38th International Conference on Information Systems (ICIS) 2017*.
- Yoo, C., & MacInnis, D. (2005). The brand attitude formation process of emotional and informational ads. *Journal of Business Research*, 58(10), 1397–1406. <https://doi.org/10.1016/j.jbusres.2005.03.011>

Publication V

[German]

**Konfrontiert mit einer Flut von Kooperationsanfragen:
So meistern Sportorganisationen die systematische Analyse und Bearbeitung**



Author(s)

Arne Grüttner

Thomas Braschler

Andrea Back

Published in Wirtschaftsinformatik & Management – Issue on: "IT und Sport", 2020.

Kurzfassung

Die zunehmende Digitalisierung und tiefgreifende Kommerzialisierung der Sportindustrie hat dazu geführt, dass viele Sportorganisationen zu nationalen und globalen Marken mit enormer Reichweite und Popularität geworden sind. Als Folge erhalten Sportorganisationen eine Vielzahl von Kooperationsanfragen (z.B. von Start-ups). Die sich daraus ergebenden Möglichkeiten zur Verbesserung und Neuerfindung von Produkten, Dienstleistungen und Betriebsabläufen bleiben jedoch oft ungenutzt, da es an Ressourcen und an etablierten Prozessen bei der Analyse und Bearbeitung dieser Kooperationsanfragen mangelt. Wir haben auf Basis der wissenschaftlichen Methode Design Science Research ein IT-Tool konzipiert¹, das sowohl eine initiale automatisierte Bewertung von Kooperationsanfragen übernimmt als auch nachgelagerte Arbeitsschritte steuert und optimiert.

Einführung

Viele Sportorganisationen erhalten aufgrund ihrer enormen Reichweite und Popularität und der damit verbundenen wirtschaftlichen Attraktivität, eine Vielzahl von Kooperationsanfragen von externen Organisationen. In Gesprächen mit fünf Vertretern aus der Deutschen Fußball Bundesliga und mit zwei Vertretern aus der Schweizerischen Fußball Super League wurde die Anzahl auf etwa 20 bis 60 Kooperationsanfragen im Monat beziffert. Beispielsweise bieten Medienagenturen ihre Dienste an, um die Social-Media-Aktivitäten einer Sportorganisation zu optimieren oder Start-ups ihre innovativen Technologien zur Leistungsverfolgung von Athleten. Solche Kooperationsanfragen variieren in ihrer Form, in ihrem Umfang und in ihren möglichen Auswirkungen auf die Sportorganisation. Kooperationsanfragen stellen Sportorganisationen heutzutage vor große Herausforderungen, da es an internen Ressourcen (z.B. personeller Natur und Know-how) und an etablierten Prozessen bei der Analyse und Bearbeitung dieser Kooperationsanfragen mangelt. Als Konsequenz bleiben Möglichkeiten zur Verbesserung und Neuerfindung von Produkten, Dienstleistungen und Betriebsabläufen häufig ungenutzt.

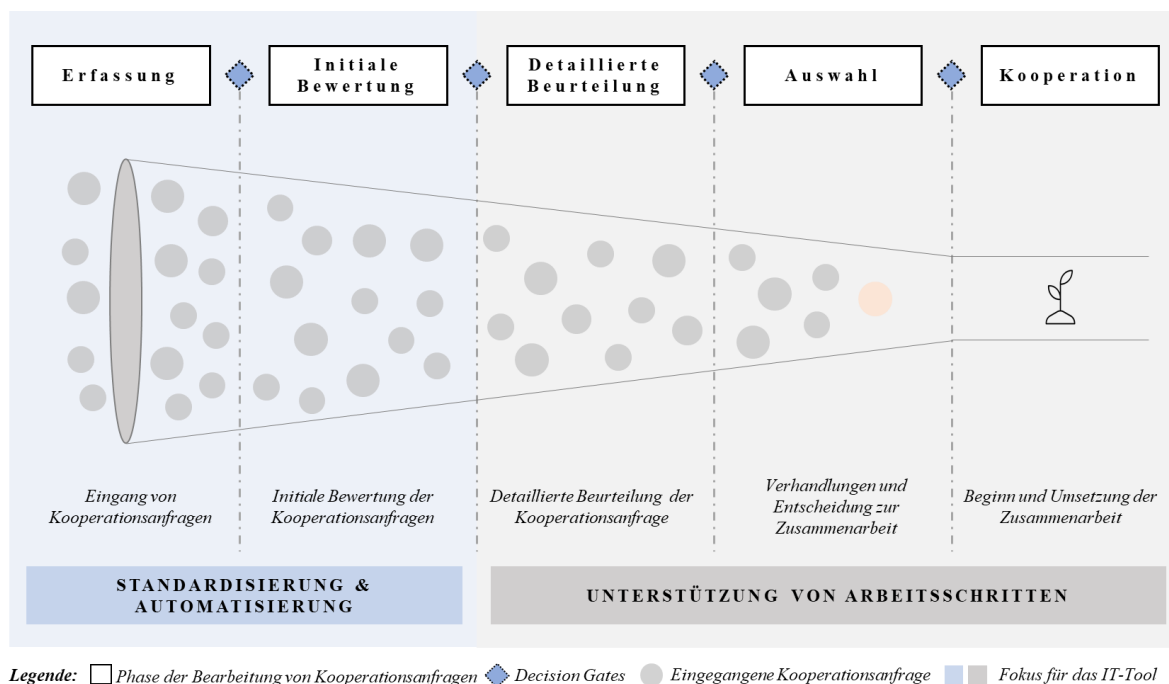
Die zunehmende Kommerzialisierung und Professionalisierung der Sportindustrie hat allerdings gezeigt, dass es gerade für Sportorganisationen immer wichtiger wird diese Möglichkeiten zu nutzen, um nachhaltig wettbewerbsfähig zu bleiben (Gruettner, 2019). Der Fußballklub FC Barcelona zum Beispiel hat dies erkannt und daher im März 2017

