



Technische Universität München

Fakultät für Sport- und Gesundheitswissenschaften

Lehrstuhl für Trainingswissenschaft und Sportinformatik

**Innovative technische Hilfsmittel zur Unterstützung von
Schiedsrichtern in Spielsportarten als Gegenstand von
Evaluationsforschung**

Otto Klaus Kolbinger

Vollständiger Abdruck der von der Fakultät für Sport- und Gesundheitswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Philosophie genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Jörg Königstorfer

Prüfer der Dissertation:

1. Prof. Dr. Martin Lames
2. Prof. Dr.-Ing. Veit Senner

Die Dissertation wurde am 19.07.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Sport- und Gesundheitswissenschaften am 25.06.2018 angenommen.

*In Erinnerung an
Margarete und Anton Gröner*

Inhaltsverzeichnis

I	Zusammenfassung	I
II	Abbildungsverzeichnis	II
III	Auflistung der eingebundenen Publikationen.....	III
1	Einleitung	1
1.1	Bedeutung von Regeln im Sportspiel und Fehlerquellen bei deren Anwendung durch Schiedsrichter	1
1.2	Innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern	6
1.2.1	Begriffsbestimmung	6
1.2.2	Historischer Abriss	8
1.2.3	Taxonomie	10
1.3	Problem- und Zielstellungen.....	11
1.4	Verortung der Einzelbeiträge.....	14
2	Methoden.....	16
2.1	Innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in den Spilsportarten und Evaluationsforschung	16
2.2	Evaluationsstrategien für innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Spilsportarten	21
2.2.1	Spezifisches Grundgerüst für Evaluationsstrategien.....	22
2.2.2	Beitrag der Wettkampfdiagnostik am Beispiel der Systematischen Spielbeobachtung	26
2.2.3	Verortung der einzelnen Publikationen im evaluationsstrategischen Grundgerüst.....	27
2.3	Weiterführende Anmerkungen zu der Methodik der einzelnen Studien.....	28
2.3.1	Do we need Goal Line Technology?	29
2.3.2	The use of vanishing spray reduces the extent of rule violations in soccer	

2.3.3	Technological Officiating Aids	31
3	Einzelbeiträge	33
3.1	Do we need Goal Line Technology?	33
3.1.1	Eigene Leistungsbeiträge:.....	33
3.1.2	Zusammenfassung.....	33
3.2	The use of vanishing spray reduces the extent of rule violations in soccer	35
3.2.1	Eigene Leistungsbeiträge.....	35
3.2.2	Zusammenfassung.....	35
3.3	Technological Officiating Aids	37
3.3.1	Eigene Leistungsbeiträge.....	37
3.3.2	Zusammenfassung.....	37
4	Diskussion	39
4.1	Retrospektive Diskussion der Einzelbeiträge	39
4.1.1	Do we need Goal Line Technology?	39
4.1.2	The use of vanishing spray reduces the extent of rule violations in soccer 42	
4.1.3	Technological Officiating Aids	44
4.2	Evaluationstheoretische Position des vorliegenden Grundgerüsts.....	45
4.2.1	Natur von Interventionen im sozialen Setting Spielsportarten.....	46
4.2.2	Verwendung von Wissen aus Evaluationen	50
4.2.3	Bewertung von Interventionen.....	52
4.2.4	Wissenschaftstheoretische Auffassungen.....	55
4.3	Einschränkungen der Dissertation.....	57
5	Fazit.....	59
6	Literaturverzeichnis.....	61
Anhang A	Publikationen	66

A.1	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	66
A.1.1	Erlaubnis zur Einbindung	66
A.1.2	Originalveröffentlichung	67
A.2	The use of vanishing spray reduces the extent of rule violations in soccer	79
A.2.1	Erlaubnis zur Einbindung	79
A.2.2	Originalveröffentlichung	80
A.3	Scientific approaches to technological officiating aids in game sports	87
A.3.1	Erlaubnis zur Einbindung	87
A.3.2	Originalveröffentlichung	88

I Zusammenfassung

Mehr und mehr (Spiel-) Sportverbände setzen auf innovative technische Hilfsmittel zur Unterstützung ihrer Schiedsrichter. In der Wissenschaft wurde sich mit diesem Thema bisher nur wenig auseinandergesetzt, ohne den Forschungsgegenstand zu spezifizieren oder Strategien zur Bewertung dieser Interventionen aufzuzeigen. Hier setzt die vorliegende Dissertation an, indem das Phänomen als Gegenstand von Evaluationsforschung diskutiert, Einzelfälle analysiert und ein evaluationsstrategisches Grundgerüst entwickelt wird.

Dieses basiert auf drei als dynamisch angesehenen Faktorenströmen: dem Wettkampfgeschehen, dem technischen Fortschritt und dem Meinungsbild der Stakeholder. Die Ergebnisse und Erfahrungen aus der praktischen Umsetzung von darauf aufbauenden Evaluationsstrategien für zwei technische Hilfsmittel, die Torlinientechnologie und das Freistoßspray, zeigen die Berechtigung des Grundgerüsts. Damit liefert diese Dissertation ein wissenschaftliches Fundament für Evaluationen von technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern in Spielsportarten.

II Abbildungsverzeichnis

Abb. 1: Chronologische Darstellung der Einführung innovativer technischer Hilfsmittel zur Unterstützung von Schiedsrichtern in Sportsportarten.....	9
Abb. 2: Skizze des idealtypischen Grundgerüsts für Evaluationsstrategien zur Unterstützung von Schiedsrichtern in Sportsportarten.....	22
Abb. 3: Grafische Darstellung der Verortung der Einzelbeiträge.....	27

III Auflistung der eingebundenen Publikationen

Bei der vorliegenden Arbeit handelt es sich um eine kumulative Dissertation, wobei drei Full-Paper (peer-reviewed) aus englischsprachigen, international verbreiteten Publikationsorganen eingebunden wurden:

1. 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. In Cabri, J., Barreiros, J. & Pezarat-Correia, P. (Hrsg.), *Sports Science Research and Technology Support* (S. 107 – 118). Schweiz: Springer International Publishing. DOI: 10.1007/978-3-319-25249-0_8
2. Kolbinger, O. & Link, D. (2016). The use of vanishing spray reduces the extent of rule violations in soccer. *SpringerPlus*, 5:1572, 1-7. DOI: 10.1186/s40064-016-3274-2
3. Kolbinger, O. & Lames, M. (2017). Scientific approaches to technological officiating aids in game sports. *Current Issues in Sport Science*, 2:001, 1-10. doi: 10.15203/CISS_2017.001

1 Einleitung

Die Bezeichnung „*innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten*“ beschreibt ein bereits mehrere Jahrzehnte auftretendes Phänomen der Sportpraxis, welches als Untersuchungsgegenstand der Sportwissenschaft bisher nur sehr limitiert in Erscheinung trat. Dementsprechend notwendig ist eine detaillierte Abhandlung dieses Phänomens im Rahmen der Einleitung, ehe die Relevanz der Arbeit dargelegt wird. Dabei soll zuerst die Grundlage für das Bedürfnis nach diesen Hilfsmitteln beschrieben werden, welche aus der Bedeutung von Regeln im Sport und der Fehlbarkeit von Schiedsrichtern abgeleitet wird (1.1). Eine genaue Klärung des Begriffs, eine Auflistung der historisch und gegenwärtig in Sportspielen eingesetzten technischen Hilfsmittel sowie eine Systematik des Untersuchungsgegenstands finden sich unter 1.2. Es folgt die Beschreibung der damit verbundenen Problematik sowie die Formulierung der Zielstellung dieser Arbeit (1.3), ehe unter 1.4. die Beiträge der drei Artikel dieser kumulativen Dissertation entsprechend der Zielstellung eingeordnet werden.

1.1 Bedeutung von Regeln im Sportspiel und Fehlerquellen bei deren Anwendung durch Schiedsrichter

„Sportspiele sind Sportarten mit international kodifiziertem Regelwerk, bei denen zwei Parteien (Einzel, Doppel oder Mannschaften) in einen Interaktionsprozess eintreten, der dadurch zustande kommt, dass beide Parteien gleichzeitig ihr eigenes Spielziel anstreben und verhindern wollen, dass die gegnerische Partei ihr Spielziel erreicht; das Spielziel der Sportspiele ist eine in den Regeln festgelegte, symbolische Handlung.“(Lames, 1991, S. 33)

Bereits definitionsgemäß kommt den Regelwerken in den Sportspielen eine enorme Bedeutung zu. Die zugrundeliegende Auffassung des Begriffs Sportart im Sinne dieser Definition versteht Lames (1991) als „eine institutionalisierte Form des Sporttreibens [...], die nach kodifizierten Regeln abläuft“ (S. 32). Sowohl das Spielziel als auch der Weg zu diesem, werden durch genaue Regeln festgelegt, welche eingehalten bzw. erfüllt werden müssen. Lames (1991) verweist dabei zudem auf die

besondere Bedeutung der international kodifizierten Regelwerke für den Sportspielbegriff aus leistungsdiagnostischer Sicht. Suits (1988) sieht Regeln gar als des Pudels Kern von (Sport-) Spielen, da erst das jeweilige Regelwerk einer Sportart die Entstehung sportartspezifischer Fähigkeiten hervorrufe. So führt beispielsweise die Untersagung absichtlicher Handspiele für Feldspieler im Fußball zur Entwicklung der Fähigkeit, das Spielziel mit anderen Mitteln (Körperteilen) zu verwirklichen. Zur Einhaltung der Regeln werden von den jeweiligen Sportverbänden Schiedsrichter bzw. Schiedsrichtergespanne eingesetzt. Diese sind als menschliche Wesen von Natur aus fehlbar, wobei die Quellen bzw. Ursachen der Fehler zwei Obergruppen zugeordnet werden können: der Wahrnehmung und der Informationsverarbeitung. Von Nafziger (2004) wurden diese, ohne den expliziten Versuch einer Einteilung, treffend als Launen der menschlichen Wahrnehmung und des Gewissens beschrieben.

Unter den Einschränkungen aufgrund der menschlichen Wahrnehmung lassen sich bereits intuitiv triviale Fehlerquellen identifizieren. Spielen sich Szenen außerhalb des Sichtfeldes des Schiedsrichters ab, beispielsweise hinter seinem Rücken, kann er bereits keine Entscheidung aufgrund seiner visuellen Wahrnehmung treffen. Wiederum für andere Szenen sind Schiedsrichter speziell auf die Verarbeitung auditiver Reize angewiesen, zum Beispiel Foul-Tips im Baseball¹, welche durch andere Lärmquellen (Zuschauer) mindestens erschwert wird.

Daneben gibt es komplexere Zusammenhänge und Phänomene, welche zu fehlerhafter Wahrnehmung führen können. Bereits mehrere Studien zeigten den Einfluss der Positionierung auf die Entscheidungsqualität von Schiedsrichtern in verschiedenen Sportarten (Baseball: Ford, Goodwin & Richardson, 1996; Fußball: Oudejans et al., 2005, 2000; Basketball: Schmidt & Bloch, 1980). Die Arbeiten von Oudejans et al. (2005; 2000) demonstrieren beispielsweise, wie die Positionierung des Schiedsrichters zu optischen Täuschungen bei Abseitsentscheidungen führen kann, was die Autoren als Hauptursache für derartige Fehlentscheidungen beschreiben. Letzteres wird von Helsen, Gilis und Weston (2006) in Frage gestellt, welche den sogenannten Flash-Lag-

¹ Die aufgeführte Sportspieldefinition gilt laut Lames (1991) nicht uneingeschränkt für u.a. Baseball, welches als Schlagball- und Abwurfspiel nach Döbler klassifiziert wird. Unter 1.2. wird erläutert, inwiefern Baseball, Cricket und Softball aus Sicht des Autors die genannte Definition erfüllen und deshalb für diese Arbeit als Sportspiel betrachtet werden.

Effekt als häufigste Fehlerursache identifizieren. Dieser Effekt beschreibt das Phänomen, dass das menschliche Auge sich schnell bewegende Gegenstände als weiter in Bewegungsrichtung fortgeschritten wahrnimmt, als sie tatsächlich fortgeschritten sind. Unabhängig von der hierfür nicht relevanten Frage nach der Hauptursache, welche beide Autorengruppen darauffolgend diskutieren (Helsen, Gilis & Weston, 2007; Oudejans, Bakker & Beek, 2007), werden für beide Phänomene stichhaltige Beweisführungen geliefert. Zudem kann, wie bereits aufgezeigt, jeweils die Gültigkeit für weitere Entscheidungen im Sport beansprucht werden, beispielsweise für Vorwärtspässe im Rugby (optische Täuschung aufgrund der Positionierung) oder knappe Torlinienentscheidungen im Handball (Flash-Lag-Effekt).

Fehlentscheidungen, die auf die Launen des menschlichen Gewissens (Nafziger, 2004) zurückzuführen sind, lassen sich in zwei Arten unterteilen. Werden Fehlentscheidungen bewusst getroffen, liegen korrupte oder kriminelle Absichten zugrunde, wovon (leider) eine Reihe von Fällen in der Sportgeschichte zeugt. Ein prominentes Beispiel in Deutschland ist Robert Hoyzer, der 2006 aufgrund von Beihilfe zum Betrug in nachweislich sechs Fällen schuldig gesprochen wurde (BGH, Urteil vom 15.12.2006, 5 StR 181/06, juris, RN 1). Diese Beihilfe bestand unter anderem aus der Manipulation des Ausgangs von Fußballspielen durch absichtliche Fehlentscheidungen, darunter Spiele des DFB-Pokals und der 2. Bundesliga.

