Current Issues in Sport Science 2 (2017)

Scientific approaches to technological officiating aids in game sports

Otto Kolbinger[1,*] & Martin Lames[1]

1 Department of Sport and Health Sciences, Technical University of Munich, Munich, Germany

* Corresponding author: Department of Sport and Health Sciences, Technical University of Munich, Georg-Brauchle-Ring 60/62, München, Germany , Tel: +49 89 28924502,
Email: otto.kolbinger@tum.de

REVIEW ARTICLE

Article History:
Submitted 4th October 2016
Accepted 13th December 2016
Published 23th February 2017

Handling Editor:
Otmar Weiß
University of Vienna, Austria
Martin Kopp
University of Innsbruck, Austria

Editor-in-Chief:
Martin Kopp
University of Innsbruck, Austria

Reviewers:
Keith Lyons
University of Canberra, Australia
Roland Leser
University of Vienna, Austria

Citation:
Kolblinger, O. & Lames, M. (2017). Scientific approaches to technological officiating aids in game sports. *Current Issues in Sport Science, 2:001*. doi: 10.15203/CISS_2017.001

ABSTRACT

An increasing number of game sports use technological officiating aids to support their umpires and referees. The aim of this review is to survey the respective literature to extract universal issues of these aids, which are used in different ways in a wide range of settings. We identified 23 studies, of which the majority was published in the current decade. These studies embraced, beside empirical works, contributions of the fields of philosophy and jurisprudence. Based on the approaches and findings of the selected studies we identified seven major issues: the underlying phenomena, usage patterns, accuracy, standard of review, influence on the nature of the game, material as well as immaterial costs and the amount of authority that is granted to the officiating aid. Further, we found regularly some overlapping between these issues, but also that some matters of interest have not been addressed so far, for example studying the influence of technological officiating aids on stakeholders' opinions. Empirical as well as theoretical evaluations of technological officiating aids have to deal with this complexity. As this seems to be neglected currently, we suggest that further studies should show awareness of this in their approaches as well as in their conclusions.

Keywords:
technological aids – officiating errors – umpires – referees

Introduction

Game sports are defined by sets of rules, which have to be obeyed in order to actually perform these sports (Suits, 1988). To enforce these rules in competitions, the respective associations appoint umpires and referees (these terms are used interchangeably throughout this review). Excluding strict interpretations, not every single incorrect judgement by an umpire might result in an unsuccessful attempt of playing a sport in the views of the majority of the respective stakeholders. Nevertheless, some incorrect calls could rather affect this result, especially if they touch another element of sport defined by Suits (1988): success and failure in competitive sports should be determined by the skills of the participating players or teams. A sporting event that is decided by a fallible umpire could be seen as executed unsuccessfully.

Beyond these fundamental considerations about officiating calls, professional sports have to deal with an additional issue. Incorrect decisions by an umpire that affect the outcome of a match can have significant financial implications. The Union for European Football Associations (UEFA), for instance, shares out a total amount of 1.257 billion € among the teams participating in the UEFA Champions League. The difference of fixed incomes for clubs winning or losing in the play-off stages is about

9.6 million €. This does not include additional income from the "market pools" which is 4.75 times higher for teams participating in the UEFA Champions League compared to those in the UEFA Europa League.

There are several intersecting, sources for judgement errors in officiating, which can be summarized in two groups (for this study we exclude judgement errors that serve criminal purposes): bias and perceptional limitations. The latter group embraces an empirical research base as well as trivial phenomena. Umpires will not be able to judge actions, or at very least perceive these actions in a less ideal way, that take place outside of their field of view as well as actions where they suffer from an obstructed view. Thereby, visual perception does not universally represent the most ideal way, as for some calls auditory perception can be even more important (e.g. foul tips in baseball). All these trivial phenomena can be assigned to perceptional limitations, as our sensory system is not able to fully overcome these obstacles. Oudejans, Verheijen, Gerrits, Steinbruckner, and Beek (2000) point out how poor positioning of assistant referees affects offside calls in football. As soon as the assistant referee is not positioned on the offside line "errors are optically inevitable" (Oudejans et al., 2000, p. 33) and even the kind of error can be predicted. In a follow-up study, Oudejans et al. (2005) identified that assistant referees are positioned off the offside line for 86.5 % of the detected offside decisions. Therefore, they concluded that the majority of erroneous offside calls are affected by poor positioning, which was doubted by Helsen, Gilis, and Weston (2006). According to them, the phenomenon that contributes most to those calls is the flash-lag effect, the tendency of the human eye to perceive a moving object ahead of its actual position. Regardless of the discussion about the prime cause for wrong offside decisions (Helsen, Gilis, & Weston, 2007; Oudejans, Bakker, & Beek, 2007), the underlying experimental research showed empirical evidence for both phenomena. Furthermore, both represent problems that are likely to occur in a variety of officials' decisions in other sports, e.g. the flash-lag effect for line calls in tennis. The same applies for two further sources for incorrect decisions mentioned by Oudejans et al. (2005). Firstly, fast motions of umpires can affect retinal image stabilization and consequently lead to less visual acuity (Crane & Demer, 1997). Secondly, an increasing distance between the umpire and the action to evaluate leads to a decline in the perception of differences in depth (Cutting & Vishton, 1995).