¹ Es ist geplant, dass bis Q2/2021 das IT-Tool in Form einer ersten Webanwendung vorliegt.

das Barça Innovation Hub ins Leben gerufen – eine Plattform, die als Ökosystem zur Förderung von Wissen und Innovation zwischen dem Klub und führenden Marken, Universitäten, Forschungszentren und Start-ups fungiert. Die systematische Analyse und Bearbeitung von eingehenden Kooperationsanfragen bieten die Möglichkeit, nachhaltig Wissen anzusammeln, Innovationen zu erkennen und damit verbunden neue Geschäftsmöglichkeiten in Sportorganisationen zu fördern. Um diese systematische Vorgehensweise zu ermöglichen, haben wir in unserem Forschungsprojekt auf Basis der wissenschaftlichen Methode Design Science Research (kurz DSR) ein IT-Tool konzipiert. Ziel ist es, dass das IT-Tool als zentralisiertes und standardisiertes Eingangstor Kooperationsanfragen erfasst, diese initial automatisiert bewertet und im Anschluss unterstützend die nachgelagerten Arbeitsschritte zwischen Mitarbeitern und Abteilungen innerhalb von Sportorganisationen und den anfragenden externen Organisationen steuert und optimiert (siehe Abbildung 1). Der DSR-Forschungsansatz, die zugehörigen Arbeitsschritte sowie das IT-Tool werden auf den nachfolgenden Seiten detailliert präsentiert.

Abbildung 1

Lebenslauf eingehender Kooperationsanfragen und Fokus unseres IT-Tools



Design Science Research – Wissenschaftlich fundierte Artefakte für die Praxis

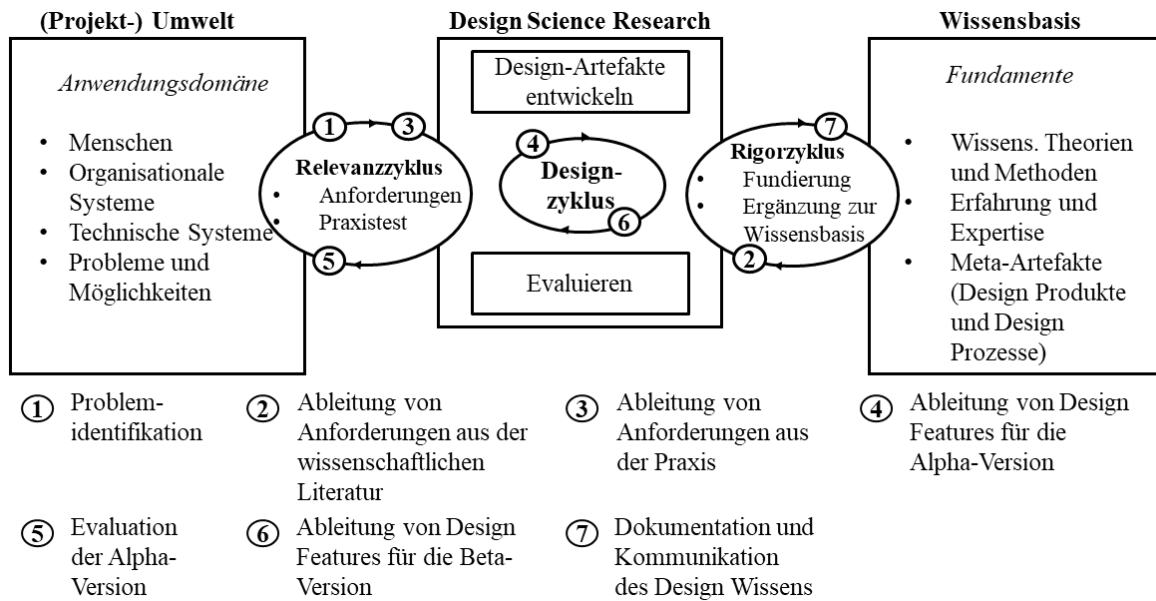
DSR ist ein Forschungsansatz, der aus der Disziplin der Informationswissenschaften stammt. Im Kern zielt DSR darauf ab, praktische Probleme mit innovativen Lösungen zu lösen. Dabei steht die Nützlichkeit des Forschungsergebnisses im Vordergrund, während beispielsweise in anderen Forschungsansätzen wie den Verhaltenswissenschaften das Ziel die "Wahrheit" ist (Hevner et al., 2004). Im Detail hat DSR die Absicht, die (Projekt-) Umwelt durch die Einführung neuer und innovativer so genannter Artefakte zu verbessern. Bei diesen Artefakten handelt es sich in der Regel selten um vollwertige, ausgewachsene Informationssysteme. So können Artefakte auch Konstrukte, Modelle, Methoden und Instanziierungen umfassen. Um diese Artefakte zu "designen", zeichnet sich DSR durch das Prinzip aus, sowohl praktische Relevanz als auch wissenschaftliche Strenge (engl. rigor) zu verbinden. Hevner (2007) bietet hierfür ein DSR-Prozessmodell an, das aus drei eng miteinander verbundene Aktivitätszyklen besteht (siehe Abbildung 2) und wie folgt erklärt ist:

- (1) Der Relevanzzyklus beschreibt die Verbindung der (Projekt-) Umwelt mit der Entwicklung des Artefakts. Er liefert den Input bezüglich des Forschungsbedarfs (d.h. der Problemstellung) sowie die Akzeptanzkriterien für die abschließende Beurteilung der Forschungsergebnisse.
- (2) Der Rigorzyklus überbrückt die Lücke zwischen der Entwicklung des Artefakts und der bestehenden Wissensbasis hinsichtlich bestehender wissenschaftlicher Theorien, Methoden und Erfahrungen.
- (3) Der Designzyklus bildet einen iterativen Prozess zwischen der Entwicklung und dem Evaluieren des Artefakts. In diesem iterativen Prozess werden die Evaluation und ihr anschließendes Feedback zur weiteren Verfeinerung des Artefakts verwendet.