In dieser wissenschaftlichen Abhandlung soll sich aber vor allem mit den unbewussten Einflüssen des Informationsverarbeitungsprozesses beschäftigt werden. Diese hängen in der Regel mit unbeabsichtigten Neigungen zusammen, identische Handlungen verschiedener Personen oder Gruppen unterschiedlich zu bewerten, welche wissenschaftlich mit unter dem Begriff *Bias* gesammelt werden. Ein bereits intensiv untersuchtes Thema ist der sogenannte *Home Bias*, die Tendenz, Entscheidungen eher zu Gunsten der gastgebenden Mannschaften oder Athleten zu treffen. Dohmen und Sauermann (2015) liefern eine umfassende Literaturarbeit zu diesem Thema², in der sie sich aber vor allem der entsprechenden Ausprägung im Fußball widmen. Sie tragen jeweils eine Reihe von Studien zur Bevorteilung der

² Da dieser Absatz einen Themenüberblick darstellt und die Zusammenfassung von Dohmen und Sauermann (2015) bezüglich des *Home Bias* als erschöpfend angesehen werden kann, wird für die entsprechenden Erkenntnisse auf eine Aufführung der Primärquellen verzichtet.

Heimmannschaft im Fußball bezüglich Nachspielzeit, zugesprochener Tore, Elfmeter und Verwarnungen zusammen. Die zusammengetragenen Primärquellen liefern darüber hinaus aber auch Beispiele für *Home Bias* im American Football, Baseball, Basketball und Eishockey. Dohmen und Sauermann (2015) geben zudem den aktuellen Stand (mit einer Ausnahme) einer anderen Ausprägung von *Bias* wieder, deren Grundlage das Teilen - bzw. Unterschiede in - der ethnischen Gruppe oder Nationalität sind. Dieses heikle Phänomen wurde bereits in den Sportspielen Australian Football, Baseball, Basketball, Rugby und inzwischen auch Fußball (Pope & Pope, 2015; nicht Teil des Reviews) nachgewiesen. In eine weitere Art des *Bias* unterscheiden Dohmen und Sauermann (2015) hingegen nicht, und zwar den Einfluss der Reputation von Athleten, obwohl zumindest eine derartige Studie genannt wird (Mills, 2014). So zeigen Kim und King (2014), dass Pitches von All-Star-Pitchern sowie Pitches von Werfern, denen eine hohe Wurfgenauigkeit attestiert wird, tendenziell freundlicher bewertet werden. Reputation kann demzufolge in zwei Erscheinungsformen auftreten, bezüglich des Status i.S.v. Erfolg innerhalb der Domäne sowie i.S.v. bestimmten zugeschriebenen Attributen.

Es gibt also eine Vielzahl von Gründen, zusammengefasst in den zwei übergeordneten Kategorien Wahrnehmung und Informationsverarbeitung, für eine fehlerhafte Anwendung von Regeln, welche zu wiederum vielfältigen Konsequenzen führen (können). Basierend auf einer formalistischen Auslegung des Verhältnisses zwischen Regeln und Sportspielen, wonach Letztere ausschließlich durch Erstere definiert werden, würde eine fehlerhafte Regelanwendung zu einem erfolglosen Versuch das Sportspiel zu betreiben bedeuten (Reid, 2012). Wobei *erfolglos* in diesem Fall nicht eine Niederlage im entsprechenden Wettkampf beschreibt, sondern dass das entsprechende Sportspiel schlicht nicht ausgeübt wurde (sondern maximal ein anderes Spiel). Laut Bordner (2015) würde dies jedoch die Frage aufwerfen, ob jemals ein Sportspiel erfolgreich durchgeführt wurde.

Auch Reid (2012) lehnt eine derartige strenge Auslegung ab und stellt vielmehr die Frage, welches Ausmaß und welche Art von Regelverstößen (unabhängig von der Ahndung des Schiedsrichters) zu einem solchen verfehlten Versuch führen. Eine Lösung für dieses Problem könnte die Unterscheidung in konstitutive und andere (zum

Beispiel regulative) Regeln liefern. Reids (2012) Beispiel für einen konstitutiven Regelbruch, Diego Maradonas *Hand Gottes*, ist gleichzeitig ein gutes Beispiel für die Verletzung einer weiteren metaphysischen Grundlage von Sportspielen nach Simon (2007). Sportliche Wettkämpfe, in unserem Fall Sportspiele, sollen durch die spezifischen Fähigkeiten der Athleten entschieden werden. Dies liegt natürlich nicht alleine in der Hand des Schiedsrichters, wie beispielsweise Untersuchungen zum Zufall im Fußball zeigen (u.a. Lames, 2006). Trotzdem kann der Schiedsrichter durch (unbeabsichtigte und beabsichtigte) Fehlentscheidungen dazu beitragen, dass das Endresultat eines Wettkamps nicht die erbrachten Leistungen widerspiegelt. Bordner (2015) führt dies vor allem darauf zurück, dass falsche Entscheidungen in ihrer Anzahl und ihrem Ausmaß sehr wahrscheinlich nicht gleichmäßig auf die teilnehmenden Parteien aufgeteilt sind, weshalb Schiedsrichterfehler zudem *unfair* sind. Darüber hinaus sind diese auch *ungerecht*, da dem Athleten bzw. der Mannschaft nicht der Vorteil zugesprochen wird, den er bzw. sie sich aufgrund der sportlichen Handlung eigentlich verdient hätte/n (Berman, 2011; Bordner, 2015).

Neben diesen internen Auswirkungen können Fehlentscheidungen auch Konsequenzen für externe Ziele des Sports bedeuten, zum Beispiel direkt durch verpasste Preisgelder oder Einnahmen aus dem weiteren (oder nächsthöheren) Wettbewerb. Hinzu kommen indirekt verpasste Einnahmen durch beispielsweise einen Imageverlust bzw. ausbleibenden Imagegewinn (Stadtman, Weigand & Wissmann, 2004). Ob Imageverluste aufgrund von Fehlentscheidungen auch die jeweiligen Verbände hinnehmen müssten, wurde noch nicht empirisch geklärt, wird jedoch beispielsweise von Svantesson (2014) implizit angenommen.

Die fehlerhafte Anwendung von Regeln ist also in vielerlei Hinsicht negativ, weshalb Sportverbände besonderes Interesse daran haben sollten, den Konsequenzen vorzubeugen bzw. diese zu mindern. Letzteres könnte beispielsweise durch Kompensationszahlungen oder Wiederholungsspiele bewerkstelligt werden, wobei Kompensationszahlungen ausschließlich dem Ausgleich finanzieller Einbußen dienen würden³. Eine Wiederholung des Wettkampfs knüpft auch an den internen

³ So leistete beispielsweise die FIFA dem irischen Fußballverband, aufgrund einer Fehlentscheidung im entscheidenden Qualifikationsspiel zur WM 2010, eine Kompensationszahlung in Höhe von 3.6 Mio. Pfund. Bei der Fehlentscheidung handelte es sich um ein absichtliches Handspiel des

Konsequenzen an, erfordert aber erhöhten Organisationsaufwand speziell in laufenden Wettbewerben⁴. Für ein vorbeugendes Vorgehen müssen die Verbände an der Ursache ansetzen, sprich fehlerhafte Regelanwendungen der Schiedsrichter minimieren. Dies könnte zum Beispiel durch weitere Investitionen in die Ausbildung der Schiedsrichter oder der Erhöhung der Schiedsrichteranzahl pro Spiel bewerkstelligt werden. Das Mittel, mit dem sich aber in dieser Arbeit im Speziellen beschäftigt werden soll, sind innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten, welche im folgenden Kapitel näher beschrieben und systematisiert werden.

1.2 Innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern

1.2.1 Begriffsbestimmung

„Innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten“ scheint auf den ersten Blick ein Begriff mit eher geringer Trennschärfe zu sein, was mit an den vielfältigen Verwendungsmöglichkeiten der Wörter Innovation und Technik liegen mag. So beschreibt beispielsweise der Begriff Technik laut Duden die „Gesamtheit der Maßnahmen, Einrichtungen und Verfahren, die dazu dienen, die Erkenntnisse der Naturwissenschaften für den Menschen praktisch nutzbar zu machen“, aber auch die „besondere, in bestimmter Weise festgelegte Art, Methode des Vorgehens, der Ausführung von etwas“ (Duden online, 2016b). Diese Arbeit beschäftigt sich mit Technik der erstgenannten Bedeutungsweise i.S.v. technischer Ausrüstung. Um diese technischen Hilfsmittel noch genauer einzuschränken, wurde der Begriff um „innovativ“ ergänzt, in Anlehnung an die soziologische Bedeutung von Innovation als „geplante und kontrollierte Veränderung, Neuerung in einem sozialen System durch Anwendung neuer Ideen und Techniken“ (Duden online, 2016a). Damit soll der Untersuchungsgegenstand der Arbeit abgegrenzt werden von trivialen bzw. traditionellen Ausrüstungsgegenständen oder Hilfsmitteln des Schiedsrichters, beispielsweise Trillerpfeifen oder Korbnetzen. Kurz gesagt behandelt diese Arbeit

französischen Angreifers Thierry Henry, welches das Tor zum 1:1 Endstand nach Verlängerung vorbereitete.

⁴ Trotzdem gibt es auch hier Beispiele aus der Sportgeschichte, zum Beispiel das Wiederholungsspiel zwischen den Fußballbundesligisten FC Bayern München und 1. FC Nürnberg in der Saison 1993/1994, aufgrund des sogenannten *Phantomtors*.

neuartige, aber bereits offiziell angewandte technische Hilfsmittel, welche zur Unterstützung von Schiedsrichtern in den verschiedenen Sportspielen eingeführt wurden. Im weiteren Verlauf der Arbeit wird aus stilistischen Gründen der Begriff nicht immer vollständig wiedergegeben. Ist also von beispielsweise „technischen Hilfsmitteln“ die Rede, handelt es sich ausschließlich um solche im Sinne des definierten Begriffes.

Unter den Sportspielbegriff fallen hierbei nicht nur die von Lames (1991) aufgeführten Sportarten, sondern auch die nach Döbler als Schlagball- und Abwurfspiele klassifizierten. Lames führt die Ausgrenzung auf die Besonderheit zurück, dass das in einem Moment verfolgte Ziel der beiden Mannschaften nicht „das gleiche ist“ (Lames, 1991, S. 33). Im Hinblick auf die Aufnahme der sogenannten Schlag- und Abwurfspiele ließe sich hier in zwei Richtungen argumentieren. Erstens, anhand einer sehr genauen, durchaus kleinkarierten Betrachtung der Lames'schen Definition. Diese enthält nicht explizit die Forderung, dass es sich gleichzeitig um die Erreichung des gleichen Spielziels handeln muss. Selbst wenn er dies implizit unterstellt, wie sich dem oben aufgeführten wörtlichen Zitat entnehmen lässt, stellt sich die Frage nach der Auslegung des Adverbs *gleichzeitig*. Es kann schlicht nicht ausschließlich i.S.v. exakt zur jeweils selben (Plank-) Zeit gesehen werden. Greift man Lames' (1991) Beispiel auf, wonach zwei Tennisspieler gleichzeitig versuchen, einen Punkt zu erzielen, lässt sich entgegengesetzten, dass es in einem Zyklus zwischen zwei Schlägen eines Spielers Phasen gibt, in denen ausschließlich der jeweils gegnerische Spieler aktiv sein eigenes Spielziel anstrebt. Derartige Phasen lassen sich selbst in den sogenannten Tor-, Mal- und Korbspielen finden, in jeweils unterschiedlichen Ausmaßen. *Gleichzeitig* muss also auch für die Lames'sche Definition eher weich gesehen werden, aber freilich weisen Schlagball- und Abwurfspiele eine striktere zeitliche Trennung auf, in welchen Phasen die Parteien ihr Spielziel im Sinne von Erreichung der festgelegten symbolischen Handlung erreichen können. Die zweite Argumentation basiert auf der Tatsache, dass Lames (1991) diese Sportspiele ohnehin nicht pauschal ausgrenzt, sondern vielmehr für seinen bestimmten Zweck „einer trainingswissenschaftlichen Leistungsdiagnostik auf der Spielwirksamkeitsebene“ (Lames, 1991, S. 34). Für den Zweck der Untersuchung technischer Hilfsmittel zur Unterstützung von Schiedsrichtern ergibt sich die Notwendigkeit dieser Ausgrenzung aber nicht. Hierfür bietet sich die Einteilung per

Wettkampfstruktur nach Hohmann, Lames und Letzelter (2014) an, welche nach Kraft- und Ausdauersportarten, technisch-kompositorische Sportarten sowie Spiel- und Kampfsportarten unterscheidet. Alle Wettkämpfe haben gemein, dass eine siegende Partei nach festen Regeln ermittelt wird, jedoch werden lediglich in der letzten Gruppe Schiedsrichter als eine Art Spielleiter eingesetzt. Den Unterschied zwischen Kampf- und Sportspielen stellt dabei die Tatsache dar, dass bei Letzteren das Spielziel eben durch die Erfüllung einer symbolischen Handlung erreicht wird. Die Frage nach der Erfüllung dieses Spielziels ist vielfach der Grund für die Einführung technischer Hilfsmittel zur Unterstützung von Schiedsrichtern, weswegen sich diese Arbeit auf die Betrachtung von Sportspielen konzentriert.