Whereas perceptional limitations are not directly interfering with the umpire's duty of acting impartially, this is true for the second group of sources for judgement errors: phenomena that are usually referred to in scholarly studies as *bias*. Dohmen and Sauermann (2015) summarized the current research base with focus on home bias in football. They identify social forces as source for this kind of bias, leading to a shift in the perceived social rewards of the referee. This is not just supported by the prevalence of beneficial calls in favor of the home team, but also by the findings about the influence of different crowd settings. Pettersson-Lidbom and Priks (2010), as an example, showed that referees did not favor the home team in the same way when spectators were excluded. Similar to other real life settings, like economy or education, another source of bias are race, origin and ethnicity. Parsons, Sulaeman, Yates, and Hamermesh (2011) found that baseball umpires are more likely to call a pitch a strike if they share the race/ ethnicity of the pitcher. Pope and Pope (2015) showed the prevalence of own-nationality bias in the UEFA Champions League by investigating the ratio of foul calls. The third category of bias is reputation, occurring in two ways. Using the same criteria, the evaluation of pitches in baseball, Kim and King (2014) found that umpires are likely to overestimate the quality of pitches of high-status athletes, for example players voted into the All Star game. In addition, this effect is moderated by properties that are attributed to the pitching style of the respective pitchers.

There is no evidence that sport associations are aware of these reasons for officiating errors, but at very least they are aware of the latter's prevalence. More and more game sports and associations are introducing technological officiating aids, as due to the technological progress an increasing amount of devices is available (and also affordable). The NFL (American Football) introduced a replay review system as early as 1986, steadily increasing its field of applications (excluding a temporally ban of the technology in the 1990s). Nowadays there is a wide range of sports using replay reviews, embracing games of different structural patterns like Baseball and Field Hockey. One of the further kind of technologies used to support umpiring are ball tracking devices, used for example in tennis, cricket and football. As these examples show, there are different technologies that are used in a wide range of sports or different settings. Nevertheless, considering that all technologies are introduced to support the officiating process, we assume that there are overarching issues that concern all (or at least several) such aids. Therefore the aim of this review is to survey the literature on technological officiating aids in game sports to extract these universal issues in order to create a fundamental and conceptual basis for further investigations.

Methods

Technological officiating aids in game sports embrace different technologies, used in a wide range of settings. In addition, it turned out that essential contributions were provided by papers of non-empirical fields. Both these issues are not reflected in common guidelines, for instance, the so- called PRISMA statement. This statement, like other guidelines, was developed to overcome the problem of poor reporting of key information that diminishes the value of meta-analysis and systematics reviews (Liberati et al., 2009). Such guidelines help to ensure complete and transparent reporting and authors are more and more expected to use these or similar guidelines. Unfortunately, none of the common guidelines are completely appropriate for studies synthesizing information of such different scientific fields. Nevertheless, this study conforms to the PRISMA guide-

lines as close as appropriate, providing all key information to fulfill the requirements for systematic reviews to the authors' best knowledge and conscience.

Search strategy

The electronic databases Web of Science™ and Scopus were searched through the last week of February 2016. An update literature search was run on May 30 and 31, 2016. The searched terms were inserted in the search field type "Topic" for Web of Science™ and "Article Title, Abstract, Keywords" for Scopus. These search terms were based on two groups. The first group included general terms, not specific to certain sports or technologies, in combination with the term "sport". The used general terms were "Umpir* Technology", "Referee* Technology" and "Technological Aids". The same procedure was applied to a group of specific terms, which still could refer to several settings, including "Line call*", "Hawk*Eye", "Decision Review System", "Snickometer", "Hot Spot", "Replay Review", "Video Replay", "Instant Replay" and "Goal Line Technology". For terms that were just relevant for one specific game sport, the name of this sport replaced the term "sport", e.g. "Cricket" for "Snickometer". No limits were applied for the publication date but only papers in English were considered.

Study selection

Several selection criteria were applied before a study was included in the review. All studies had to be published as proceedings or journal papers, with the respective source ranked by Web of Science™, Scopus or the Washington and Lee University ranking (Law Journals: Submissions and Ranking). Due to the goals of this review, the respective papers had to focus on technological officiating aids that were already approved and used by a sport association. In addition, the study had to refer to real world settings, meaning that pure laboratory studies were excluded.

Organization of the synthesis

The main issues, which are described in detail in the discussion section, were identified through an inductive approach. First, all the selected studies were screened to identify the goals and main research questions of each study. Based on this, overarching issues were independently extracted and discussed. After consensus was reached, all studies were screened again and associated to these issues.

The result section gives the numbers of the complete study selection process and lists the identified studies, including general information as well as information about the investigated technological officiating aid. The discussion section summarizes the evidence and is organized conceptually based on the extracted universal issues. Due to readability, the main findings of each study are stated in the respective sections of the discussion.

Results

Figure 1 shows the results of the selection process, leading to the inclusion of a total of 23 studies for the review. The search procedure provided 401 different citations. 371 studies of these were excluded after checking the respective title and abstracts. After reviewing the full-texts of the articles, eleven more studies were discarded, one of those due to poor quality. Finally, four studies that met all criteria were identified by screening the reference lists of the accessed papers. Nine of the included studies used empirical methods, nine studies used a philosophical approach and five studies originated from the field of jurisprudence. Table 1 provides further details about these studies, including the examined technology and the respective (main) objectives.