Das gesamte DSR-Prozessmodell nach Hevner (2007) umfasst sieben spezifische Schritte über die drei zuvor vorgestellten Aktivitätszyklen hinweg. Ziel unseres Forschungsprojektes ist es, ein Artefakt in Form eines IT-Tools zu entwickeln. Um die spezifischen Richtlinien für die Durchführung eines DSR-Projekts zu gewährleisten, haben wir uns an den sieben Schritten orientiert, die zusammen mit den Ergebnissen auf den nächsten Seiten des Artikels präsentiert werden. Aufgrund der Kürze und der praktischen Orientierung dieses Beitrags, verzichten wir auf die Ableitung von Design Features für die Beta-Version (Schritt 6) und die Dokumentation und die Kommunikation des Design Wissens für die Wissensbasis (Schritt 7).

Abbildung 2

DSR-Prozessmodell und zugehörige sieben Schritte in Anlehnung an Hevner (2007)

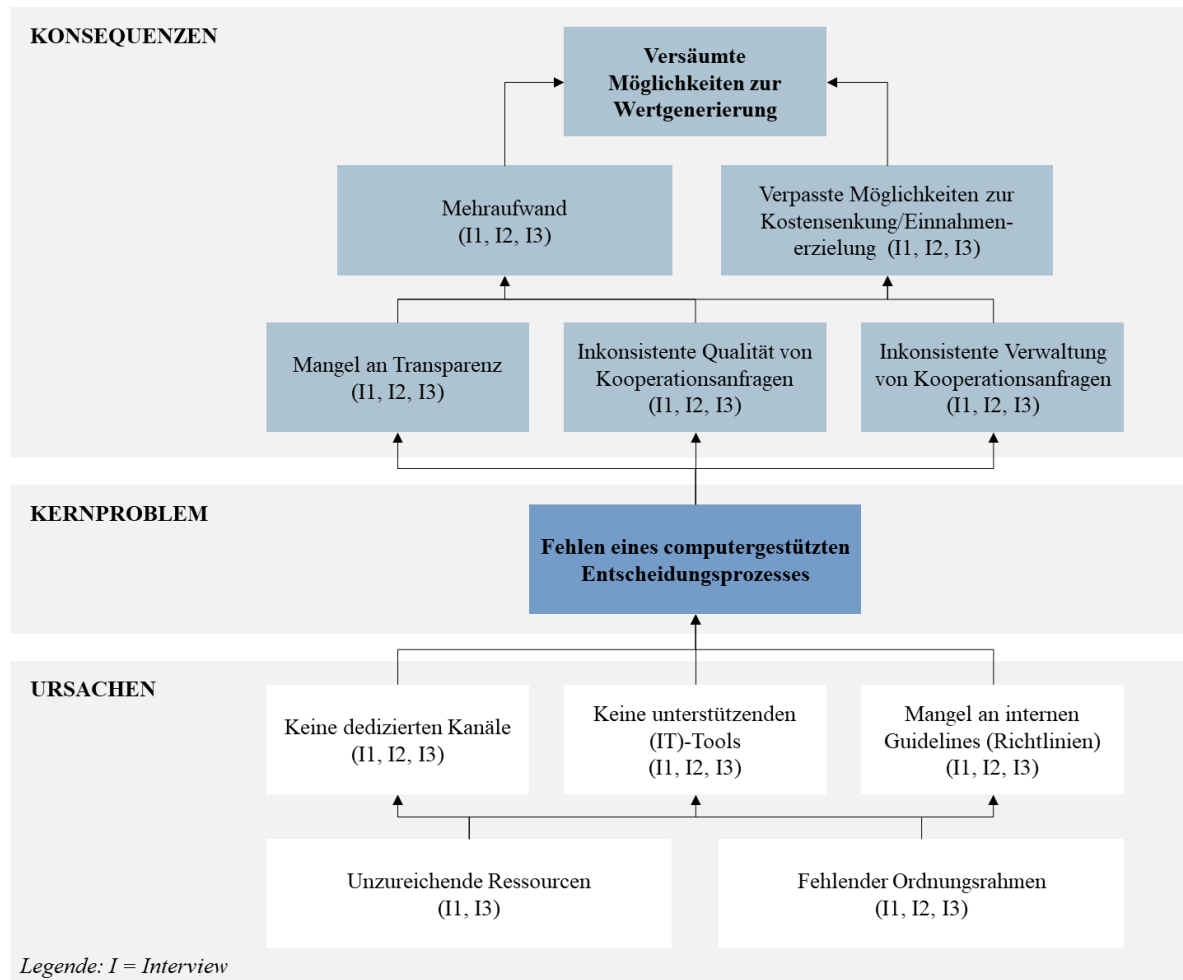


Schritt 1 – Problemidentifikation

Der erste Schritt des DSR-Prozessmodells besteht darin, ein tieferes Verständnis über das zu lösende Problem zu erlangen. Dafür wurden drei initiale Experteninterviews mit Vertretern von Fußballvereinen aus der Deutschen Fußball Bundesliga (zwei Interviews) und der Schweizerischen Fußball Super League (ein Interview) geführt. Ziel der Interviews war es, (1) das aktuelle Vorgehen bei der Analyse und Bearbeitung von Kooperationsanfragen zu verstehen und (2) ein besseres Verständnis dafür zu erlangen, wie ein IT-Tool zur Unterstützung eingesetzt werden könnte. Die aus den Experteninterviews gewonnenen Erkenntnisse lassen sich in einer so genannten Problembaumanalyse zusammenfassen, wie sie in Abbildung 3 dargestellt wird.