1.2.2 Historischer Abriss

Anhand eines historischen Abrisses sollen nun die eingesetzten Hilfsmittel aufgeführt und systematisiert werden, welche unter den Sammelbegriff „Innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten“ fallen. Den Anfang machte die Einführung des Videobeweises in der National Football League (abgekürzt NFL, Sportart American Football) 1986, welcher aber 1992 vorübergehend abgeschafft wurde. Berman (2011) sieht die hauptsächlichen Gründe für diese Aussetzung in der damaligen Fehlerquote und der zeitraubenden Prozedur. Noch ehe das sogenannte „Instant Replay“ in der NFL Ende des Jahrtausends wieder eingeführt wurde, griffen bereits Offizielle in der National Hockey League (NHL, Eishockey), ab 1991, und im Rugby, ab 1996, auf diese Technik zurück. Seitdem wurde der Videobeweis in acht weiteren Spielsportarten eingesetzt (siehe Abbildung 1). Damit stellt er das am weitesten verbreitete technische Hilfsmittel für Schiedsrichter in Spielsportarten dar.

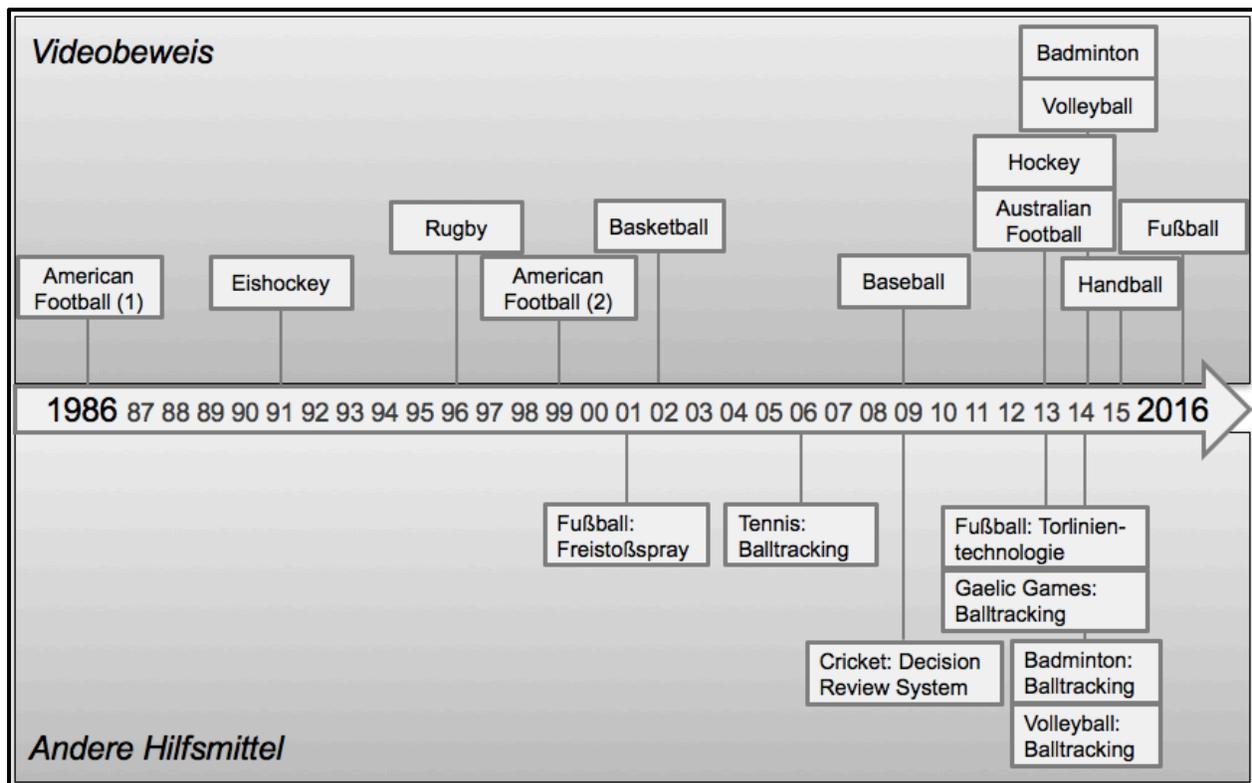


Abb. 1: Chronologische Darstellung der Einführung innovativer technischer Hilfsmittel zur Unterstützung von Schiedsrichtern in Sportsportarten

Immer häufiger im Einsatz sind auch verschiedene Positionserfassungssysteme, welche sich aber bezüglich der hier besprochenen Thematik auf das Tracking des Spielgerätes beschränken. Tennis begann 2006 mit der Nutzung eines derartigen Systems der Firma Hawk Eye Innovations Ltd. zur Überprüfung von Linienentscheidungen. Diesem Beispiel folgten speziell im aktuellen Jahrzehnt (siehe Abbildung 1) ein halbes Dutzend weiterer Sportarten. Cricket nimmt dabei eine Sonderrolle ein, da dort dieses System mit anderen technischen Hilfsmitteln kombiniert wird, um zu einer Entscheidung zu gelangen. Im Rahmen des *Decision Review Systems*, welches 2009 offiziell eingeführt wurde, werden zusätzlich ebenso der Videobeweis sowie Aufnahmen spezifischer Mikrofone (*Snickometer*) und Wärmebildkameras (*Hot Spot*) eingesetzt. Zu den Sportarten, die mehrere Systeme nutzen, wenn auch nicht bezogen auf eine Entscheidung, gehört inzwischen auch Fußball, wo neben der Torlinientechnologie auch das sogenannte Freistoßspray eingesetzt wird, um den erforderlichen Mindestabstand der verteidigenden Mannschaft

bei Freistößen zu markieren. Insgesamt fällt die Häufung der Einführung von technischen Hilfsmitteln in den letzten fünf Jahren auf, in denen sich die Anzahl mehr als verdoppelt hat.

1.2.3 Taxonomie

Die technischen Hilfsmittel werden dabei für eine Vielzahl unterschiedlicher Gründe und Zwecke eingesetzt, natürlich auch bedingt durch die unterschiedlichen Bedürfnisse und Rahmenbedingungen der jeweiligen Sportarten. Unter anderem werden dabei identische Fragen mit unterschiedlichen Hilfsmitteln geklärt. So wird die Frage nach einer potenziellen Torlinienüberschreitung im Eishockey über einen Videobeweis geklärt, während im Fußball dafür Positionserfassungssysteme eingesetzt werden.

Bezüglich einer Taxonomie der Hilfsmittel gibt es in der Literatur zwei Vorschläge, welche jeweils zwei Klassen identifizieren. Dyer (2015) unterteilt, unter der wenig aussagekräftigen Bezeichnung *nicht menschliche Entscheidungsfindung im Sport*, in Videobeweis und Linienentscheidungs-Technik. Die Benennung der zweiten Klasse ist dabei schlicht falsch, was an dem zwar von ihm aufgeführten, aber scheinbar nicht verstandenen, Beispiel Cricket aufgezeigt werden kann. Das dort eingesetzte Positionserfassungssystem, von Dyer (2015) als Linienentscheidungs-Technik aufgeführt, überprüft keine Entscheidungen, die Linien betreffen. Genau an diesem Beispiel lässt sich auch die Problematik der Taxonomie an sich, unabhängig von Dyers (2015) Benennung der Klassen, verdeutlichen. Das Positionserfassungssystem wird in Cricket grundlegend anders eingesetzt als beispielsweise im Fußball oder Tennis. Im Cricket wird das technische Hilfsmittel zur Unterstützung der Entscheidungsfindung herangezogen, während es in den beiden eben genannten Sportarten die Entscheidung quasi selbst trifft.

Letztere Art des Einsatzes wird von Collins (2010) und Collins und Evans (2008, 2012) kritisiert, welche auf die Fehlbarkeit der Positionserfassungssysteme hinweisen. Im Rahmen ihrer Kritik, auf welche im Detail an einer späteren Stelle dieser Arbeit eingegangen werden soll, nehmen sie implizit eine Taxonomie für technische Hilfsmittel vor, die sich aus der Art ihres Beitrages zur Entscheidungsfindung ableitet: entscheidungstreffende Hilfsmittel und entscheidungsunterstützende Hilfsmittel. Diese Einteilung bietet sich aus oben genannten Gründen eher an, ist aber noch nicht

erschöpfend. Das Freistoßsray im Fußball wird weder zur Unterstützung der Entscheidungsfindung, geschweige denn als Quasi-Substituierung des Schiedsrichters eingesetzt. Vielmehr geht es darum, den Schiedsrichter bei der Durchsetzung von Regeln zu unterstützen⁵. Dementsprechend wird für den eigenen Vorschlag einer Taxonomie in drei Klassen unterteilt: Technische Hilfsmittel...

1. ... zur Quasi-Substituierung des Schiedsrichters
2. ... zur Unterstützung der Entscheidungsfindung
3. ... zur Unterstützung der Durchsetzung von Regeln

1.3 Problem- und Zielstellungen

Wie oben erwähnt, liegen dem Einsatz innovativer technischer Hilfsmittel unterschiedliche Rahmenbedingungen und Bedürfnisse zugrunde. Bei allen handelt es sich aber um Innovationen i.S.v. neuen Techniken, welche (positive) Änderungen in einem sozialen System herbeiführen sollen. Dementsprechend plausibel mag es erscheinen, dass deren Einführung von umfassenden Evaluationen bzw. umfassender Evaluationsforschung⁶ begleitet wird, deren Gegenstand die wissenschaftliche Untersuchung von praktischen Interventionen ist. Was die Sportpraxis bzw. die handelnden Verbände betrifft, kann nicht abschließend geklärt werden, ob bzw. in welchem Ausmaß und welcher Form dies der Fall ist. Dies würde eine komplett transparente Kommunikationspolitik der Verbände oder ggf. der jeweiligen (Technik-) Anbieter erfordern, welche schlicht nicht gegeben ist. Dennoch lässt sich das Ausmaß an Evaluationen einiger Verbände über deren Pressemitteilungen, -materialien oder veröffentlichte Richtlinien für Anbieter zumindest grob charakterisieren.

Ein Fokus der Verbände bei der Einführung innovativer Hilfsmittel scheint auf der jeweiligen Funktionsweise der Technik zu liegen, speziell in Bezug auf Genauigkeit und Anwendbarkeit. Öffentlich zugängliche Dokumente des Weltfußballverbandes Fédération Internationale de Football Association (FIFA) berichten ausführlich über die zu absolvierenden Testabläufe für potenzielle Torlinientechnologie-Anbieter sowie die

⁵ Derzeit ist das Freistoßsray die einzige Technik, die hauptsächlich der Durchsetzung einer Regel dient. Allerdings ist ein Einsatz ähnlicher Hilfsmittel in anderen Sportarten gut vorstellbar und wird teilweise bereits diskutiert; so zum Beispiel die Einblendung der sogenannten *Line of Scrimmage* im American Football.

⁶ Eine genaue Klärung der beiden Begriffe sowie der Verwendung in dieser Arbeit folgt unter 2.1.

erforderliche Genauigkeit ($\pm 1,5$ cm) und Latenzzeit (< 1 Sekunde) (FIFA, 2014). Weitere Informationen bezüglich zusätzlicher Maßnahmen zu Evaluationen dieses Hilfsmittels werden nicht aufgeführt. Auch der internationale Tennisverband International Tennis Federation (ITF) veröffentlicht ein Dokument zur Beschreibung der durchgeführten Evaluationen für das angewandte Positionserfassungssystem mit gleichem Fokus (Fehlertoleranz von ± 5 mm). Ergänzend wird untersucht, ob vom System ausgehende Behinderungen oder Sicherheitsrisiken für Stakeholder auftreten können (ITF, 2015). Dabei wird allerdings nicht genau spezifiziert, wie diese Evaluationsprozesse ablaufen sollen.

Ein weiterer Fokus liegt auf der Einsatzhäufigkeit bzw. Beschreibung des Nutzens anhand von korrigierten bzw. beeinflussten Entscheidungen. Steen (2011) greift in seinem Artikel über das Decision Review System Untersuchungen des Weltcricketverbandes International Cricket Council (ICC) auf, welche einen Anstieg korrekter Schiedsrichterentscheidungen von 91,3% auf 97,4% in 13 *Tests*⁷ ergaben. Leider wird nicht näher auf die Methodik der Untersuchung eingegangen, beispielsweise die Operationalisierung von Entscheidungen. Diese Problematik ergibt sich nicht für die von der NFL (2016) veröffentlichten Angaben über die Nutzung des Videobeweises in dieser Liga, da dort schlicht die Häufigkeit des Einsatzes, sowie die Rate an diesen *Challenges*, welche zu einer Änderung der ursprünglichen Entscheidung führten, beschrieben werden. So verlangen beispielsweise im Zeitraum von 1999 bis 2013 die hauptverantwortlichen Trainer im Schnitt etwas weniger als ein Mal pro Spiel einen Videobeweis, was 66% der insgesamt durchgeführten Videobeweise entspricht. Laut Berman (2011) erlangen Trainer durch ihre *Challenges* dabei häufiger eine Abkehr von der ursprünglichen Entscheidung (44%), als dies bei den von der Liga vorgegebenen Videobeweisen der Fall war (25%).

Wie oben beschrieben, ist das Ausmaß der Evaluationen in der Praxis nicht bekannt. Gleiches gilt für den gewählten Ansatz bzw. die Frage, ob es überhaupt ein bewusstes methodisch-strategisches Konzept gibt. Die eingeschränkte Dimensionalität sowie die Unklarheiten bezüglich des methodischen Vorgehens der oben aufgeführten

⁷ *Tests* sind die älteste Wettkampfform im Cricket und kann nicht wörtlich i.S.v. Testspielen übersetzt werden. *Tests* genießen eine besondere Stellung im Cricket und dürfen beispielsweise lediglich von derzeit zehn Verbänden gespielt werden.