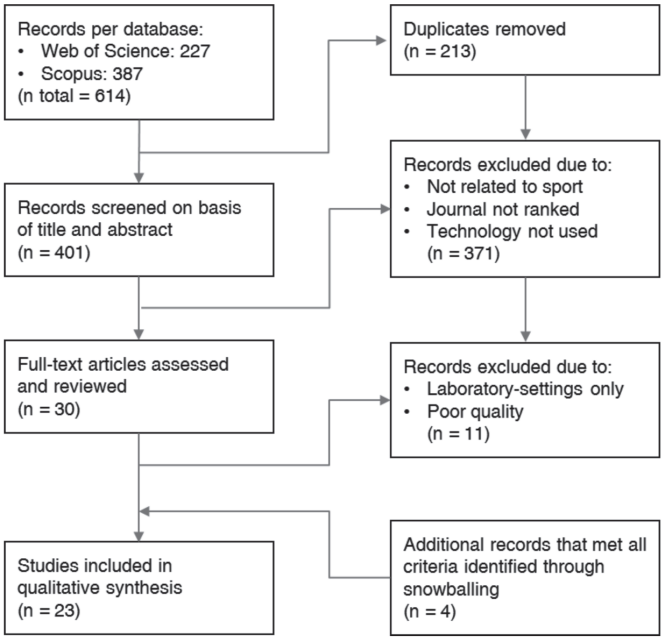

Figure 1: Flow diagram of the study selection

The nine empirical studies covered a wide range of different analyses, with some of the studies touching on more than one technological officiating aid. The hawk eye technology in tennis, which is used for the verification of controversial line calls, was investigated by seven of these studies (Abramitzky, Einav, Kolkowitz, & Mill, 2012; Anbarci, Lee, & Ulker, 2014; Carboch, Vejvodova, & Suss, 2016; Clarke & Norman, 2012; Cross, 2014; Mather, 2008; Nadimpalli & Hasenbein, 2013). The remaining two studies focused on the Decision Review System, covering various decisions in Cricket (Borooah, 2016), and the goal line technology in football (Kolbinger, Linke, Link, & Lames, 2015).

The nine studies that addressed philosophical issues focused on four different technological officiating aids. These included the aids mentioned above; the decision review system in cricket (Steen, 2011), the use of hawk eye technology in tennis as well as in cricket (Collins, 2010; Collins & Evans, 2008, 2012) and

Table 1: Chronological listing of all studies included in the qualitative analysis, including the investigated technological officiating aid, the respective sports in brackets and the main topic

Authors (Year)	Officiating Aid (Sport)	Main Topic
Guggenheim (2000)	Instant Replay (American Football)	Appellate Model
Nafziger (2004)	Video Replay (Several)	Impact of video replay on sports
Collins and Evans (2008)	Hawk-Eye (Tennis & Cricket)	Public Understanding of Science
Mather (2008)	Hawk-Eye (Tennis)	Perceptual Uncertainty
Oldfather and Fernholz (2009)	Instant Replay (American Football)	Appellate Model
Collins (2010)	Hawk-Eye (Tennis & Cricket)	Relationship between umpires and technology/ Impact on justice
Berman (2011)	Instant Replay (American Football)	Appellate Modell
Steen (2011)	Decision Review System (Cricket)	Impact on game's credibility and justice
Vannatta (2011)	Instant Replay (American Football)	Improvement of Umpiring through Instant Replay
Abramitzky et al. (2012)	Hawk-Eye (Tennis)	Optimality of line-call challenges
Clarke and Norman (2012)	Hawk-Eye (Tennis)	Optimality of line-call challenges
Collins and Evans (2012)	Hawk-Eye (Tennis & Cricket)	Public Understanding of Science
Nlandu (2012)	Goal line technology (Football)	Evaluating arguments for goal line technology
Royce (2012)	Video Replay (Several)	Impact on justice
Ryall (2012)	Goal line technology (Football)	Evaluating arguments against goal line technology
Nadimpalli and Hasenbein (2013)	Hawk-Eye (Tennis)	Optimality of line-call challenges
Anbarci et al. (2014)	Hawk-Eye (Tennis)	Gender differences in line-call challenges
Cross (2014)	Hawk-Eye (Tennis)	Accuracy of the footprint
Svantesson (2014)	Several (Football)	Impact on Football
Bordner (2015)	Several (Several)	Impact on justice
Kolbinger, Linke, Link, and Lames (2015)	Goal line technology (Football)	Prevalence of critical goal line decisions
Borooah (2016)	Decision Review System (Cricket)	Imperfections of the Decision Review System
Carboch et al. (2016)	Hawk-Eye (Tennis)	Umpire errors

goal line technology (Nlandu, 2012; Ryall, 2012). In addition, two studies discussed the use of video replay systems in several sports (Royce, 2012; Vannatta, 2011) and two focused on the use of technology to correct calls in general (Bordner, 2015; Collins, 2010).

Legal theorists tended to consider video replay. Three studies analyzed the instant replay in the NFL (Berman, 2011; Guggenheim, 2000; Oldfather & Fernholz, 2009), one paper discussed the use of video replay and other technologies in football (Svantesson, 2014) and one study observed the use of replay reviews in general (Nafziger, 2004).

Thus, jurisprudence was the first scientific field to address the issue of technological officiating aids already in 2000 (Guggenheim, 2000) and contributed three of only five studies overall before 2010. Consequently, video replay represented the most examined technology in this period. Since then, 18 further studies were conducted in the mentioned fields, with six of them published alone in 2012 (see Table one).

Discussion

Based on the emphasis and findings of these studies we identified seven major issues that will be discussed. First, the preconditions and reasons that lead to the introduction of technological officiating aids in game sports, which we refer to as *underlying phenomena*. Further, the respective *usage and "success" rates,* where success is meant in the sense of corrected initial calls. The latter is also affected by the so-called *standard of review*, the influence of the initial call of the referee on the review process and its outcome. This is also true for the *accuracy* of a technology (a fact that is not considered by all users), which of course also represents an essential prerequisite for a system to provide any merit. Changes in the *nature of the game* concerning officiating and flow are vice versa a consequence of the introduction of an aid. Based on the kind of effect, these might overlap with the issue of *costs and threats* that are affected by such a technology. All of these issues obviously intersect

with the question about the amount of *authority granted* to the technological officiating aids.