Abbildung 3

Problembaumanalyse: Ursachen, Kernproblem und Konsequenzen



Ursachen: Zwei Gründe wurden als Hauptursachen für die mangelnde Analyse und Bearbeitung von Kooperationsanfragen in Sportorganisationen identifiziert:

- (1) **Unzureichende Ressourcen:** Zwei der drei Experten gaben an, dass es ihrer Sportorganisation oft an Ressourcen fehlt, um die Masse der eingehenden Kooperationsanfragen richtig zu analysieren und zu bearbeiten. Die Experten schätzen, dass ihre Sportorganisationen jeweils zwischen 20 und 60 Kooperationsanfragen im Monat erhalten. Eine gründliche Analyse und Bearbeitung dieser Kooperationsanfragen erfordern Ressourcen (z.B. personeller Natur und Know-how), die oft nicht ausreichend zur Verfügung stehen.
- (2) **Fehlender Ordnungsrahmen:** Alle drei Experten gaben an, dass es keinen etablierten Ordnungsrahmen für die Analyse und Bearbeitung von Kooperationsanfragen in ihrer Sportorganisation gibt. Daher werden die Mitarbeiter, die Kooperationsanfragen erhalten, in ihrer Entscheidung, ob und wie sie eine Kooperationsanfrage bearbeiten, allein gelassen.

Kernproblem: Die zwei identifizierten Hauptursachen drücken sich in einem Kernproblem aus: das Fehlen eines computergestützten Entscheidungsprozesses. Auf die Frage, wie die gesamte Analyse und Bearbeitung von Kooperationsanfragen in seiner Sportorganisation beschrieben werden könnte, antwortete ein Experte mit dem einfachen, aber aufschlussreichen Satz, dass “Kooperationsanfragen nach dem Zufallsprinzip, unstrukturiert und ohne jedes Konzept behandelt werden” (P. Grabolle, personal communication, February 6, 2020). Ein anderer Experte beschrieb den internen Prozess zur Analyse und Bearbeitung von Kooperationsanfragen allgemein als "Kür" und erklärte dies folgenderweise: “Jeder, der eine Anfrage erhält, kann frei entscheiden, was er damit macht. Es gibt weder eine Regel, dass man jede Anfrage beantworten muss, noch gibt es eine Standard-Mail oder -Antwort” (A. Cüppers, personal communication, January 30, 2020).

Konsequenzen: Aus dem beschriebenen Kernproblem ergeben sich verschiedene direkte und indirekte Konsequenzen für Sportorganisationen, die in ihrer Gesamtheit darin münden, dass Möglichkeiten zur Verbesserung und Neuerfindung von Produkten, Dienstleistungen und Betriebsabläufen (d.h. eine Wertgenerierung) ungenutzt bleiben.

Schritt 2 – Ableitung von Anforderungen aus der wissenschaftlichen Literatur

In Schritt 2 des DSR-Prozessmodells werden die Meta-Anforderungen für das zu entwickelnde Artefakt aus der Wissensbasis abgeleitet. Hierfür haben wir eine systematische Literaturrecherche nach vom Brocke et al. (2009) durchgeführt. Insgesamt haben wir mittels der Keyword-Suche nach „Strategic Alliance“, „Partner Assessment/Evaluation“ und „Partner Selection“ in drei Datenbanken 20 für den Kontext dieses Forschungsprojektes relevante Studien identifiziert. Im Kern führte die Analyse dieser Studien zu drei spezifischen Erkenntnissen, die die nachfolgenden Meta-Anforderungen für unser IT-Tool ableiten:

Ein IT-Tool zur Analyse und Bearbeitung von Kooperationsanfragen für Sportorganisationen muss ...

- einen strategischen Fit eines potenziellen Partners berücksichtigen: Die Auswahl und die Bewertung der Kooperationsanfragen müssen unter Berücksichtigung ihrer strategischen Eignung für die Sportorganisation erfolgen. Dabei soll eine mögliche Kooperation an übergeordnete Unternehmensziele gebunden sein (Holmberg & Cummings, 2009).

- die Bedeutung des Geschäftskontextes berücksichtigen: Bei der Partnerwahl haben zahlreiche konzeptionelle und empirische Untersuchungen gezeigt, dass es eine nahezu unbegrenzte Anzahl potenzieller Kriterien für die Partnerauswahl gibt. Daher raten Forscher, dass die Kriterien für die Partnerauswahl stark von den Merkmalen und dem Kontext der an der Beziehung beteiligten Partner abgeleitet werden sollten (Nielsen & Gudergan, 2012).
- Aufgaben- und partnerbezogene kritische Erfolgsfaktoren berücksichtigen: Einer der führenden Forschungsbeiträge auf dem Gebiet der Partnerwahl in Allianzen ist die konzeptionelle Arbeit von Geringer (1991). In dieser Studie werden die beiden Begriffe aufgaben- und partnerbezogene kritische Erfolgsfaktoren eingeführt, die es erlauben, spezifische Partnerauswahlkriterien in zwei Kategorien zu gruppieren. Aufgabenbezogene Erfolgsfaktoren können als operative Fähigkeiten, Ressourcen und Fertigkeiten beschrieben werden, die ein Unternehmen für den zukünftigen Erfolg benötigt, während partnerbezogene Erfolgsfaktoren sich auf die Effektivität und Effizienz des Partners selbst beziehen.

Schritt 3 – Ableitung von Anforderungen aus der Praxis

Schritt 3 des DSR-Prozessmodells konzentriert sich auf die (Projekt-) Umwelt. Ziel ist es, weitere Meta-Anforderungen für das zu entwickelnde Artefakt zu identifizieren, die sich aus den Erkenntnissen der ersten drei Experteninterviews ableiten. Die relevantesten fünf praktischen Erkenntnisse, die in den Interviews identifiziert wurden, werden wie folgt zusammengefasst:

Ein IT-Tool zur Analyse und Bearbeitung von Kooperationsanfragen für Sportorganisationen muss ...

- Kooperationsanfragen zentral speichern und zugänglich machen.
- einen Überblick über die eingegangenen, in Bearbeitung befindlichen und bereits bearbeiteten Kooperationsanfragen bieten.
- eine strukturierte Detailübersicht zu jeder einzelnen Kooperationsanfrage ermöglichen.
- eine initiale automatisierte Bewertung von Kooperationsanfragen hinsichtlich ihrer Relevanz für Sportorganisationen ermöglichen.
- Funktionen bereitstellen, die die nachgelagerten Arbeitsschritte zur Analyse und Bearbeitung der Kooperationsanfrage innerhalb der Sportorganisation und mit den anfragenden Organisationen unterstützen.