Beispiele lässt zumindest den Verdacht aufkommen, dass sich auf einer eher trivialen, wissenschaftlich nicht fundierten Evaluationsebene bewegt wird. Ein Grund hierfür mag in der bisher nahezu nicht stattfindenden Betrachtung dieses Phänomens in den empirischen Wissenschaften liegen. Cross (2014) setzt sich mit der Genauigkeit des entscheidenden Outputs des Positionserfassungssystems im Tennis auseinander. Dafür vergleicht er den vom Anbieter ausgegebenen Ballabdruck auf dem Boden in Form und Größe mit den Befunden aus einer Reihe von Laborexperimenten. Laut Cross' (2014) Vergleichen ist dabei der offiziell genutzte Abdruck tendenziell breiter bei dominant horizontal auftreffenden Bällen, während für eher im vertikalen Winkel einfallende Flugbahnen genau das Gegenteil eintritt. Borooh (2016) untersucht das Kosten-Nutzen-Verhältnis des Decision Review Systems im Cricket, in dem er der Zunahme an korrekten Entscheidungen die Kosten pro Test gegenüberstellt. Aufgrund der von ihm so errechneten Kosten von 54.545 US\$ pro Prozentpunkt Zunahme, in Verbindung mit Simulationen unter Einbezug von Annahmen über menschliche und maschinelle Fehlerwahrscheinlichkeiten, stellt Borooh (2016) die Berechtigung der Investitionen in dieses technische Hilfsmittel in Frage. Weitere sechs empirische Arbeiten nutzen Auswertungen zum Einsatz von technischen Hilfsmitteln als Instrument zum Erkenntnisgewinn für andere Fragestellungen. Das technische Hilfsmittel stellt dabei in allen sechs Studien das Positionserfassungssystem zur Überprüfung von Linienentscheidungen im Tennis dar. Mather (2008) und Carboch, Vejvodova und Suss (2016) nutzen Daten dieser Technik zur Beschreibung von Schiedsrichterfehlern. Abramitzky, Einav, Kolkowitz und Mill (2012), Clarke und Norman (2012) sowie Nadimpalli und Hasenbein (2013) beschäftigen sich mit dem optimalen Einsatz des Hilfsmittels aus Sicht der Spieler. Anbarci, Lee und Ulker (2014) untersuchen ebenfalls Einsatzmuster, jedoch mit dem Fokus auf Unterschieden zwischen den Geschlechtern.

Evaluationen von innovativen technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern in Spielsportarten fanden also bisher auch in der Scientific Community entweder implizit oder ohne Darlegung eines konzeptionellen Hintergrunds statt. An diesen Versäumnissen versucht die vorliegende Dissertation anzugreifen, beginnend mit der Klärung, welche Kriterien erfüllt sein müssen, um in diesem Kontext den an Evaluationsforschung gestellten Ansprüchen gerecht zu werden, und ob es sich hier

überhaupt um einen legitimen Gegenstand für Evaluationsforschung handelt. Darauf aufbauend soll ein Grundgerüst für eine sportartenübergreifende Evaluationsstrategie von technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern erarbeitet werden, u.a. mit einem Schwerpunkt auf Verfahren und Werkzeugen der Wettkampfdiagnostik. Exemplarisch wurden damit Evaluationsforschungen für zwei technische Hilfsmittel im Fußball, die Torlinientechnologie und das Freistoßspray, durchgeführt, wobei jeweils auf die Methode der Systematischen Spielbeobachtung zurückgegriffen wurde. Die gewonnenen Erkenntnisse werden anschließend sowohl bezüglich der jeweiligen Technik als auch bezüglich der Evaluationsstrategie diskutiert. Unter anderem vor dem Hintergrund dieses Erfahrungswissens wird anschließend das erarbeitete evaluationsstrategische Grundgerüst evaluationstheoretisch eingebettet und werden Konsequenzen für die Evaluationspraxis gezogen.

1.4 Verortung der Einzelbeiträge

Bei der vorliegenden Arbeit handelt es sich um eine kumulative Dissertation, wobei drei Full-Paper (peer-reviewed) aus englischsprachigen, international verbreiteten Publikationsorganen eingebunden werden. Inhaltlich trugen diese drei akzeptierten Erstautorenveröffentlichungen an folgenden Stellen dieser Dissertation bei: *Scientific Approaches to Technological Officiating Aids* (Kolbinger & Lames, 2016) extrahiert anhand einer Literaturanalyse, unter Einbezug philosophischer und rechtswissenschaftlicher Beiträge, Problemfelder, welche bereits von der Scientific Community identifiziert wurden und unterstützt dabei primär die Entwicklung des Grundgerüsts zu eben jener sportartenübergreifenden Evaluationsstrategie. In *Do we need Goal Line Technology in Soccer or could Video Proof be a more suitable Choice*⁸ (Kolbinger, Linke, Link & Lames, 2015) und *The use of vanishing spray reduces the extent of rule violations in soccer* (Kolbinger & Link, 2016) wurde die internationale Fachöffentlichkeit erstmals für die Thematik der Evaluation innovativer technischer Hilfsmittel zur Unterstützung für Schiedsrichter in Spielsportarten anhand praktischer Beispiele sensibilisiert. Diese Publikationen unterstützen die Dissertation als praktische

⁸ Der Artikel wurde unter diesem Titel beim Publikationsorgan eingereicht, welches nicht nur Groß- und Kleinschreibung veränderte, sondern zudem den Untertitel zum Titel hinzufügte. All dies geschah ohne Wissen des Promovenden.

Anwendungen des evaluationsstrategischen Grundgerüsts in zweierlei Hinsicht. So wird gezeigt, wie wettkampfdiagnostische Evaluationsforschung in diesem Kontext umgesetzt werden kann und welche Bedeutung ihr zukommt. Reziprok lieferten die Anwendungen wiederum Hinweise zur Weiterentwicklung des Grundgerüsts.

2 Methoden

Die oben aufgeführten Zielstellungen der Arbeit werfen eine Reihe von Fragen auf, die es zu klären gilt, bevor das methodische Vorgehen im engeren Sinne betrachtet werden kann. Die im Titel vorgenommene Referenz der Evaluationsforschung bedarf zuerst einer Spezifizierung, was in dieser Arbeit unter Evaluationsforschung zu verstehen ist. Dass dies nicht trivial ist, wird unter 2.1. aufgezeigt, wo auch der Nachweis der Eignung der Evaluationsforschung für dieses Phänomen (und umgekehrt) geführt wird. Es folgt die Entwicklung eines Grundgerüsts für spezifische Evaluationsstrategien, chronologisch orientiert am Prozess der Einführung einer Innovation (2.2.). In diesem Grundgerüst soll ein Verfahren der Wettkampfdiagnostik, die *Systematische Spielbeobachtung*, eine tragende Rolle spielen, welches als besonders geeignet erachtet wird. Dies erfordert eine kontextbezogene Diskussion der Eignung dieser Methoden, welche bislang ausschließlich im trainingswissenschaftlichen Kontext angewendet wurden. Der zweite Unterpunkt des Methodenkapitels endet mit der Verortung der verfassten Publikationen im evaluationsstrategischen Grundgerüst. Abschließend werden unter 2.3. weiterführende Informationen über die in den einzelnen Studien verwendeten Methoden gegeben.

2.1 Innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in den Sportsportarten und Evaluationsforschung

Die Begriffe Evaluation und Evaluationsforschung werden beide jeweils mit unterschiedlichen Bedeutungen verwendet, was unter anderem dazu führt, dass beide Begriffe auch synonym gebraucht werden. Suchman führte 1967 den Begriff *evaluative Forschung* ein, um diese von der Evaluation als allgemeinem Prozess der Bewertung und Beurteilung eines Wertes abzugrenzen. Fasst man Suchmans Abgrenzung in einer Definition für evaluative Forschung zusammen, welche sich aus den jeweils ersten Begriffsbestimmungen in seiner Arbeit von 1967 zusammensetzt, würde diese lauten:

*Evaluative Forschung ist die Verwendung wissenschaftlicher Methoden und Techniken mit der Absicht, den Wert von etwas zu beurteilen.*⁹

⁹ Bettet man diese Definition in Suchmans (1967) weitere Ausführungen ein, muss diese Zusammenfassung in mindestens zwei Punkten konkretisiert werden. Erstens handelt es sich bei dem

Dies führte aber keineswegs dazu, dass die Begriffe einerseits eindeutig abgegrenzt und andererseits einheitlich verwendet werden. So definiert unter anderem die DeGEval (Gesellschaft für Evaluation e. V.) den Begriff Evaluation so, dass er im Prinzip den Ansprüchen evaluativer Forschung gerecht wird. Hinzu kommt die Einführung des Begriffs Programmevaluation (u.a. Patton, 1997; Stufflebeam & Shinkfield, 2007) welcher ebenfalls austauschbar zu evaluativer Forschung (und teilweise auch Evaluation) verwendet wird. Der Begriff Evaluationsforschung kann schlicht als eine stilistische Änderung der ursprünglichen Bezeichnung „evaluative Forschung“ gesehen werden. Diese Änderung in Evaluationsforschung hat sich so einheitlich durchgesetzt, dass im Titel von Suchmans entsprechendem Werk von 1967, „Evaluative research: Principles and practice in public service and social action programs“, das *evaluative* fälschlicherweise in vereinzelt Quellenangaben bereits durch *evaluation* ersetzt wird.

Die Suchman'sche Unterscheidung wird als Ausgangspunkt der Evaluationsforschung (Lames, 2000) bzw. Evaluationstheorie (Shadish, Cook & Leviton, 1991) genutzt bzw. gesehen. Letztere Meinung teilten auch Alkin und Christie (2004) in ihrer historisch-konzeptionellen Einordnung führender Evaluationsforscher (bzw. solche, die von diesen beiden Autoren als führend identifiziert wurden), welche anhand eines „Evaluations-Baums“ veranschaulicht wird. In ihrer Neuauflage (Christie und Alkin, 2013) verleitet die beiden die Tatsache, dass Suchmans Werk (1967) speziell darin bestand, die Arbeit zu experimentellen und quasi-experimentellen Forschungsmethoden von Campbell und Stanley (1963) in einen Kontext zu Evaluation zu stellen, zu einem durchaus radikalen Schritt. Suchman wird nicht nur die Vorreiterrolle abgesprochen, sondern er wird zudem aus der Veranschaulichung der führenden Evaluationstheoretiker entfernt. Diese Einschätzung wird hier nicht geteilt, gibt jedoch einen guten Einstieg, um exemplarisch die Kontroversen, welche Versuche eines historischer Abrisses und/oder einer Klassifikation bestehender Beiträge zur Evaluationstheorie mit sich bringen, aufzuzeigen. Der Evaluations-Baum von Alkin und

Gegenstand, welcher evaluiert werden soll, um soziale Programme und zweitens beschränkt sich Suchman bezüglich des wissenschaftlichen Vorgehens auf quantitative Methoden.

Christie (2013) teilt Evaluationstheoretiker¹⁰ gemäß ihres Hauptbeitrages für die Evaluationstheorie in drei Äste ein: *Methods*, *Valuing* und *Use*. Diese Einteilung lässt sich aus drei Richtungen angreifen, nämlich bezüglich der Auswahl der zugeteilten Autoren, deren Zuteilung und bezüglich der Sinnhaftigkeit dieser Taxonomie. Das Diskussionspotenzial betreffend der Auswahl wurde bereits oben am Beispiel Suchman aufgezeigt. Die Zuteilung der Autoren betrachten selbst Alkin und Christie (2013) angenehm selbstkritisch und teilen in ihrer Neuauflage auch die Einschätzungen der betroffenen Autoren mit, welche sich durchaus an anderen Positionen in dieser Klassifikation sehen. Alkin und Christie (2013) sind sich dabei auch bewusst, dass ihre Darstellung maximal eine drastisch vereinfachte Darstellung eines hochkomplexen Geflechts darstellt. Insgesamt betrachtet stellt sich die Frage, ob eine derartig strikte Zuteilung von Evaluationstheorien sinnvoll ist oder sich eher eine Herangehensweise wie von Cook und Shadish (1986) anbietet, welche ohne den Hintergedanken einer Klassifikation Beiträge zur Evaluationstheorie aus ausgewählten, wesentlichen Blickwinkeln betrachten. Auf diese Einteilung wird zu einem späteren Punkt detaillierter eingegangen, wenn anhand dieser die spezifische eigene Evaluationsstrategie diskutiert und eingeordnet werden soll.

Die Notwendigkeit einer solchen Systematisierung ergibt sich aus der Tatsache, dass eine Vielzahl theoretischer Beiträge vorliegt, womit eine Vielzahl an Definition einhergeht. Als diejenige Definition, die sich in den Sozialwissenschaften durchgesetzt hat, nennt Lames (2000) die auf Suchman (1967) basierende Definition von Rossi und Freeman (1993), welche er folgendermaßen ins Deutsche übersetzt:

„Evaluationsforschung ist die systematische Anwendung der Vorgehensweise der Sozialforschung für die Ermittlung der Konzeptionierung, der Gestaltung, der Einrichtung und der Nutzeneffekte sozialer Interventionsprogramme.“(Lames, 2000, S.8 nach Rossi und Freeman, 1993, S.5)

Bevor die Eignung dieser Definition für den hier vorliegenden Kontext besprochen wird, müssen zwei Vorbemerkungen gemacht werden. In der inzwischen siebten

¹⁰ Alkin (2013) zweifelt zudem an, dass es sich um Theorien im eigentlichen Sinne handle. Er führt die Beiträge zwar anschließend unter dieser Bezeichnung, sieht sie aber lediglich als richtungsgebende Modelle.