Underlying phenomena

Six studies created knowledge about the preconditions and reasons that lead to the introduction of technological officiating aids. Mather (2008) and Carboch et al. (2016) both looked for line judging errors of umpires in tennis. Comparing the pattern of line call challenges to a simple model of uncertainty, Mather (2008) found that the majority of erroneous line calls are affected by perceptual uncertainty. Judgement errors of umpires as well as players increase with decreasing distance between the ball impact and the lines, peaking within 5 mm of the line. Based on Mather's model, 8.2 % of all line calls within 100 mm of the respective line are judged incorrectly. Carboch et al. (2016) added information about the errors of umpires. They found a mean distance of the line for erroneous umpire calls of 33.2 mm in general as well as 20.9 mm and 42.8 mm respectively for long and cross lines (represented by all lines running parallel to the net). Umpire errors occur on average once per 17.4 service games. Furthermore, more errors occur on cross line calls and at the earlier stages of a tournament.
Concerning the use of a similar technology in a different setting, the goal line technology, Kolbinger et al. (2015) examined the prevalence of critical goal line decisions. Based on the investigation of 1167 matches, they found 19.6 critical goal line decisions per season in the Bundesliga and 13.9 in the 2nd Bundesliga. By checking those critical calls through video footage, they concluded that only 5.0 (Bundesliga) and 2.8 (2nd Bundesliga) calls per season could have been solved exclusively by goal line technology.
Beside these empirical studies, three philosophical papers by Collins (2010), Nlandu (2012) and Bordner (2015) focused on the underlying phenomena. Collins (2010) described six constructs of judging and justice in relation to umpiring in general as well as its relationship to technological officiating aids. Umpires are granted with *ontological authority*, which means that their decisions define what happens in a game. This is based on the assumption they own an *epistemological privilege* because of a superior view and their respective skills. These judging constructs lead to *presumptive justice*, which describes that one can fairly assume that justice is seen to be done. Technological progress led to three further constructs of justice. *Transparent justice* and *transparent injustice* respectively, if it can be proved or disproved that a call was judged correctly, and *false transparency* if it is wrongly suggested that justice has been done.
Nlandu (2012) identifies the scapegoating of referees as the overarching reason for the introduction of goal line technology. He identifies two kind of – according to him – false assumptions on which the arguments for technological officiating aids are based: Firstly, umpire decisions affect the results of competitions more than players or teams. Secondly, technological officiating aids ensure the fairness of game outcomes by eliminating wrong decisions. Bordner (2015), in contrary to this, elaborates on the consequences of incorrect decisions of umpires for sports, using four different approaches. Based on a strict interpretation of the relationship between rules and sport, incorrect decisions make it impossible to actually perform the sport. More moderately, but still interfering with the fundamentals of sportive contests, blown calls lead to an unequal enforcement of rules, as one party is put at a disadvantage, which is unfair according to Bordner (2015). Furthermore, it is declared as unjust that such calls deny the athletes something they should have achieved. All these issues are seen as a threat for sports, even if the referee is not acting corruptly.

Usage and "success" rate

Six of the nine empirical studies provided information about the frequency of use of hawk-eye technology in tennis (referred to as usage rate) and the amount of changed calls ("success" rate), except one study that focused exclusively on the latter. Findings about the amount of challenged decisions and the respective share of overturned calls are listed in Table 2. The studies covered different settings in terms of tournament design, surface and gender. Consequently, the "success" rates show a decent range from under 30 % to almost 40 %. As the frequency is presented based on different parameters, which cannot be completely transformed, the results cannot be compared (sets can consist of different numbers of games which again were decided after different number of points). However, it can be extracted that only a small share of all points is checked by the technology. Due to the amount of points in a match, this still means that they occur on average at least about four times per best-of-three match.
These results concerning usage and "success" rate are used for various intentions. As mentioned above, the studies of Mather (2008) and Carboch et al. (2016) looked for perceptual uncertainty and patterns of umpire errors. Anbarci et al. (2014) examined the influence of gender on the use of the challenge system. They found similar utilization patterns, as the usage rate for both increased with the overall number games played and the number of games won by the opponent. Furthermore, challenges at the end of a set are less likely to be successful, whereas second challenges are more successful than first challenges. Gender differences were found as men showed a lower "success" rate in tiebreaks, for balls on the opponents half of the court and made more "embarrassing" challenges (wrong by more than 50 mm). This kind of unsuccessful challenges especially increases as higher the respective player or his opponent was ranked, while higher ranked women were less likely to make such challenges. In addition, women made more unsuccessful challenges if the opponent had a higher ranking.
The remaining three studies all focused on the optimality of line call challenges. According to the model of Clarke and Norman (2012), an optimal challenge strategy would increase the player's probability to win an even set by 5 %, and an even best-of-three match by 8 % (9 % for best-of-five). Therefore,

Table 1: Detailed information about the empirical studies covering the use of hawk eye in tennis, listing the respective sample and results for usage and success rate (in the sense of overturned initial umpire's decisions)