Schritt 4 – Ableitung von Design Features für die Alpha-Version

Das Ziel in Schritt 4 des DSR-Prozessmodells besteht darin, die aus der Wissensbasis und (Projekt-) Umwelt abgeleiteten Meta-Anforderungen in einen ersten Satz von Designfeatures zu übersetzen, sodass sie als konkrete Richtlinien für die Entwicklung einer Alpha-Version des Artefakts verwendet werden können. Dabei wird die Entwicklung des Artefakts mit der Definition eines übergreifenden Designprinzips eingeleitet. Wir haben das übergreifende Designprinzip in unserem Forschungsprojekt wie folgt definiert:

- Bei dem zu entwickelnden Artefakt handelt es sich um ein webbasiertes IT-Tool, das externen Organisationen die systematische Einreichung von Kooperationsanfragen bei Sportorganisationen ermöglicht sowie Sportorganisationen unterstützt, Kooperationsanfragen effizient zu analysieren und zu bearbeiten.

Wir haben uns auf Basis der abgeleiteten Meta-Anforderungen dazu entschieden, das zu entwickelnde IT-Tool in zwei Hauptkomponenten zu unterteilen: (1) ein Formular für die Einreichung von Kooperationsanfragen und (2) das eigentliche IT-Tool zur Analyse und Bearbeitung von Kooperationsanfragen. Screenshots der Alpha-Version beider Hauptkomponenten sowie eine kurze Erklärung zu einzelnen Designfeatures sind in den nachfolgenden Abbildung 4 bis 8 dargestellt.

Das Formular dient als zentrales Eingangstor für die einzelnen Kooperationsanfragen und gewährleistet ein einheitliches, standardisiertes Format. Es ist von den anfragenden externen Organisationen auszufüllen und enthält eine Vielzahl von Fragen, die auf den Sportorganisationskontext des Forschungsprojekts angepasst sind. Die abgefragten Datenfelder und Datentypen sind so gewählt, dass sie die nachgelagerte Analyse und initiale automatisierte Bewertung von Kooperationsanfragen mittels Data Mining- und Natural Language Processing-Techniken erleichtern. Nach dem Einreichen der Kooperationsanfrage werden die gesammelten Daten in eine Datenbank zur Analyse und weiteren Bearbeitung eingespeist.

Abbildung 4

Formular für die Einreichung von Kooperationsanfragen

The diagram illustrates the four steps of the FC JOGA BONITO cooperation request form, connected by arrows indicating the flow from step 1 to step 2, step 2 to step 3, and step 3 to step 4.

Step 1: COMPANY & CONTACT DETAILS

FC JOGA BONITO

COMPANY & CONTACT DETAILS

MS ☐ MR ☐

LAST NAME

FIRST NAME

ROLE/POSITION

EMAIL

TELEPHONE

FIRM NAME

ADDRESS

PLACE ZIP CODE

COUNTRY

WEBSITE

INDUSTRY

IS YOUR COMPANY A STARTUP/SCALEUP?

YES ☐ NO ☐

FIRM SIZE

Back Next

Step 2: DEPARTMENT OF INTEREST

FC JOGA BONITO

DEPARTMENT OF INTEREST

WHICH OF OUR DEPARTMENT DOES YOUR PRODUCT/SERVICE MAINLY CONCERN?

PROFESSIONAL SECTION ☐

YOUTH SECTION ☐

COMMUNICATIONS / INVESTOR RELATIONS / MEDIA ☐

LOGISTICS & EVENTS ☐

SALES & MARKETING ☐

HUMAN RESOURCES ☐

IT & FINANCE ☐

MERCHANDISING & LICENSING ☐

Back Next

Step 3: BUSINESS MODEL & COST OF PRODUCT/SERVICE

FC JOGA BONITO

BUSINESS MODEL & COST OF PRODUCT/SERVICE

ADD DOCUMENTS

DRAG AND DROP OR [BROWSE](#)

DOCUMENT COMMENTS (OPTIONAL)

Add any comments/information to your document, if necessary...

0/250

DO YOU HAVE ANY CONTACT/REFERENCE IN OUR ORGANIZATION?

YES ☒ NO ☐

YOUR CONTACT(S)/REFERENCE(S)

Please name your contact(s)/reference(s) at FC Joga Bonito...

0/200

Back Complete Form

Step 4: DESCRIPTION OF PRODUCT / SERVICE AND ADDED VALUE

FC JOGA BONITO

DESCRIPTION OF PRODUCT / SERVICE AND ADDED VALUE

PRODUCT/SERVICE DESCRIPTION

Please briefly describe your product or service ...

0/1000

PROBLEM SOLUTION & ADDED VALUE FOR FC JOGA BONITO

Please briefly describe a specific problem of ours that your product/service addresses and how it adds value to FC Joga Bonito ...