Auflage ihres Werkes liegt eine weiter konkretisierte Definition vor (Rossi, Freeman & Lipsey, 2004), welche einen Schwerpunkt auf das Setting und den sozialen Nutzen von Evaluationsforschung legt¹¹. Zudem wird der Begriff Evaluationsforschung explizit austauschbar mit den Begriffen Evaluation und Programmevaluation verwendet. Bezüglich der Eignung für den Kontext in den Sportspielen kann als Hindernis angesehen werden, dass sich diese Definition implizit auf solche sozialen Interventionen bezieht, mit denen schwerwiegende gesellschaftliche Probleme bekämpft werden sollen (z.B. Defizite im Gesundheitssystem). Im wortwörtlichen Sinn betrachtet, handelt es sich bei der Einführung technischer Hilfsmittel zur Unterstützung von Schiedsrichtern um soziale Interventionen, jedoch erscheint die Nutzung dieses Begriffs, mit dem Bewusstsein, von welchen sozialen Interventionen Rossi et al. (2004) sprechen, unangebracht zu sein. In der Übertragung der Definition für den Kontext der Trainingswissenschaft benutzen Hohmann et al. (2014) schlicht den Begriff *Interventionen*¹². Der Beitrag von Hohmann et al. (2014) zeigt darüber hinaus auf, welchen wichtigen Beitrag Evaluationsforschung als Forschungsstrategie im Kontext des Leistungssports¹³ leisten kann. Von ihrem trainingswissenschaftlichen Standpunkt ausgehend beschränken sie sich dabei selbstverständlich auf die Bedeutung für den Trainingsprozess von Athleten. Ganz grob zusammengefasst behandelt Evaluationsforschung also dort die Frage, wie sich Interventionen auf die Leistung von Athleten auswirken. Es liegen evaluative, wissenschaftliche Vorerfahrungen über Interventionen im Leistungssport vor, die sich für die Fragestellung dieser Arbeit übertragen lassen. Es lässt sich also zusammenfassend sowohl festhalten, dass innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Sportarten einen legitimen Forschungsgegenstand von Evaluationsforschung darstellen als auch dass sich diese Forschungsstrategie durchaus anbietet.

¹¹ "Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions." (Rossi et al., 2004, S. 24)

¹² „Evaluationsforschung ist die systematische Anwendung wissenschaftlicher Methoden zur Bewertung des Konzeptes, der Implementation, der Wirksamkeit und der Effektivität einer Intervention.“ (Hohmann et al., 2014, S. 35 f.)

¹³ Hohmann et al. (2014) beziehen hier ihre Aussagen nicht nur auf den Leistungssport, sondern beispielsweise auch auf den Schul- und Gesundheitssport. Im Kontext dieser Arbeit sind diese Formen des Sporttreibens allerdings nicht relevant.

Das evaluationsstrategische Grundgerüst wird in dieser Arbeit ausgehend von der Begriffsbestimmung von Suchman (1967) erarbeitet, welche als Ausgangspunkt für die nachfolgende Definition gesehen werden kann. Betrachtet man gängige Definitionen, wie beispielsweise die oben genannten von Rossi und Freeman (1993) bzw. Rossi et al. (2004), deren Paraphrasierungen bzw. Übertragungen durch Lames (2000) und Hohmann et al. (2014) oder andere Definitionen (u.a. Stufflebeam & Shinkfield, 2007), so sind jeweils drei grobe Bausteine beinhaltet.

- Wie wird evaluiert?
- Was wird evaluiert?
- Wozu wird evaluiert?

Dabei agieren die Definitionen auf unterschiedlichen Abstraktionsniveaus, wobei durch die Präzisierungen in der Regel wissenschaftstheoretische Ausgangspositionen oder Fachzugehörigkeiten ausgedrückt werden. In manchen Definitionen, beispielsweise bei Hohmann et al. (2014), ist der dritte Baustein impliziter Natur, sofern man nicht unterstellen möchte, dass nur zum Selbstzweck der Bewertung bewertet werden soll. Im Hinblick auf die Einbettung dieser Definition bzw. der Evaluationsforschung, als Forschungsstrategie der Trainingswissenschaft, kann hier aber angenommen werden, dass es sich eher um ein hohes Abstraktionsniveau handelt. Auf einem sehr hohen Abstraktionsniveau agiert eben auch die Suchman'sche Unterscheidung.

Laut dieser wird Evaluationsforschung mit der Absicht durchgeführt, den Wert von etwas zu bestimmen. Auf den ersten Blick kann dies im Hinblick auf andere Definitionen als unvollständig wahrgenommen werden, weswegen an dieser Stelle aufgezeigt werden soll, warum dies nicht der Fall ist, sondern dies schlicht eine sehr hohe Abstraktionsebene darstellt. Dies wird exemplarisch anhand der sehr ausführlichen Definition von Stufflebeam und Shinkfield (2007) aufgezeigt, welche an der entsprechende Stelle neben Wert (worth) von Leistung i.S.v. Qualität (merit), Rechtschaffenheit (probity), Sicherheit (safety), Bedeutung (significance) und Gerechtigkeit (equity) spricht. Diese weiteren Begriffe werden jedoch alle vom Begriff Wert mit abgebildet bzw. stellen Vorraussetzungen zur Bestimmung des Wertes dar. So

kann ein Evaluand¹⁴ an sich qualitativ hochwertig sein, jedoch im Kontext der Intervention den gewünschten Zweck nicht erfüllen, was somit zugleich den aufgeführten Begriff der Bedeutung miteinordnet. Als Beispiel führen Mertens und Wilson (2012) die hochtalentiertere Kardiologin auf, welche keinen Wert für ein Krankenhaus hat, welches Unterstützung in der Anästhesiologie benötigt. Die Abbildung der Begriffe Rechtschaffenheit, Sicherheit und Gerechtigkeit ergeben sich moralisch intuitiv, wenn man sich die Konsequenzen für den Wert vor Augen führt, falls diese nicht gegeben sind. Reduziert man sich also in seinem definitorischen Ausgangspunkt auf den Begriff Wert, ist dies sehr abstrakt, aber keineswegs unvollständig. Als Ausgangspunkt für die evaluationstheoretische Betrachtung eines diesbezüglich noch jungfräulichen Feldes, eignet sich eine eher abstrakt gehaltene Definition, wie es die zugrunde gelegte Suchman'sche ist, sogar besonders. Abschließend sei noch angemerkt, dass auch in dieser Arbeit, sofern nicht explizit gekennzeichnet, die Begriffe Evaluation, Evaluationsforschung und Programmevaluation austauschbar verwendet werden. Diese Maßnahme ist in den sich dadurch stilistischen Vorteilen und der Tradition der synonymen Verwendung der Begriffe in der einschlägigen Scientific Community (siehe oben) begründet.

2.2 Evaluationsstrategien für innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Sportsportarten

Die in der Überschrift gewählte Bezeichnung *Evaluationsstrategien* wurde bewusst gewählt, ebenso wie die Formulierung der Zielstellung bzgl. der Entwicklung eines *Grundgerüsts für eine spezifische Evaluationsstrategie*. Zugrunde liegt die Sichtweise von Cronbach und Shapiro (1982), dass eine Evaluationsstrategie (bzw. in ihrer Wortwahl: Design) immer anhand der Anforderungen bzw. sich bietenden Möglichkeiten des jeweiligen Settings ausgewählt werden muss. Dementsprechend kann es nicht eine einzige Evaluationsstrategie für sämtliche technischen Hilfsmittel geben, weswegen der Anspruch nur die Entwicklung eines Grundgerüsts sein kann. Dabei ist das Abstraktionsniveau bewusst so gewählt, dass eine Gültigkeit für alle entsprechenden Arten von Hilfsmitteln und Settings beansprucht werden kann. In

¹⁴ Als *Evaluand* wird in der Evaluationsforschung das bezeichnet, was im Zuge einer Evaluation bzw. im Zuge evaluativer Forschung evaluiert wird.

diesem Abschnitt wird zuerst der erarbeitete Vorschlag für ein chronologisches Grundgerüst zur Evaluation vorgestellt, die Beitragsmöglichkeit der Wettkampfdiagnostik diskutiert und anschließend die einzelnen Publikationen in dem Grundgerüst verortet.

2.2.1 Spezifisches Grundgerüst für Evaluationsstrategien

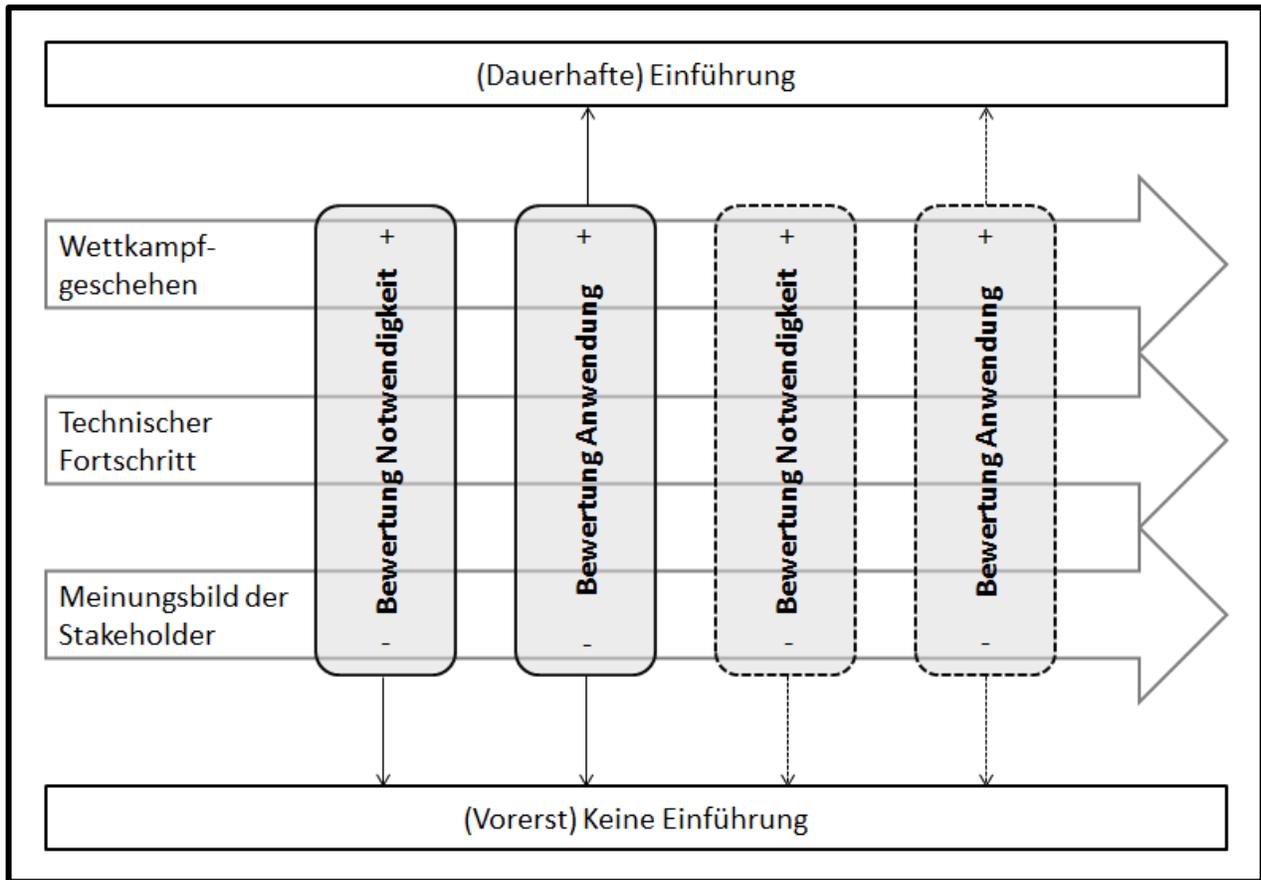


Abb. 2: Skizze des idealtypischen Grundgerüsts für Evaluationsstrategien zur Unterstützung von Schiedsrichtern in Spielsportarten

Der Modellierungsvorschlag in Abbildung 2 zeigt das chronologisch konzeptionelle Vorgehen bei der Einführung von innovativen technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern in Spielsportarten. Idealtypisch wird davon ausgegangen, dass, bevor das Hilfsmittel angewendet wird, die Notwendigkeit überprüft wird. Zudem geht das Modell davon aus, dass sich die zugrundeliegenden Faktoren verändern und dementsprechend die beiden angesprochenen Phasen, Bewertung der Notwendigkeit und Bewertung der Anwendung, mehrfach durchlaufen werden. Die Faktoren wurden für die Modellierung in drei thematische Ströme eingeteilt: das Wettkampfgeschehen, das Meinungsbild der Stakeholder und der technische Fortschritt.

Alle diese Ströme sind zur letztendlichen Bestimmung des Wertes eines technischen Hilfsmittels zu beachten, wobei das Wettkampfgeschehen die Sonderrolle einnimmt, dass es den Ursprung der Einführungsintention beinhaltet.

Zur Überprüfung der Notwendigkeit der Einführung lässt sich die Vorgehensweise folgendermaßen zusammenfassen und anschließend konkretisieren. Im Wettkampfgeschehen wird eine fehlerhafte Anwendung des Regelwerks identifiziert oder ggf. scheinbar identifiziert. Dies impliziert bereits die Meinung mindestens eines Stakeholders, wenn der- oder diejenigen Stakeholder es für relevant genug erachtet, die anderen Stakeholder über diese Tatsache zu informieren. Darüber hinaus existiert ein Werkzeug, ein Verfahren oder eine Technologie, welche/s die Aussicht bietet, die (scheinbar) vorliegende Diskrepanz zwischen den Regeln und deren Anwendung schließen oder mindestens verringern zu können. Im Sinne einer ganzheitlichen Bewertung muss der Evaluator nun Informationen aus allen drei Strömen sammeln, bereitstellen und i.S.v. Beratung bewerten.