Authors (Year)	Sample (Year)	Usage	"Success" rate
Mather (2008)	15 ATP tournaments (2006 – 2007)	n. a.	39.3 %
Abramitzky, Einav, Kolkowitz, and Mill (2012)	741 ATP matches (2006 – 2008)	3.8 challenges per match; 2.6 % of all points challenged	37.7 %
Clarke and Norman (2012)	Wimbledon (2009)	6.7/ 1.8 challenges per match/ set (men)	29.2 % for women, 29.6 % for men
Nadimpalli and Hasenbein (2013)	Australian Open & Wimbledon (2012)	2.2/ 3.2 % (Australian Open) and 2.2/ 2.9 % (Wimbledon) of all points challenged by women/ men	35.1 % and 25.7 % for women, 31.4 % and 28.0 % for men
Anbarci, Lee, and Ulker (2014)	480 ATP & WTA matches (2006-2008)	0.33 and 0.31 challenges per set respectively for women and men *	32.5 % (Women), 39.8 % (men)
Carboch, Vejvodova, and Suss (2016)	208 ATP matches (2014)	992 challenges in 4691 games (0.21 per game)	27.0 %

*These values can't be reproduced with the provided data, which would result in 0.69 (men) and 0.62 (women) challenges per set. The corresponding author was contacted but didn't reply.

players should save their challenges for important points and challenge more often in the final stages of a set or when they are trailing. They also mention that based on the 1.8 challenges per set, the chance for a player using all of his three challenges is fewer than 2 %. The non-exhaustive use of the available challenges is also discussed by Abramitzky et al. (2012). They found that the player's respective behavior is close to the optimal one, prescribed by a model weighing the importance of points and the remaining challenges in relation to the remaining points. This can be illustrated by the lower "success" rate in close compared to easy service games (37.7 % to 56.9 %), in later compared to earlier stages of sets overall (35.5 % to 41.1 %) and in tie breaks compared to regular games (24.9 % to 39.5 %). Nevertheless, based on their model the authors conclude that the players should challenge more often. Nadimpalli and Hasenbein (2013) tried to develop a Markov decision process to identify the optimal stages to use line call challenges in service games. As main findings they demonstrate that players should challenge if the referee is very likely wrong and challenge the outcome of all converted game-points of their opponent.

Accuracy

Five studies examined the accuracy of technological officiating aids and the consequents respectively, with Cross (2014) representing the only study that actually conducted an empirical evaluation. To verify line challenges in tennis, the hawk eye system uses a footprint interpolated from the measurements of the ball trajectory. This footprint was compared to the results of laboratory experiments measuring the rate of expansion of a tennis ball hitting the surface. Cross (2014) showed that the virtual footprint of predominantly flat impacts, e.g. fast serves, was actually 7.1 % wider than expected. For rather steep impacts the interpolated footprint was 36.2 mm, whereas the experiments predicted 44 mm for a footprint based on the respective kinematic parameters. Analyzing the footprint based on the fuzzy surface of a tennis ball, it was also shown that the edges of a footprint are not well defined and single fibers of the ball can extend the footprint up to 10 mm.

The accuracy concerns of Collins and Evans (2008) are, *inter alia*, based on this footprint. In their philosophical contributions they further elaborate on the process of ball tracking in general and the consequences of its use for sports as well as the public understanding of science. Thereby they are focusing on the use of the hawk eye technology in tennis and cricket. By nature, this technology is unable to provide an accuracy level of 100 %, as the ball trajectory is interpolated on the basis of a limited number of time points and a model including insufficient assumptions about the ball. Further, it is examined whether the trajectory of the ball fulfils certain conditions by embedding it in a modelled environment. To suggest that this virtual reality is actually real, leads to an untrue public understanding of science, especially concerning the accuracy of machines. The authors continued the discussion of this topic (Collins & Evans, 2012) by referencing the use in cricket, where the produced trajectory is illustrated with zones of uncertainty. This kind of use still improves the overall decision making and illustrates the limits of the accuracy of technological officiating aids in a clear way. Furthermore the authors distinguish, like in Collins (2010), between justice and accuracy. The latter is not the only dimension that should be taken into account to make a just decision and might not even represent the most important one. Instead, the continuity with the historical and universal judgement of such decisions could have much more influence on doing justice. This position is highly doubted by Royce (2012), who dedicates

a whole paper to reflect Collins work of 2010. Regarding the relationship between accuracy and justice, he defines the latter solely on the basis of the former.
Vannatta (2011) raises concerns about the accuracy of video replay reviews based on Edmund Husserl's (1929/1973) concept of static and genetic phenomenology. Due to the concept of genetic phenomenology, the best way to perceive lived movement in time is to actually perceive it in lived time. Thus, Vannatta (2011) states that replays should be limited to those decisions that demand for static phenomenology only, which is abstracted from temporality. Rephrased for sport or American football respectively, such decisions are questions like if a player is out of bounds.

Standard of review

As a process of appellate review, the standard of review is a natural topic for legal theorists. Concerning umpiring in sport, this term can be rephrased as the influence of the initial call of the referee on the review process and its outcome. All of the three jurisprudential studies covering this issue focused on the instant replay system in the NFL. In this setting, the standard of review is called "Indisputable Visual Evidence" (IVE), meaning that the call of the umpire can only be overturned if the video footage actually shows that the umpire was incorrect. Guggenheim (2000) raised early concerns regarding the standard itself as well as its application. Based on the share of overturned calls (30 % in his sample, consisting of half of the 1999 NFL Season) he concluded that IVE is not applied correctly, which is inter alia affected by the ambiguous definition. He consequently suggests to replace the current standard by "manifest weight of the evidence", which means that a call can be overturned if it is clearly not supported by the video footage. In contrast to this, Oldfather and Fernholz (2009) describe IVE as an appropriate standard for American Football, as they compare NFL's process of appellate review to those in law. IVE helps to shorten the time of single reviews and also to limit the amount of overall appeals. The latter not just helps to shorten (or rather: less prolong) the total match time, but also to protect umpire's authority.
Berman (2011), dedicating a complete essay to the standard of review, is neither fully supporting Guggenheim's (2000) view nor the one of Oldfather and Fernholz (2009). In contrast to the latter, he questions the shorter duration of review interruptions as replay officials use extra time to confirm the presence or absence of indisputable evidence. Furthermore, as every reversal represents an indisputable error on the initial call, reversals under IVE could be even more costly for the umpire's dignity. Berman (2011) provides a deep elaboration of the standard of review, principally to compare the current one to a so-called de novo standard. This standard would mean that the initial call of the umpire is irrelevant for the review process and therefore the final decision is only defined by the provided evidence of the video footage. Berman's intention is not to resolve the question, which standard of review is the optimal one, even as he makes the case for a less deferential standard. He rather points out the various impacts on and of the standards including inter alia loss aversion, corruption and types of errors. As an example, an advantage of *de novo* would be a higher amount of corrected calls, a disadvantage the increasing threat of erroneous reversals.
The standard of review is not just a question for video replays, as the use of the hawk-eye technology in tennis and cricket shows (Collins, 2010; Collins & Evans, 2012). Cricket relies on the initial umpire's decision if the system is not able to fully disprove this call, principally based on confidential intervals. This is not true in tennis where the final decision is solely determined by the respective output of the technology.