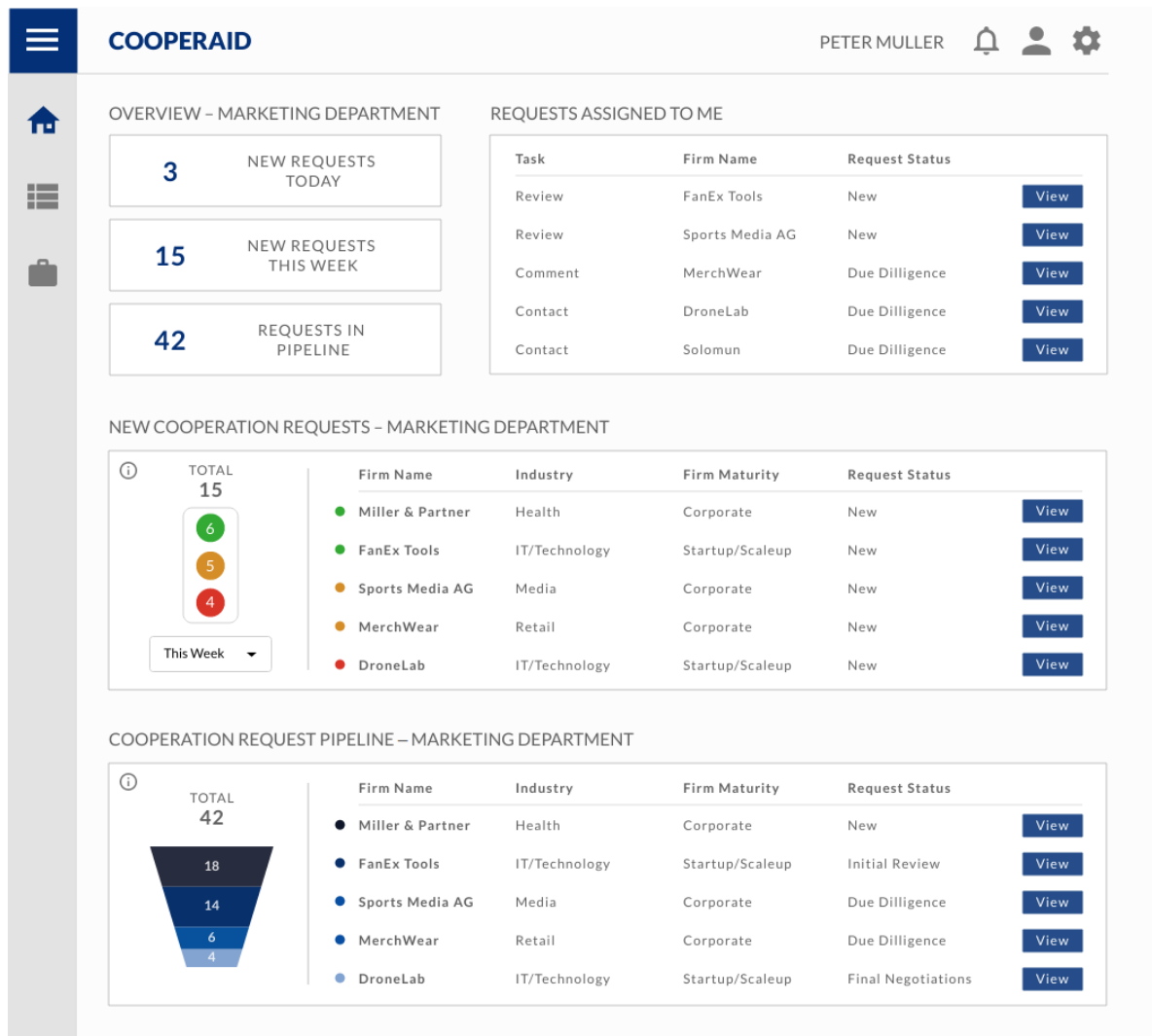
0/1000

Back Next

Das IT-Tool zur Analyse und Bearbeitung der Kooperationsanfragen ist in drei Elemente gegliedert (siehe Abbildung 5 bis 7). Das Home-Dashboard für Kooperationsanfragen bietet einen Überblick über alle eingegangenen, in Bearbeitung befindlichen und bereits bearbeiteten Kooperationsanfragen. Herzstück des Home-Dashboards ist das Ampelsystem, das alle eingegangenen Kooperationsanfragen mit Hilfe von Algorithmen des maschinellen Lernens initial automatisiert bewertet und kategorisiert: grün: hohe Relevanz; gelb: mittlere Relevanz; rot: geringe Relevanz. Im Detail basiert der Bewertungsalgorithmus auf einer Gewichtung von (1) Übereinstimmung mit Organisations- und Abteilungszielen und (2) Qualitätsmerkmalen der Kooperationsanfrage selbst.

Abbildung 5

Home-Dashboard für Kooperationsanfragen



Das zweite Element ist die zentrale Datenbank für Kooperationsanfragen. Hier werden alle aktuellen und historischen Kooperationsanfragen gespeichert und zugänglich gemacht. Das Menü beinhaltet unterschiedliche Filter. Zum Beispiel kann nach der betroffenen Abteilung, der Relevanz der Kooperationsanfrage oder nach der Industrie und der Firmengröße der anfragenden Organisation gefiltert werden. Des Weiteren existiert eine allgemeine Wortfeldsuche. Es besteht die Möglichkeit, Kooperationsanfragen, die in der Vergangenheit abgelehnt wurden, nun aber inhaltlich relevant werden, zu reaktivieren.

Abbildung 6

Zentrale Datenbank für Kooperationsanfragen

COOPERAID PETER MULLER

COOPERATION REQUEST DATABASE

Search... Go!

Department Categorization Request Status From To

Industry Firm Maturity Firm Size

☒ Last 7 days ☐ Last 30 days ☐ Last 90 days Clear 2

Department	Categorization	Request Status	Firm Name	Industry	Firm Maturity	Firm Size
☐ Marketing	Red	Rejected	Altos	IT/Technology	Startup/Scaleup	1-10
☐ Sales	Orange	Rejected	Secure IX	IT/Technology	Corporate	251-500
☐ Finances	Green	Initial Review	Yallo	IT/Technology	Startup/Scaleup	11-50
☐ Digital	Green	Due Dilligence	OneGrid	IT/Technology	Startup/Scaleup	11-50
☐ HR	Orange	Rejected	TechnoHub	IT/Technology	Startup/Scaleup	51-100
☐ Finances	Orange	Initial Review	Santoville	IT/Technology	Corporate	1001-5000
☐ IT	Red	Rejected	Solomun	IT/Technology	Corporate	5001-10000
☐ Marketing	Green	Rejected	EasyHelp	IT/Technology	Startup/Scaleup	11-50
☐ Sales	Orange	Rejected	Analytics Go!	IT/Technology	Corporate	51-100

69 entries found < 1 2 3 4 5 ... 9 10 > View Export

Das dritte Element ist der persönliche Arbeitsbereich für Kooperationsanfragen. Hier werden alle für den jeweiligen Mitarbeiter relevanten Kooperationsanfragen zur Analyse und Bearbeitung gesammelt. Es bietet Zugang zu einer Detailansicht einer Kooperationsanfrage (siehe Abbildung 8) und die Möglichkeit, Kooperationsanfragen zu anderen Mitarbeitern und Abteilungen weiterzuleiten und unterstützt so die nachgelagerten Arbeitsschritte für die Analyse und Bearbeitung. Schließlich bietet es die Möglichkeit, irrelevante Kooperationsanfragen automatisch mit einer Ablehnungsmail an die anfragende externe Organisation abzulehnen.