Bezüglich des Wettkampfgeschehens ist konkret zu überprüfen, ob und in welchem Maße die vermeintlich fehlerhafte Regelanwendung vorliegt. Dies bringt direkt die Frage des heranzuziehenden Goldstandards mit sich, an der die Anwendung der Regeln gemessen werden soll. Dieser könnte beispielsweise ein Positionserfassungssystem, eine hochauflösende Hochfrequenzkamera oder ein Pool aus Experten sein. Dies wiederum ist verbunden mit der Problematik, ob der gewählte Goldstandard diesem Anspruch wirklich genügt bzw. ob es überhaupt einen Goldstandard geben kann. Letztere Frage stellt sich speziell dann, wenn das Regelwerk dem Schiedsrichter Interpretationsspielräume ermöglicht. Unabhängig davon kann zusätzlich das Wettkampfgeschehen gegebenenfalls bereits auf die Eignung der Anwendung eines technischen Hilfsmittels untersucht werden. Dies beinhaltet auch die Abschätzung weiterer Effekte auf das Spielgeschehen durch dessen Einführung, welche positiver als auch negativer Natur sein können.

Die Überprüfung des technischen Hilfsmittels an sich muss sich ebenfalls mit der Frage des geeigneten Goldstandards beschäftigen. Um einen Nutzen zu generieren, sollte der Einsatz des Hilfsmittels eine geringere Abweichung vom Goldstandard ermöglichen, als es ohne Einsatz der Fall wäre. Hinzu können je nach Technik und

Setting andere Parameter kommen, beispielsweise die Latenzzeit oder Sicherheitsaspekte. Relevant für alle technischen Hilfsmittel ist die Ermittlung bzw. Abschätzung von materiellen Kosten, welche durch die Einführung bzw. einen dauerhaften Einsatz hervorgerufen werden. Gleiches gilt für die Überprüfung der Anwendbarkeit.

Für die Erfassung des Meinungsbildes der Stakeholder bietet sich an, diese zeitlich an den oben beschriebenen Prozessen zu orientieren¹⁵. So ist es sinnvoll, ein erstes Meinungsbild bereits vor diesen Prozessen zu erstellen, um die Methodik so anzupassen, dass Informationen, bezogen auf sämtliche Erwartungen oder Befürchtungen, generiert werden können. Wurden die relevanten Informationen der anderen beiden Ströme erfasst, können sie den Stakeholdern zur Bewertung mitgeteilt werden, welche wiederum systematisch erfasst wird. Hierbei sei ausdrücklich darauf hingewiesen, dass der Evaluator durchaus als Lehrer im Cronbach'schen (1982) Sinn zu einer Bewertung kommen und diese vor den Stakeholdern vertreten darf. Der ganze Prozess kann hermeneutisch iteriert werden, bis eine Entscheidung zur Notwendigkeit getroffen werden kann. Idealtypisch würde diese Entscheidung von einem repräsentativem Panel (vgl. Stufflebeam, 2001) getroffen, sobald bestmöglicher Konsens zwischen den Stakeholdern besteht. Als realistischer wird jedoch angenommen¹⁶, dass diese Entscheidung von den jeweiligen Entscheidungsträgern eines Wettbewerbs, unter Berücksichtigung des Meinungsbildes der Stakeholder, getroffen wird.

Wie oben bereits erwähnt, ist eine negative Entscheidung i.S.v. einer Ablehnung der Einführung nicht (zwingend) endgültig. Änderungen in den einzelnen Strömen können zu einer Neubewertung der Situation führen und einer positiven Entscheidung in einem erneuten Durchlauf der Notwendigkeits-Bewertung. Vergleichsweise unkomplizierter ist der Verlauf, wenn die Einführung eines technischen Hilfsmittels als

¹⁵ Das Vorgehen ist ähnlich dem des Evaluationskonzepts der Fourth Generation Evaluation von Guba und Lincoln (1989). Es unterscheidet sich aber nicht nur im chronologischen Vorgehen, sondern vor allem bezüglich der Bedeutung der den Stakeholdern zur Verfügung gestellten Informationen und dem Abschluss des Evaluationsprozesses.

¹⁶ An dieser Stelle wird die erwartete Abweichung von der Idealvorstellung explizit angegeben, da es sich aus Sicht des Autors um die am unwahrscheinlichsten zu erfüllende Vorstellung handelt. Die anderen Idealtypen sind mehr eine Frage des Aufwands, während dieser spezielle Typ zusätzlich umfassende Änderungen im Hierarchie- und Machtgefüge von Sportverbänden/-Ligen erfordern würde.

notwendig erachtet wird. In diesem Fall ist der nächste Schritt die Überprüfung der Anwendung des technischen Hilfsmittels im jeweiligen realen Setting. In einer ersten Phase kann dies in simulierten Wettkämpfen oder Wettkämpfen, welche aus Sicht der Stakeholder nur eine geringe Bedeutung besitzen, durchgeführt werden. Speziell letztere Möglichkeit wird von Verbänden gerne genutzt, so wurde im Tennis die Hawk-Eye Technologie erstmalig auf der ATP Champions Tour eingesetzt, einer Turnierserie, in der ausschließlich ehemalige Tennis-Profis aktiv sind. Ein anderes Beispiel ist die Major League Baseball, welche eine Saison der Arizona Fall League, eine Liga zur Weiterentwicklung von Nachwuchsspielern zwischen den Saisons der MLB, für einen Testlauf des neuen Videobeweissystems nutzte. Der Vorteil eines solchen Vorgehens ist, dass sich Anwendungsprobleme oder negative Auswirkungen erkennen oder sogar abstellen lassen, bevor das entsprechende Hilfsmittel in Wettkämpfen höherer Bedeutung eingesetzt wird. Jedoch muss man sich der Tatsache bewusst sein, dass sich das Wettkampfgeschehen zwischen diesen Wettbewerben mit hoher Wahrscheinlichkeit unterscheiden wird. Dementsprechend muss die Anwendung zusätzlich auch im realen Setting bewertet werden, sprich im dafür vorgesehenen Wettbewerb.

Die Evaluation geschieht nun wieder unter Berücksichtigung aller drei in Abbildung 2 dargestellten Ströme und ist inhaltlich naturgemäß dem Prozess der Notwendigkeitsbewertung ähnlich. Bezüglich des Wettkampfgeschehens muss untersucht werden, wie sich dieses durch den Einsatz des technischen Hilfsmittels verändert. Zum einen im Hinblick auf das zu lösende Problem, also ob und in welchem Ausmaß sich die Anwendung der Regeln in der spezifischen Situation, bzw. den spezifischen Situationen, verbessert hat. Zum anderen muss das Wettkampfgeschehen aber auch auf weitere Auswirkungen untersucht werden. Bezüglich des eingesetzten technischen Hilfsmittels an sich liegt der Fokus auf der Funktionsweise und den tatsächlich entstandenen Kosten. Auf Basis der gewonnenen Erkenntnisse kann nun ein neues Meinungsbild der Stakeholder abgefragt und übermittelt werden, welches zur Entscheidung führt, ob das technische Hilfsmittel eingeführt wird. Auch an dieser Stelle gilt, dass die Entscheidung keinen endgültigen Charakter besitzen sollte, unabhängig von der Art der Entscheidung. Die zugrundeliegenden Faktoren sind zeitlich variabel und

können damit zu jedem anderen Zeitpunkt zu einer unterschiedlichen Wertbestimmung führen.

2.2.2 Beitrag der Wettkampfdiagnostik am Beispiel der Systematischen Spielbeobachtung

Eine Konsequenz dieses Ansatzes, dass es aus Sicht des Autors eine Vielzahl unterschiedlicher Faktoren zu berücksichtigen bzw. zu bewerten gilt, ist die Notwendigkeit eines breiten Spektrums an Methoden, mit unter anderem Verfahren und Werkzeugen der Demoskopie, der Psychologie oder der Wettkampfdiagnostik. Der wissenschaftlichen Verwurzelung, als wissenschaftlicher Mitarbeiter bzw. Inhaber des Lehrstuhls für Trainingswissenschaft und Sportinformatik, des Doktoranden sowie dessen Betreuer geschuldet, lag der Fokus auf der Anwendung wettkampfdiagnostischer Methoden. Präziser ausgedrückt, da diese Methodologie wiederum eine Reihe von Verfahren und Werkzeugen beinhaltet, wurde speziell die Methode der Systematischen Spielbeobachtung (u.a. Lames, 1994) eingesetzt. Diese Methode eignet sich besonders zur systematischen und objektiven Erfassung von Verhaltensweisen im Wettkampf. Dies gilt nicht nur für die Trainingswissenschaft, die Wissenschaft, in der dieses Verfahren vor allem entwickelt wurde (Hohmann et al., 2014), sondern auch für die hier vorliegende Problematik. Sowohl die Anwendung von Regeln als auch die damit zusammenhängenden Verhaltensweisen bzw. Konsequenzen für das Spielgeschehen lassen sich über entsprechende Beobachtungssysteme abbilden. Dementsprechend wurde diese Methode in den Studien „Torlinientechnologie“ und „Freistoßsray“ angewendet.

Wie bereits im Kapitel zur Evaluationsforschung (2.1) beschrieben, ergibt sich auch bei der Nutzung der systematischen Spielbeobachtung ein Vorteil aus den Vorerfahrungen der Anwendung im Setting des Leistungssports. Allerdings können dort vorliegende Modelle und Begrifflichkeiten nicht uneingeschränkt für das hier vorliegende Setting übernommen werden, was kurz anhand der Qualitativen Spielbeobachtung (Hansen & Lames, 2001) aufgezeigt werden soll. Diese weist von den verwendeten Methoden und auch der evaluationstheoretischen Herangehensweise viele Gemeinsamkeiten mit dem hier gewählten Ansatz auf, unterscheidet sich aber grundlegend in der Intention des Einsatzes. Dort steht die Unterstützung des

Betreuungsprozesses im Sinne der Lames'schen Kopplung von Training und Wettkampf (1994) im Vordergrund und bildet dementsprechend die Basis für methodische und konzeptionelle Entscheidungen. Dies ist bei der Evaluierung innovativer technischer Hilfsmittel zur Unterstützung von Schiedsrichtern nicht der Fall. Die Abkehr von etablierten sportwissenschaftlichen Begrifflichkeiten, wie beispielsweise „Qualitative Spielbeobachtung“, ist also schlicht dieser Tatsache geschuldet und nicht einer ablehnenden Position diesen gegenüber.

2.2.3 Verortung der einzelnen Publikationen im evaluationsstrategischen Grundgerüst

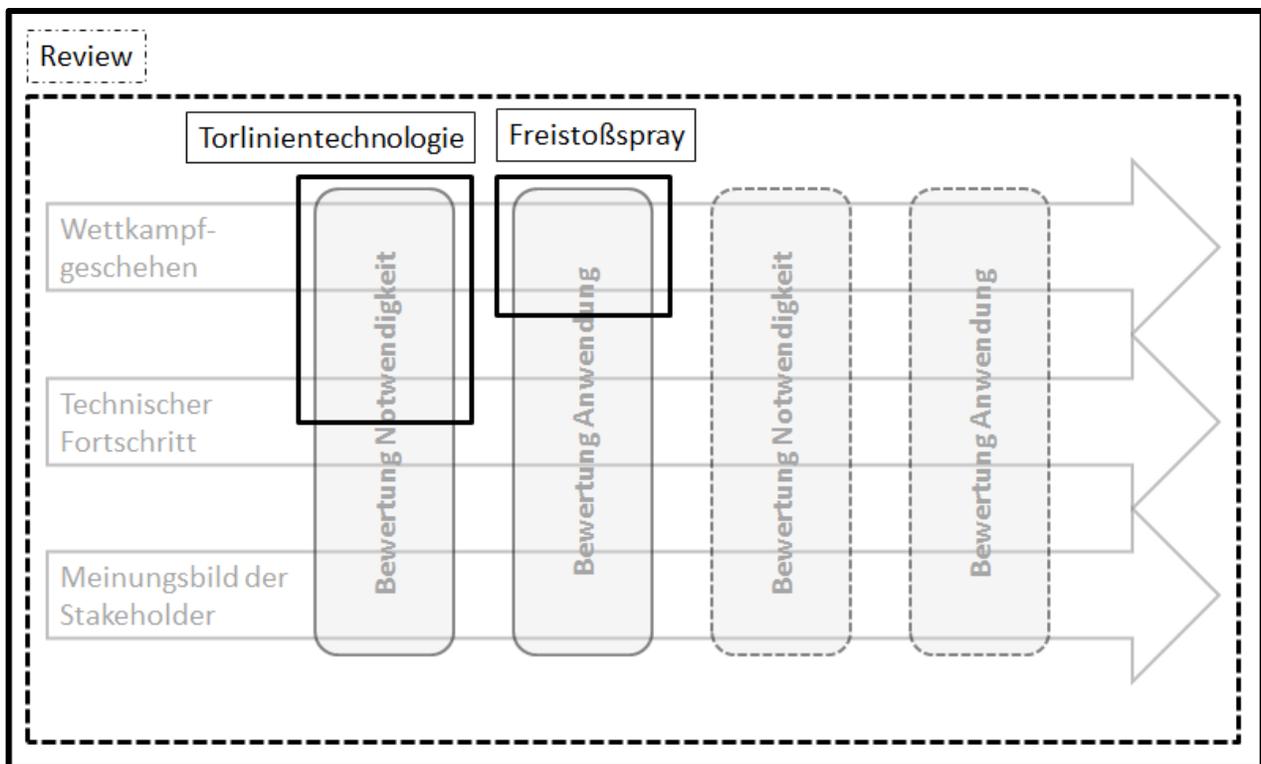


Abb. 3: Grafische Darstellung der Verortung der Einzelbeiträge

Anhand von Abbildung 3 wird bildhaft dargestellt, wie die drei Publikationen dieser Dissertation im entwickelten Grundgerüst einzubetten sind bzw. in welchem Zusammenhang sie damit stehen. Wie im abschließenden Teil der Einleitung beschrieben, war die Intention des systematischen Reviews bezüglich des Grundgerüsts die Unterstützung von dessen Entwicklung. Dementsprechend umrahmt es in der

bildhaften Darstellung dieses Grundgerüst, wobei anhand der gestrichelten Linien dargestellt werden soll, dass es dieses aber nicht explizit abbildet.