Nature of the game

The influence of technological officiating aids on the nature of the game was touched by seven studies, which can be pooled in two groups. Four studies examined if there is a shift in the overall pattern of calls, among them the studies of Collins (2010) and Collins and Evans (2008, 2012). Those raise concerns that the introduction of hawk eye technology led to a discrepancy in the way calls are judged in professional and amateur games as well as to a shift in the historical judgement. As an example, ambiguous leg before wicket decisions in cricket have always been judged in favor of the batsman, which was (at least for some time) no longer true after the introduction of the respective technological officiating aid. Steen (2011) confirms this trend and provided some more in-depth analysis of the effect in cricket. Investigations of the 2011 world cup showed that the innovation not just benefits the fielding teams in general, but also a specific type of bowler (so-called spinners) and therefore consequently the respective teams.
Three studies discussed the influence on the flow of the game. In an article that tries to disprove Sepp Blatter's arguments (former head of the FIFA) against goal line technology, Ryall (2012) picks up the concerns about the rhythm of the game. Regarding this respective technology, the problem can be solved easily as this fact can be checked in real time. This is not true for all technologies, which is also stated by Bordner (2015). Referencing media reports about the NFL, which actually showed an increase of the total match duration of one second, he further argues that matches are not substantially prolonged. Bordner (2015) also focus on the major sports in the US for his third argument, as he points out that there are already several interruptions in the game for less meaningful reasons like commercial breaks. Furthermore, it is possible that fans rather enjoy the interruptions or that the benefit of correcting calls at least outweighs the increase of interruptions. This is also addressed in Svantesson's (2014) elaborations about the use of technological officiating aids by the FIFA. On the other hand, he raises concerns about an increasing risk for injuries due to the interruptions.

Costs and threats

The introduction of technological officiating aids affects physical as well as immaterial costs. The latter could also include the changes of the nature of the game, of course, depending on the respective opinions. This would be true for Collins (2010) and Collins and Evans (2012) which also identify a further threat. As the current use of the hawk-eye system in tennis implies that the outcomes are 100 % accurate, it creates false transparency. Especially for close calls it is likely that justice appears to be done when it actually is not. Another threat caused by technological officiating aids presents the undermining of the referee's authority. Nlandu (2012) identifies this concern as one of the main concerns of high-level representatives of associations, in his example the FIFA and UEFA. Consequently this concern is also discussed by Ryall (2012), who denies this phenomenon as technologies in other sports rather demonstrated umpires are right most of the time. In contrast to this, Nafziger (2004) already sees referee's authority and legitimacy undermined as the referee is not, by nature, right all the time anymore.
Two empirical studies set up material cost-benefit analyses for the use of technological officiating aids. Borooah (2016) calculated such an analysis for leg before wicket decisions in cricket, using the costs for the Decision Review System (about 60,000 US$ per match day) and the number of corrected decisions. Thus, he identified an average cost of 54,545 US$ per one percent more correct decisions for his (small) sample, the Ashes series of 2013. By a series of simulations he concluded that the goal of more correct decision could be more economically achieved by improving the umpires instead of using technology. Concerns about an appropriate cost-benefit-relation are also stated by the study of Kolbinger et al. (2015), examining the necessity of goal line technology. To equip a stadium with this technology requires expenses of 135,000 € per season. Based on the frequency of relevant incidents mentioned above (4.0 and 2.8 respectively per season for the whole Bundesliga and 2nd Bundesliga), the Bernoulli probability for one mandatory use of the technology in a stadium of the Bundesliga doesn't reach 95 % until the 11th season. For the 2nd Bundesliga it would even take 20 seasons.