Abbildung 7

Persönlicher Arbeitsbereich für Kooperationsanfragen

COOPERAID PETER MULLER

REQUESTS ASSIGNED TO ME – OVERVIEW

5 NEW ASSIGNED REQUESTS

3 REQUESTS IN PROGRESS

11 COMPLETED REQUESTS

MY REQUESTS

NEW (5) IN PROGRESS (3) **COMPLETED (11)** ARCHIVED (4)

	Firm Name	Industry	Lead	Request Status		
☐	● Miller & Partner	Health	Peter Muller (You)	Rejected	View	Delete
☐	● FanEx Tools	IT/Technology	Peter Muller (You)	Rejected	View	Delete
☐	● Sports Media AG	Media	Patrick Roser	Rejected	View	Delete
☐	● MerchWear	Retail	Abigail Bloom	Accepted	View	Delete
☐	● DroneLab	IT/Technology	Peter Muller (You)	Rejected	View	Delete
☐	● Sports Media AG	Media	Peter Muller (You)	Rejected	View	Delete
☐	● MerchWear	Retail	Peter Muller (You)	Rejected	View	Delete
☐	● DroneLab	IT/Technology	Peter Muller (You)	Accepted	View	Delete
☐	● Sports Media AG	Media	Peter Muller (You)	Rejected	View	Delete
☐	● MerchWear	Retail	Peter Muller (You)	Rejected	View	Delete
☐	● DroneLab	IT/Technology	Peter Muller (You)	Rejected	View	Delete

Zuletzt bietet das IT-Tool eine Detailansicht zu jeder eingegangenen Kooperationsanfrage. Hier können detailliertere Informationen zu der jeweiligen Kooperationsanfrage abgerufen werden. Zum Beispiel Kontaktdaten der anfragenden Organisation, eine präzise Beschreibung des Produktes und des zu lösenden Problems, die Kosten, die mit einer möglichen Kooperation verbunden sind sowie existierende Referenzen innerhalb der Sportorganisation und allgemein in der Sportindustrie. Zusätzlich werden mehrere Funktionen zur Unterstützung der nachgelagerten Arbeitsschritte angeboten, darunter die automatisierte Beantwortung von Kooperationsanfragen. Weitere Funktionen sind das Ablehnen, Kommentieren, Weiterleiten, Archivieren und Exportieren von Kooperationsanfragen.

Abbildung 8

Detailansicht einer Kooperationsanfrage

COOPERAID PETER MULLER

COOPERATION REQUEST DETAILS Close

SPORTS MEDIA GMBH - MEDIA
 Department: Marketing | Categorization: ● | Request Status: Due Dilligence | Request Submitted: June 20, 2020

FIRM DETAILS [Data from cooperation request form]	CONTACT DETAILS [Data from cooperation request form]
DESCRIPTION OF PRODUCT/SERVICE [Data from cooperation request form]	PITCH VALUE ADDED [Data from cooperation request form]
BUSINESS MODEL [Data from cooperation request form]	COST OF PRODUCT/SERVICE [Data from cooperation request form]
REFERENCES [Data from cooperation request form]	INTERNAL CONTACT(S)/REFERENCE(S) [Data from cooperation request form]
INTERNAL COMMENTS [Data from cooperation request form]	

SportsMedia_PitchDeck

Reject
Comment
Forward
Archive
Export
Contact

Schritt 5 – Evaluation der Alpha-Version

Der fünfte Schritt des DSR-Prozessmodells stellt die Evaluation der Alpha-Version des Artefakts in Form von erneuten Experteninterviews und Usability-Tests (siehe, Barnum (2011)) dar. Hierfür haben wir das konzipierte IT-Tool mit drei weiteren Experten von Sportorganisationen aus der Deutschen Fußball Bundesliga und einem weiteren Experten von einer Sportorganisation aus der Schweizerischen Fußball Super League evaluiert. Um Hauptkomponente (1) - das Formular für die Einreichung von Kooperationsanfragen - zu testen, haben wir zusätzlich Experteninterviews und Usability-Tests mit zwei Start-ups durchgeführt. Somit soll gewährleistet werden, dass

die abgefragten Informationen auch aus Sicht der anfragenden externen Organisationen vollständig sind und einen praktischen Nutzen stiften. Insgesamt haben die Ergebnisse der Evaluation gezeigt, dass alle Experten das konzipierte IT-Tool als sehr positiv wahrgenommen haben. Alle Experten gaben an, dass sie das IT-Tool für die Analyse und Bearbeitung von Kooperationsanfragen in ihrer Sportorganisation respektive ihrem Start-up verwenden würden. Darüber hinaus gaben zwei Experten an, dass das IT-Tool auch für andere Anwendungsfälle in ihrer Sportorganisation von Nutzen sein könnte, wie z.B. dem Beschwerdemanagement oder der Sponsorenakquise.

Insgesamt waren die Experten der Ansicht, dass die Effizienzgewinne, die durch die Analyse, initiale automatisierte Bewertung sowie die Unterstützung bei den nachgelagerten Arbeitsschritten erzielt werden, den größten Mehrwert darstellen. Darüber hinaus bewertete ein Experte die durch das IT-Tool gewonnene Transparenz bei der Analyse und Bearbeitung von Kooperationsanfragen sehr positiv. Schließlich stellte ein weiterer Experte fest, dass das zu entwickelnde IT-Tool eine kollaborative Denkweise und den Informationstransfer fördert und damit generell den kulturellen Wandel hin zu mehr Wissensteilung unterstützt, den die Sportorganisation derzeit durchläuft.