Die „Studie Torlinientechnologie“ entstand aus einem Projekt zur Überprüfung der Notwendigkeit von Torlinientechnologie in den Fußball-Bundesligen in Kooperation mit der Deutschen Fußball Liga GmbH. Diese Überprüfung wurde anhand der Frequenz der entsprechenden Szenen im Wettkampfgeschehen durchgeführt. Zusätzlich wurde vor dem Hintergrund der Kosten dieses Hilfsmittels überprüft, ob und in welchem Maße dieses einem Videobeweis überlegen wäre. Die Studie hatte also neben dem Wettkampfgeschehen auch explizite Berührungspunkte mit Aspekten, die dem Strom „Technischer Fortschritt“ zuzuordnen wären, was durch den partiellen Einschluss in der entsprechenden Illustration abgebildet werden soll.

Exklusiv den Strom des Wettkampfgeschehens bei der Bewertung der Anwendung einschließend ist die „Studie Freistoßspray“ im Grundgerüst bildhaft verortet. Diese Evaluation wurde zu einem Zeitpunkt durchgeführt, zu dem das Spray bereits über mehrere Monate im Wettkampf eingesetzt wurde und dementsprechend mit der Hauptintention, die Auswirkungen auf das Wettkampfgeschehen zu überprüfen.

2.3 Weiterführende Anmerkungen zu der Methodik der einzelnen Studien

Die jeweils verwendete Methodik wurde naturgemäß in den entsprechenden Artikeln ausführlich erläutert. Dementsprechend liegt die Intention hier nicht darin, die jeweiligen Abschnitte der betroffenen Publikationen zu wiederholen. Vielmehr werden die Hintergründe über getroffene Entscheidungen offengelegt, methodische Besonderheiten hervorgehoben und speziell die Beobachtungssysteme in Bezug auf deren Anforderungen diskutiert. Die bezüglich Letzterem zwischen den beiden betroffenen Studien vorliegenden Überschneidungen werden vorneweg aufgeführt, ehe die weiteren Anmerkungen zu den verwendeten Methoden gesondert für jede der einzelnen Studien dargestellt werden. Die Artikel wurden chronologisch nach dem Erscheinungsdatum sortiert, als Kapitelüberschriften dienen – soweit vorhanden – jeweils die sogenannten *Running Heads*.

Wie erwähnt, bedienen sich die Studien zur Torlinientechnologie und zum Freistoßspray der Methode der systematischen Spielbeobachtung, wobei es sich bei beiden Beobachtungssystemen um sogenannte Zeichensysteme handelt. Es werden

einzelne Beobachtungseinheiten erfasst, jedoch ohne den Anspruch, aus diesen den kompletten zugrundeliegenden Wettkampf zu rekonstruieren. Dies stellt keine Schwäche der genutzten Beobachtungssysteme dar, sondern ergibt sich bzgl. der bearbeiteten Fragestellungen schlicht aus den Gründen der Relevanz und Ökonomie. Einzig in der Studie zum Freistoßsprint lassen sich die relevanten Aspekte des Verhaltenskontinuums innerhalb der jeweiligen Beobachtungseinheiten rekonstruieren, um die entsprechenden Merkmale im Zusammenhang mit dem Verhalten des Schiedsrichters interpretieren zu können.

2.3.1 Do we need Goal Line Technology?

Im Zuge der Erstellung beider Beobachtungssysteme mussten zur Sicherung der Objektivität operationale Definitionen und Messvorschriften formuliert werden, welche hier nochmals gesondert diskutiert werden sollen. Bei dieser Studie betraf dies die Kriterien zur Erfüllung der operationalen Definition von „torlinien-kritischen Entscheidungen“, welche als eher großzügig angesehen werden können. Ein möglicher Kritikpunkt i.S.v. Validität stellt damit dar, ob über das Messinstrument tatsächlich „torlinien-kritische Entscheidungen“ erfasst werden. Allerdings kann dieser Kritik mit einem Verweis auf die genaue Definition dieser Merkmalstufe begegnet werden, welche nur den Anspruch hat, sämtliche Szenen abzubilden, die den Verdacht erregen, dass eine Torlinienentscheidung kritisch i.S.v. zweifelhaft ist. Man kann darüber hinaus argumentieren, dass diese Merkmalstufe nur so definiert sein kann, da die Wahrnehmung als kritische Entscheidung subjektiver Natur ist. Diese Aufweichung des Merkmals sollte dazu beitragen, ein höheres Maß an Objektivität und der - bei systematischen Beobachtungen damit besonders verbundenen - instrumentellen Konsistenz zu gewährleisten.

Zur Erfassung der exklusiv nur durch die Torlinientechnologie aufklärbaren Fälle sowie der Abschätzung der Eignung eines Videobeweises, wurde die Systematische Spielbeobachtung um Expertenurteile i.S.v. subjektiver Eindrucksanalyse ergänzt. Damit wurde jede kritische Entscheidung dahingehend bewertet, ob die ursprüngliche Beurteilung des Schiedsrichters als richtig, falsch oder anhand des Bildmaterials nicht eindeutig aufgelöst werden konnte. Die Verwendung dieser Methode war aufgrund der

politischen Rahmenbedingungen¹⁷ mit einer weiteren Besonderheit verbunden, nämlich dass die Entscheidungen jeweils von allen beteiligten Experten gemeinsam bewertet wurden und dementsprechend keine Objektivitätsüberprüfungen durchgeführt werden konnten.

2.3.2 The use of vanishing spray reduces the extent of rule violations in soccer

Die größte methodische Schwierigkeit dieser Studie lag in der Identifizierung der Freistöße vor der Einführung des Freistoßsprays, welche im Sinne einer Kontrollgruppe mit den gesprayten Freistößen („Treatmentgruppe“) nach der Einführung verglichen werden sollten. Anders ausgedrückt: Es mussten Freistöße gefunden werden, für die mit hoher Wahrscheinlichkeit angenommen werden konnte, dass das Freistoßspray eingesetzt worden wäre, sofern es zur Verfügung gestanden hätte. Dafür wurden zunächst alle Freistöße nach der Einführung auf Muster bzgl. des Sprayeinsatzes untersucht. Die ursprünglich geplante Parallelisierung exklusiv über Ortsvariablen musste verworfen werden, da diesbezüglich keine Konsistenz des Sprayeinsatzes vorlag. Letztendlich wurde der Parallelisierungsprozess in drei Schritten durchgeführt. Zuerst wurden die Freistöße sieben, an den Spielfeldmarkierungen orientierten, Ortsklassen zugeteilt und anschließend anhand der Anzahl der Defensivspieler in der Mauer aufgeteilt, was als Indikator für das von der Defensivmannschaft empfundene Risiko der Freistoßsituation gesehen wurde. Final wurden die Freistöße vor der Einführung dem am jeweils kürzesten entfernten gesprayten Freistoß zugeordnet. Fand sich eine Situation, in der beispielsweise ein gesprayter Freistoß für mehrere der örtlich nächste war, so wurden die Freistoßpaare so erstellt, dass die Summe aller Abstände zwischen den einzelnen Paaren die geringstmögliche Distanz ergab.

Speziell mit dieser Absicht der Sicherung der instrumentellen Konsistenz waren auch besondere Messvorschriften in der „Studie Freistoßspray“ verbunden, welche Reliabilitätseinschränkungen aufgrund des zur Verfügung gestellten Bildmaterials bestmöglich entgegen wirken sollten. Bei diesem handelte es sich um Mitschnitte aus Fernsehübertragungen mit eingeschränkter Bildqualität, welche zudem nicht speziell auf die wissenschaftlichen Bedürfnisse zugeschnitten waren. Daraus ergaben sich

¹⁷ Auf diese wird im Rahmen der evaluationstheoretischen Einbettung im Kapitel „Diskussion“ nochmals ausführlich eingegangen.

Schwierigkeiten bei der Erfassung von Regelverstößen i.S.v. Verletzungen des erforderlichen Mindestabstands seitens der Defensivmannschaft.

Erstens konnte aufgrund der eingeschränkten Bildauflösung und der Kameraführung nur selten der ursprüngliche, vom Schiedsrichter festgelegte, Mindestabstand anhand von Markierungen oder anderen Besonderheiten des Untergrunds vermerkt werden. Somit konnten Verletzungen des Mindestabstands nicht über die Differenz zur ursprünglichen Position erfasst werden. Dementsprechend mussten Abstandsverkürzungen über Bewegungen der Verteidiger, in der Regel kleine Schritte in Richtung Ball, ermittelt werden.

Dies wurde durch eine zweite Eigenschaft der Fernsehmitschnitte erschwert, nämlich durch Wechsel auf andere Kameras oder die Einspielung von Wiederholungen. Die Mauer bzw. relevante Spieler der Defensivmannschaft waren nicht in allen Fällen die gesamte Zeitspanne, von der Positionierung durch den Schiedsrichter bis hin zur Ausführung des Freistoßes, im Bild. Die Erfassung der Regelverstöße i.S.v. Verletzungen des Mindestabstandes musste sich demzufolge auf die Abstandsverkürzungen beschränken, welche auch tatsächlich vom bereitgestellten Bildmaterial gezeigt wurden. Auch wenn durch die entsprechenden Entscheidungen zur Datenerhebung, aus Sicht des Autors, eine bestmögliche Reliabilität im allgemeinen Sinn und speziell im Hinblick auf instrumentelle Konsistenz gewährleistet wurde, müssen die Einschränkungen der Reliabilität aufgrund des Bildmaterials u.a. bei der Interpretation der Ergebnisse zwingend berücksichtigt werden.

Zudem sei auch für diese Studie nochmals explizit auf die zusätzlich zur Systematischen Spielbeobachtung eingesetzte Methode hingewiesen, welche zur Erfassung des vom Schiedsrichter festgelegten Mindestabstands eingesetzt wurde. Diese Variable wurde über eine eigens entwickelte Software via Homographie bestimmt, also über die Umrechnung eines im Videobild erfassten Abstandes (von Pixeln) in die Realität.

2.3.3 Technological Officiating Aids

Das dritte Paper unterscheidet sich nicht nur von der konzeptionellen Herangehensweise, sondern als systematisches Review auch in der Methodik grundlegend von den beiden empirischen Studien. Darüber hinaus wurde auch

betreffend des Vorgehens innerhalb dieses Typus von wissenschaftlichen Beiträgen Neuland betreten. Dies ist zum einen der Natur des vorliegenden Forschungsgegenstands geschuldet, welcher eine Reihe von verschiedenen Hilfsmitteln beinhaltet, welche wiederum in einer Vielzahl von unterschiedlichen Settings eingesetzt werden. Die daraus resultierende Vielfalt an anzuwendenden Methoden schließt a priori Formen der klassischen Ergebnissynthese aus. Zum anderen kam das Bedürfnis des Autors hinzu, Beiträge aus der Philosophie und der Rechtswissenschaft einzubeziehen, welche als essenziell in Bezug auf den Forschungsgegenstand betrachtet wurden. Für Reviews mit einem derartig interdisziplinären Ansatz liegen weder Richtlinien noch dem Autor bekannte Referenzpublikationen vor. Das Vorgehen wurde dementsprechend so gewählt, dass ein international anerkanntes Protokoll, die PRISMA-Richtlinien (Liberati et al., 2009), so weit wie möglich umgesetzt wurde und die nötige Abweichung offensiv in der Publikation kommuniziert wurde. Das erfolgreich abgeschlossene Peer-Review-Verfahren wird als Bestätigung gesehen, dass der gewählte Ansatz berechtigt war.

3 Einzelbeiträge

In diesem Kapitel werden die drei Einzelbeiträge zusammengefasst sowie die jeweiligen Leistungsbeiträge des Promovenden hervorgehoben. Als Kapitelüberschriften dienen wieder jeweils die jeweiligen *Running Heads* bzw. der tatsächliche Beitragstitel, sofern ersteres nicht vorhanden ist.

3.1 Do we need Goal Line Technology?

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. In Cabri, J., Barreiros, J. & Pezarat-Correia, P. (Hrsg.), *Sports Science Research and Technology Support* (S. 107 – 118). Schweiz: Springer International Publishing. DOI: 10.1007/978-3-319-25249-0_8

3.1.1 Eigene Leistungsbeiträge:

Die Studie war Teil eines Projekts zur Notwendigkeit der Torlinientechnologie, welche im Auftrag der Deutschen Fußball Liga GmbH (DFL) durchgeführt wurde. Der Promovend trug maßgeblich zur Entwicklung der Forschungsidee bei, speziell in Bezug auf die Erweiterung um die Untersuchungen zum Videobeweis. Das Beobachtungssystem wurde selbstständig unter der Aufsicht von Prof. Martin Lames und Dr. Daniel Link entworfen. Die Datenerfassung erfolgte unter Leitung des Doktoranden, wobei dieser einen Teil der Daten selbst erhob und zudem die Schulung des anderen Beobachters, Daniel Linke, durchführte. Sämtliche Auswertungen wurden selbstständig durchgeführt, ebenso die Verfassung des Papers. Dies schließt die Literaturrecherche, die Darstellung und Interpretation der Ergebnisse und die theoretische Verortung mit ein.