Authority granted

As already shown, several intersections between the different topics may occur. This is especially true for the question, which amount of authority should be awarded to the technological officiating aid. The most dismissive attitude towards granting authority to such technologies is stated by Nlandu (2012), who rather suggests the sport associations should focus on the underlying ethical fallacies. Therefore, he suggests installing a system of sport ethics education, even providing incentives for good behavior.
Other skeptical contributions are not going that far, but would reduce or respectively limit the use of these aids. As already shown in the accuracy section, Vannatta (2011) uses concepts of phenomenology to show why instant replays should be limited to static incidents. Collins (2010) as well as Collins and Evans (2008, 2012), based on their view of the relationship between accuracy and justice, see technologies as a decision aid, not a decision maker. They consequently conclude that the authority to make the calls should just shift to the technology for obvious errors, but stick with the umpire for close calls. This view is completely supported by Borooah (2016) and at least partly by Royce (2012). The latter criticizes the approach concerning accuracy, justice and error definitions, but states similar conclusions. Technologies can just provide further information, but not how the respective incident has to be judged. Therefore, human decision-making can't be replaced in refereeing. A human element of judging is also raised by Nafziger (2004), who argues that the material and immaterial costs of such technologies often seem to outweigh the flaws of human decision making. Nevertheless, he suggests a "carefully managed use … to overcome the vicissitudes of the human eye and conscience"(Nafziger, 2004, p. 27).
On the other hand, two authors would use technological officiating aids as much as possible. Svantesson (2014) states that there is no need to grant the referees more discretion than necessary. The referees' decisions should be based on facts as much as possible and the settings of sport events, in his case football matches, provide an ideal setting to do so. Bordner (2015) bases his positivistic view on the ideal to get all calls consistently right. This should be an ultimate goal, as we devote our time to sports, and often the best way to get calls right is to use technological aids. For both authors the benefits of such aids outweigh the costs and interestingly both suggest to expand the use of technological officiating aids also to lower levels as soon as (or respectively: as long as) it is affordable.

Conclusions

The purpose of this review was to create a basis for further investigations of technologies that support referees in game sports, which includes semantics and a respective taxonomy. In a review about technologies in sport in general, outside the scope of this review, Dyer described two groups of "non-human decision-making technology" called "video replay technology" and "line judgement technology" (Dyer, 2015). We think this classification is neither appropriate nor sufficient. First, the technologies pooled as line judgement technologies also resolve problems that do not include line calls. In addition, one technology can be used in fundamentally different ways, as inter alia Collins (2010) shows for the use of hawk eye in tennis and cricket.
Therefore, we think it is more appropriate to build the taxonomy for technological officiating aids based on the different kinds of contribution to the officiating process. Collins and Evans (2012) have shown two of these. On the one hand, there are technologies that support the decision-making process of referees; on the other hand, technologies are used to replace

umpires for specific decisions. We argue that there is a third kind of contribution, technological aids that help the referee to enforce rules. This category includes for example the currently introduced vanishing spray, to mark the required distance of defending players in football. The denotation "sport decision aids", as introduced by Collins and Evans (2012), consequently does not include all these kinds of contributions and technologies. Therefore we suggest *Technological Officiating Aids* as overarching term for all technologies that support referees and umpires respectively in game sports.

Screening the studies that examined such aids in empirical fields as well as in philosophy and jurisprudence, we extracted seven major issues, which are described above: *underlying phenomena, usage - and "success" rates, standard of review*, *accuracy*, *nature of the game*, *costs and threats* and *authority granted*. Further fundamental demands of common evaluation research theories have not been addressed with respect to technological officiating aids by the scientific community so far, for example surveys on and off stakeholders' opinions (Guba & Lincoln, 1989). However, the number of different issues as well as their numerous intersections already demonstrates the wide range of factors that need to be taken into account to evaluate the use of technological officiating aids. The existing literature shows implicitly as well as explicitly a lack of awareness of the complexity, but also that exhaustive investigations are necessary to deal with single matters of technological officiating aids in game sports. Thus, we do not suggest that further studies should obligatory examine the complete spectrum of issues. We rather suggest that studies examining aspects of technological officiating aids in game sports should be aware of the complexity as well as of the various intersections and demonstrate this awareness in their approach and their conclusions.

Acknowledgements

The term Technological Official Aids was suggested by an anonymous reviewer and we are very grateful for this. We think it is a substantially better fit than our previous term Technological Umpiring Aids.

Funding

This work was supported by the German Research Foundation (DFG) and the Technische Universität München within the funding programme Open Access Publishing.

Competing Interests

The authors have declared that no competing interests exist.

Data Availability Statement

All relevant data are within the paper.

References

Abramitzky, R., Einav, L., Kolkowitz, S., & Mill, R. (2012). On the optimality of line call challenges in professional tennis. *International Economic Review*, *53*(3), 939-964. doi: 10.1111/j.1468-2354.2012.00706.x

Anbarci, N., Lee, J., & Ulker, A. (2014). Win at all costs or lose gracefully in high-stakes competition? Gender differences in professional tennis. *Journal of Sports Economics*, *17*(4), 323-353. doi: 10.1177/1527002514531788

Berman, M. N. (2011). Replay. *California Law Review*, *99*(6), 1683-1743. doi: 10.2307/41345443

Bordner, S. S. (2015). Call 'em as they are: What's wrong with blown calls and what to do about them. *Journal of the Philosophy of Sport*, *42*(1), 101-120.

Borooah, V. K. (2016). Upstairs and downstairs the imperfections of cricket's decision review system. *Journal of Sports Economics*, *17*(1), 64-85.

Carboch, J., Vejvodova, K., & Suss, V. (2016). Analysis of errors made by line umpires on ATP tournaments. *International Journal of Performance Analysis in Sport*, *16*(1), 264-275.

Clarke, R. S., & Norman, M. J. (2012). Optimal challenges in tennis. *Journal of the Operational Research Society*, *63*(12), 1765-1772. doi: 10.1057/jors.2011.147

Collins, H. (2010). The philosophy of umpiring and the introduction of decision-aid technology. *Journal of the Philosophy of Sport*, *37*(2), 135-146.

Collins, H., & Evans, R. (2008). You cannot be serious! Public understanding of technology with special reference to "Hawk-Eye". *Public Understanding of Science*, *17*(3), 283-308.

Collins, H., & Evans, R. (2012). Sport-decision aids and the "CSI-effect": why cricket uses Hawk-Eye well and tennis uses it badly. *Public Understanding of Science*, *21*(8), 904-921.