Fazit und Ausblick

Ziel unseres Forschungsprojektes ist es, ein IT-Tool zu entwickeln, welches als zentralisiertes und standardisiertes Eingangstor Kooperationsanfragen in Sportorganisationen erfasst, diese initial automatisiert bewertet und im Anschluss die nachgelagerten Arbeitsschritte steuert und optimiert. Auf Basis der wissenschaftlichen Methode DSR ist es uns gelungen, ein solches IT-Tool initial zu konzipieren. Die durchgeführten neun Experteninterviews (inkl. sechs Usability-Tests) haben sowohl die Dringlichkeit für solch ein IT-Tool als auch die Nützlichkeit unseres konzipierten IT-Tools unterstrichen. Basierend auf den Usability-Tests von Schritt 5, werden in einer Beta-Version des IT-Tools einige kleinere funktionale und gestalterische Verbesserungsvorschläge eingebaut. Um unser IT-Tool nun in Sportorganisationen etablieren zu können, benötigt es nicht nur (1) die reine Implementierung, Systemintegration und Anpassung des IT-Tools, sondern gerade auch (2) Change-Management-Prozesse in der Sportorganisation selbst. So müssen neue Rollen und Verantwortlichkeiten definiert und die Bereitschaft der Mitarbeiter gestärkt werden, mit dem entwickelten IT-Tool zu arbeiten. Um unser IT-Tool zu entwickeln und weiter zu verbessern, suchen wir Sportorganisationen, die an einer Zusammenarbeit interessiert

sind. Falls Sie und ihre Sportorganisation sich von unserem IT-Tool angesprochen fühlen, können Sie sich gerne unter arne.gruettner@unisg.ch bei uns melden.

Referenzen

- Barnum, C. M. (2011). *Usability Testing Essentials: Ready, Set... Test!* Morgan Kaufmann Publishers.
- Geringer, J. M. (1991). Strategic Determinants of Partner Selection Criteria in International Joint Ventures. *Journal of International Business Studies*, 22(1), 41–62. <https://doi.org/10.2307/155239>
- Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. *Proceedings of the 25th Americas Conference on Information Systems (AMCIS) 2019*.
- Hevner, A. R. (2007). A Three Cycle View of Design Science Research. In *Scandinavian Journal of Information Systems* (Vol. 19, Issue 2).
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Holmberg, S. R., & Cummings, J. L. (2009). Building Successful Strategic Alliances. Strategic Process and Analytical Tool for Selecting Partner Industries and Firms. *Long Range Planning*, 42(2), 164–193. <https://doi.org/10.1016/j.lrp.2009.01.004>
- Nielsen, B. B., & Gudergan, S. (2012). Exploration and exploitation fit and performance in international strategic alliances. *International Business Review*, 21(4), 558–574. <https://doi.org/10.1016/j.ibusrev.2011.07.001>
- vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. *Proceedings of the 17th European Conference on Information Systems (ECIS) 2009*.

*Zu viele Sommer gesehen durch
Fenster von Bibliotheken.*

— Prinz Pi,
Kompass ohne Norden, 2013.

Curriculum Vitae

Arne Grüttner

Personal Data

Date of Birth	09/11/1992
Nationality	German

Education

04/2018 – today	University of St.Gallen, St. Gallen, Switzerland Ph.D. in Management
10/2015 – 03/2018	University of Cologne, Cologne, Germany Master of Science in Information Systems
09/2016 – 01/2017	Tsinghua University, Beijing, China Exchange Semester
09/2012 – 09/2015	Traineeship and Studies at Deutsche Telekom AG, Münster, Germany Bachelor of Science in Business Informatics

Professional Academic and Practical Experience

04/2018 – today	Research Associate at the University of St.Gallen, Institute of Information Management, St. Gallen, Switzerland
11/2017 – 04/2018	Research Assistant at the University of Cologne, Chair for Information Systems and Systems Engineering, Cologne, Germany
08/2017 – 10/2017	Consultant at Phoenix Contact Cyber Security AG, Berlin, Germany
05/2016 – 08/2016	Consultant at KPMG Consulting Cyber Security, Cologne, Germany
09/2012 – 09/2015	IT Project Management Trainee at T-Systems International GmbH, Münster, Germany
07/2014 – 10/2014	Sales and Marketing Trainee at T-Systems Singapore Pte Ltd, Singapore
01/2014 – 05/2014	Software Development Trainee at T-Systems International GmbH, Berlin, Germany

Publications

2020 (in press)	Gruettner, A., Wambsganss, T., & Back, A. (2020). From Data to Dollar – Using the Wisdom of an Online Tipster Community to Improve Sports Betting Returns. <i>European Journal of International Management (EJIM)</i> .
2020	Gruettner, A., Braschler, T., & Back, A. (2020). Konfrontiert mit einer Flut von Kooperationsanfragen: So meistern Sportorganisationen die systematische Analyse und Bearbeitung. <i>Wirtschaftsinformatik & Management</i> .
2020	Gruettner, A., Vitisvorakarn, M., Wambsganss, T., Rietsche, R., & Back, A. (2020). The New Window to Athletes' Soul – What Social Media Tells Us About Athletes' Performances. <i>HICSS Proceedings</i> .
2019	Gruettner, A. (2019). What We Know and What We Do Not Know About Digital Technologies in the Sports Industry. <i>AMCIS Proceedings</i> .
2017	Gruettner, A., Richter, J., & Basten, D. (2017). Explaining the Role of Service-Oriented Architecture for Cyber-Physical Systems by Establishing Logical Links. <i>ECIS Proceedings</i> .

Awards and Scholarships

2018	Award – Top 5% Percentile in Information Systems (M.Sc.)
2016	Award – Member of the KPMG highQ-Young-Talent-Program
2015	Award – Top Trainee of Deutsche Telekom AG (B.Sc.)
2014	Scholarship – German Academic Exchange Service (DAAD)

St. Gallen, June 02, 2020