Crane, B. T., & Demer, J. L. (1997). Human gaze stabilization during natural activities: translation, rotation, magnification, and target distance effects. *Journal of Neurophysiology*, *78*(4), 2129-2144.

Cross, R. (2014). The footprint of a tennis ball. *Sports Engineering*, *17*(4), 239-247.

Cutting, J. E., & Vishton, P. M. (1995). The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Handbook of perception and cognition. Vol 5: Perception of space and motion* (pp. 69-117). San Diego, CA: Academic Press.

Dohmen, T., & Sauermann, J. (2015). Referee bias. *Journal of Economic Surveys*, *30*(4), *679-695*. doi: 10.1111/joes.12106

Dyer, B. (2015). The controversy of sports technology: a systematic review. *SpringerPlus*, *4*(1), 1-12. doi: 10.1186/s40064-015-1331-x

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Guggenheim, J. A. (2000). Blowing the whistle on the NFL's new instant replay rule: Indisputable visual evidence and a recommended Appellate Model. *Vermont Law Review*, *24*, 567.

Helsen, W., Gilis, B., & Weston, M. (2006). Errors in judging "offside" in association football: test of the optical error versus the perceptual flash-lag hypothesis. *Journal of Sports Science*, *24*(5), 521-528. doi: 10.1080/02640410500298065

Helsen, W., Gilis, B., & Weston, M. (2007). Helsen, Gilis, and Weston (2006): do not err in questioning the optical error hypothesis as the only major account for explaining offside decision-making errors. *Journal of Sports Science*, *25*(9), 991-994. doi: 10.1080/02640410601150488

Husserl, E. (1973). *Cartesianische Meditationen und Pariser Vorträge* [Cartesian meditations and the Paris lectures]. The Hague, Netherlands: Martinus Nijhoff. (Original work published 1929)

Kim, J. W., & King, B. G. (2014). Seeing stars: Matthew effects and status bias in major league baseball umpiring. *Management Science*, *60*(11), 2619-2644. doi: 10.1287/mnsc.2014.1967

Kolbinger, O., Linke, D., Link, D., & Lames, M. (2015). Do we need goal line technology in soccer or could video proof be a more suitable choice: A cost-benefit-analysis of goal line technology in soccer and thoughts about an introduction of video proof. *Sports Science Research and Technology Support*, *556*, 107-118. doi: 10.1007/978-3-319-25249-0_8

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., &. Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. doi: 10.1371/journal.pmed.1000100

Mather, G. (2008). Perceptual uncertainty and line-call challenges in professional tennis. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1643), 1645-1651. doi: 10.1098/rspb.2008.0211

Nadimpalli, V. K., & Hasenbein, J. J. (2013). When to challenge a call in tennis: A Markov decision process approach. *Journal of Quantitative Analysis in Sports*, *9*(3), 229-238.

Nafziger, J. (2004). Avoiding and resolving disputes during sports competition: of cameras and computers. *Marquette Sports Law Review*, *15, 13-27*.

Nlandu, T. (2012). The fallacies of the assumptions behind the arguments for goal-line technology in soccer. *Sport, Ethics & Philosophy*, *6*(4), 451-466.

Oldfather, C. M., & Fernholz, M. M. (2009). Comparative procedure on a sunday afternoon: Instant replay in the NFL as a process of appellate review. *Indiana Law Review*, *43*, 45.

Oudejans, R. R., Bakker, F. C., & Beek, P. J. (2007). Helsen, Gilis and Weston (2006): err in testing the optical error hypothesis. *Journal of Sports Science*, *25*(9), 987-990. doi: 10.1080/02640410600778610

Oudejans, R. R., Bakker, F. C., Verheijen, R., Gerrits, J. C., Steinbrückner, M., & Beek, P. J. (2005). How position and motion of expert assistant referees in soccer relate to the quality or their offside judgements during actual match play. *International Journal of Sport Psychology*, 36(1), 3-21.

Oudejans, R. R., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbruckner, M., & Beek, P. J. (2000). Errors in judging 'offside' in football. *Nature*, *404*(6773), 33. doi: 10.1038/35003639

Parsons, C. A., Sulaeman, J., Yates, M. C., & Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation. *American Economic Review, 101*(4), 1410-1435. doi: 10.1257/aer.101.4.1410

Pettersson-Lidbom, P., & Priks, M. (2010). Behavior under social pressure: Empty Italian stadiums and referee bias. *Economics Letters*, *108*(2), 212-214.

Pope, B. R., & Pope, N. G. (2015). Own-nationality bias: Evidence from Uefa Champions League football referees. *Economic Inquiry*, *53*(2), 1292-1304. doi: 10.1111/ecin.12180

Royce, R. (2012). Refereeing and technology – reflections on Collins' proposals. *Journal of the Philosophy of Sport*, *39*(1), 53-64.

Ryall, E. (2012). Are there any good arguments against goal-line technology? *Sport, Ethics & Philosophy*, *6*(4), 439-450.

Steen, R. (2011). Going upstairs: The decision review system – velvet revolution or thin edge of an ethical wedge? *Sport in Society*, *14*(10), 1428-1440.

Suits, B. (1988). Tricky triad: Games, play, and sport. *Journal of the Philosophy of Sport*, *15*(1), 1-9.

Svantesson, D. J. B. (2014). Could technology resurrect the dignity of the FIFA World Cup refereeing? *Computer Law & Security Review*, *30*(5), 569-573.

Vannatta, S. (2011). Phenomenology and the question of instant replay: A crisis of the sciences? *Sport, Ethics and Philosophy*, *5*(3), 331-342.