3.1.2 Zusammenfassung

Das Ziel der Studie war die Überprüfung der Notwendigkeit von Torlinientechnologie und die Untersuchung der Eignung eines Videobeweises für die Klärung kritischer Schiedsrichterentscheidungen im Zusammenhang mit Toren. Im Grundgerüst zur Evaluationsforschung zu technischen Hilfsmitteln wird sie deshalb als

Bewertung der Notwendigkeit anhand des *Wettkampfgeschehens* verortet, unter expliziter Einbeziehung des *Technischen Fortschritts* (siehe Abbildung 3).

Mit Hilfe eines eigens erstellten Beobachtungssystems wurden 1167 Spiele der Fußball-Bundesligen aus drei Spielzeiten (2011/2012 bis 2013/2014) auf kritische Torlinienüberschreitungen, welche auch nicht gegebene Tore enthalten konnten, und kritische Tore generell überprüft. Alle als kritisch identifizierten Ereignisse wurden anschließend von einem aus zwei Personen bestehenden Expertenpool anhand der Videobilder auf die Korrektheit der ursprünglichen Schiedsrichterentscheidung untersucht. Die Ergebnisse sind allesamt deskriptiver Natur, da der Anspruch der Arbeit die explorative Untersuchung von Häufigkeiten der verschiedenen Ereignisse war.

Identifiziert wurden 19,6 torlinien-kritische Ereignisse pro Saison in der Bundesliga und 13,9 in der 2. Bundesliga. Von allen kritischen Toren sind nur 5,0% aufgrund der Torlinienüberschreitung kritisch, während Abseitsentscheidungen mit 84,3% die häufigste Quelle darstellen (Hand 3,3%, Fouls 7,3%). Mit Hilfe der verfügbaren, unspezifischen Videobilder konnten 86,6% aller kritischen Szenen und 76,6% aller torlinien-kritischen Ereignisse aufgeklärt werden. Dementsprechend lagen pro Saison 5,0 (BL) bzw. 2,8 (2. BL) Szenen vor, welche nur durch den Einsatz von Torlinientechnologie hätten aufgeklärt werden können.

Aufgrund dieser Seltenheit werden Zweifel bezüglich des Kosten-Nutzen-Verhältnisses für den Einsatz von Torlinientechnologie geäußert. Erst nach zehn bzw. 19 Saisons erreicht die Bernoulli-Wahrscheinlichkeit eines einmaligen Einsatzes 95%, wobei aber in jeder Saison Aufwendungen in Höhe von ca. 135.000 € entstehen. Vor diesem Hintergrund werden Vor- und Nachteile des Videobeweises sowie die Hindernisse und Rahmenbedingungen bzgl. einer möglichen Einführung diskutiert. Der Fokus liegt dabei auf der Ableitung von Hinweisen aus den Erfahrungen anderer Sportarten vor dem Hintergrund der Wettkampfstruktur im Fußball, wobei eine Anwendung ähnlich zum sogenannten „Booth Review“ in der National Football League als am geeignetsten angesehen wird.

3.2 The use of vanishing spray reduces the extent of rule violations in soccer

Kolbinger, O. & Link, D. (2016). The use of vanishing spray reduces the extent of rule violations in soccer. *SpringerPlus*, 5:1572, 1-7. DOI: 10.1186/s40064-016-3274-2.

3.2.1 Eigene Leistungsbeiträge

Auch dieser Artikel entsprang aus einem Projekt im Auftrag der DFL, nämlich zur Evaluierung des Freistoßsprayeinsatzes in der Bundesliga. Die Evaluationsstrategie wurde nahezu ausschließlich vom Promovenden entwickelt, der zusammen mit Dr. Daniel Link die Projektleitung übernahm. Das Beobachtungssystem wurde mit diesem zusammen entwickelt, wobei der Doktorand die Schulung der eingesetzten Beobachter alleine durchführte. Die Auswertung der Daten und Aufbereitung der Ergebnisse wurde ebenso hauptsächlich eigenständig durchgeführt und nur punktuell unterstützt. Das gleiche gilt für die Fertigstellung des Papers und insbesondere für die Erstellung des theoretischen Grundgerüsts sowie Interpretation und Einordnung der Ergebnisse. Zudem betreute der Promovend eine aus dem Datensatz entstandene Bachelorarbeit, bei der Regelverstöße in Abhängigkeit von Mannschaft und Schiedsrichter untersucht wurden (Scheller, 2016).

3.2.2 Zusammenfassung

Die Intention der Studie war die Evaluation des Freistoßsprayeinsatzes in der Bundesliga hinsichtlich des Wettkampfgeschehens und kann dementsprechend im evaluationsstrategischen Grundgerüst verortet werden (siehe Abbildung 3). Im Detail wurde dafür – größtenteils ausgehend von den veröffentlichten Zielstellungen der Verbände – der Einfluss auf Regelverstöße i.S.v. Verletzungen des Mindestabstands, deren Ahndung, die Festsetzung jener Distanz durch den Schiedsrichter, sowie den Erfolg von Freistößen untersucht.

Die Stichprobe bestand aus allen verfügbaren Freistößen der Bundesligasaisons 2013/2014 und 2014/2015, die als Torschuss oder Flanken mit weniger als 35 Metern Abstand zur gegnerischen Grundlinie ausgeführt wurden. Eine besondere Schwierigkeit bestand in der Identifizierung einer geeigneten „Kontrollgruppe“ für die „Treatmentgruppe“ gesprayerter Freistöße. Dies wurde über einen stringenten

Parallelisierungsprozess auf Basis von u.a. der Position und der Anzahl der Spieler in der Mauer bewerkstelligt und führte zu einer Stichprobe von 299 Paaren. Der Vergleich der beiden Gruppen wurde je nach zugrundeliegender Variable via Chi-Quadrat-Test oder t-test für gepaarte Stichproben durchgeführt.

Dieser Vergleich zeigte, dass nach wie vor ca. 25% der Freistöße von Regelverstößen i.S.v. Verletzungen des Mindestabstands betroffen sind. Jedoch ging die Anzahl massiver Regelverstöße signifikant um ca. 60% zurück ($\chi^2 = 4.58$; $p = .032$). Keiner der Regelverstöße, weder vor noch nach der Einführung des Freistoßsprays, wurde vom Schiedsrichter mit einer gelben Karte oder einer Freistoßwiederholung geahndet. Das Verhalten des Schiedsrichters änderte sich auch nicht in Bezug auf die Festlegung des Mindestabstands, welcher im Mittel weiterhin über 65 cm von der geforderten Distanz abweicht. Ebenso konnte weder für die Erfolgswahrscheinlichkeit von Torschuss- als auch von Flankenfreistößen ein signifikanter Effekt durch die Einführung des Freistoßsprays nachgewiesen werden.

Zusammenfassend kann festgehalten werden, dass die hauptsächliche Zielstellung des Freistoßsprays, eine bessere Einhaltung des geforderten Mindestabstands, zumindest im Sinne einer Eindämmung der Regelverstöße erreicht wurde. Die weiteren Ergebnisse zeigten nicht nur, dass weitere erwartete positive Effekte ausblieben, sondern deckten zudem interessante Phänomene im Zusammenhang mit den betroffenen Regeln auf. So soll das Spray die Durchsetzung eines Mindestabstands unterstützen, welcher aber oftmals nicht den in den Regeln geforderten 9,15 Metern entspricht. Eine weitere Diskrepanz zwischen dem Regelwerk und dessen Anwendung liegt in der Ahndung der entsprechenden Regelverstöße. Vom Regelwerk explizit gefordert wird eine Verwarnung (Gelbe Karte) des fehlbaren Spielers sowie eine Wiederholung des Freistoßes. Jedoch wurde keiner der begangenen Regelverstöße dementsprechend geahndet. Da es sich um ein vergleichsweise einfach beobachtbares Setting für den Schiedsrichter handelt, wird eine Art unbewusste Übereinkunft zwischen den Akteuren i.S.v. *Unwritten Rules* vermutet, wonach die Schwere des vorliegenden Vergehens keine derartige Ahndung erfordert.

3.3 Technological Officiating Aids

Kolbinger, O. & Lames, M. (2017). Scientific approaches to technological officiating aids in game sports. *Current Issues in Sport Science*, 2:001, 1-10. doi: 10.15203/CISS_2017.001

3.3.1 Eigene Leistungsbeiträge

Das systematische Review wurde nahezu komplett selbstständig vom Promovenden angefertigt. Der Doktorvater unterstützte bei methodischen Entscheidungsfindungen, der konzeptionellen Gestaltung sowie kritisch bei der Anfertigung des Manuskripts, bzw. dessen Überarbeitungen. Die Hauptverantwortung für diese Punkte lag aber weiterhin beim Doktoranden. Dieser zeichnet sich auch verantwortlich für die ursprüngliche Forschungsidee.

3.3.2 Zusammenfassung

Die Zielsetzung des Reviews bestand darin, den aktuellen Forschungsstand zu technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern darzustellen, um sportartenübergreifende Problemfelder zu extrahieren. Damit unterstützt diese Literaturanalyse die Entwicklung des evaluationsstrategischen Grundgerüsts für derartige Interventionen.

Die systematische Suche wurde in der letzten Februarwoche 2016 über die Datenbanken *Web of ScienceTM* und *Scopus* durchgeführt. Aufgrund des zahlen- und bedeutungsmäßig hohen Anteils philosophischer und rechtswissenschaftlicher Artikel wurden diese mit einbezogen. Anhand qualitativer Inhaltsanalysen der selektierten Artikel wurden induktiv systemübergreifende Problemfelder für technische Hilfsmittel zur Unterstützung von Schiedsrichtern extrahiert.

Final entsprachen 23 wissenschaftliche Beiträge den Selektionskriterien der Autoren. Diese setzten sich aus neun empirischen Arbeiten, neun philosophischen Papern und fünf rechtswissenschaftlichen Publikationen zusammen. Die chronologische Betrachtung der Veröffentlichungen zeigte dabei zum einen die Neuheit des Forschungsgegenstandes (erste Publikation im Jahr 2000) sowie das aktuell zunehmende Auftreten und Interesse an diesem Phänomen auf (knapp unter 80% der Beiträge aus der aktuellen Dekade). Die am häufigsten betrachtete Intervention stellte

die Überprüfung von Linienentscheidungen im Tennis via Balltracking dar, welche Untersuchungsgegenstand in zehn Studien war.

System- und sportartenübergreifend wurden sieben Haupt-Problemfelder identifiziert: die *underlying phenomena*, also die Grundlage für den Bedarf der Intervention; die Anwendungsfrequenz und die Häufigkeit korrigierter Entscheidungen (*usage- and „success“ rates*); die Schwelle, um jene Korrektur auszulösen (*standard of review*); die Genauigkeit (*accuracy*) der eingesetzten technischen Hilfsmittel; die Auswirkungen der Intervention auf das Wettkampfgeschehen (*nature of the game*); hervorgerufene materielle und immaterielle Kosten (*costs and threats*); und final die Frage nach dem Ausmaß der an das jeweilige System übergebenen Autorität (*authority granted*). In der Komplexität des sozialen Settings existieren naturgemäß Wechselwirkungen zwischen diesen Problemfeldern. Daneben gibt es jedoch weitere Felder, welche für technische Hilfsmittel zur Unterstützung von Schiedsrichter in Sportsportarten noch nicht betrachtet wurden, etwa Einflüsse von und auf Stakeholder. Zukünftige evaluative Forschung zu solchen Interventionen sollte dementsprechend dieser Komplexität Rechnung tragen.

4 Diskussion

Im Rahmen dieser Dissertation wurde ein evaluationsstrategisches Grundgerüst zur Bewertung innovativer technischer Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten erarbeitet. Dieses stand zu den durchgeführten Studien in einem reziproken Verhältnis. So griffen die Beiträge natürlich auf das Grundgerüst zurück oder genauer gesagt auf den zum jeweiligen Zeitpunkt der Verfassung aktuellen Entwicklungsstand. Andersherum beeinflussten die gewonnenen Erkenntnisse aus und um die Artikel auch die Entwicklung des Grundgerüsts, wobei das Review sogar explizit diese Zielstellung verfolgte. Vor diesem Hintergrund sollen nun zuerst jeweils die einzelnen Beiträge retrospektiv diskutiert werden. Anschließend wird im Sinne der themenübergreifenden Diskussion das Grundgerüst evaluationstheoretisch anhand der Kriterien von Cook und Shadish (1986) eingeordnet. Das Kapitel endet mit dem Aufzeigen der globalen Einschränkungen dieser Arbeit.

4.1 Retrospektive Diskussion der Einzelbeiträge

4.1.1 Do we need Goal Line Technology?

In dieser Studie wurde anhand des Wettkampfgeschehens und der technischen Gegebenheiten die Notwendigkeit der Torlinientechnologie untersucht sowie der Videobeweis als mögliche Alternative diskutiert. Die Untersuchung des Wettkampfgeschehens ergab eine niedrige Frequenz von torlinien-kritischen Ereignissen, besonders im Vergleich zu anderen kritischen Schiedsrichterentscheidungen. Bezogen auf die technischen Möglichkeiten wurde das Potenzial des Videobeweises zur Aufklärung sämtlicher kritischer Torentscheidungen, also inklusive torlinien-kritischer Szenen, aufgezeigt sowie die Vorbedingungen und Hürden bzgl. dessen Einführung diskutiert. Dabei wurde sich speziell auf das Wettkampfgeschehen bezogen, wobei der Einfluss des Meinungsbildes der Stakeholder nur an einer Stelle explizit angesprochen wird. An keiner Stelle des Artikels hingegen wird explizit auf die Bedeutung der Stakeholder bezüglich der Bestimmung des Wertes des Evaluanden hingewiesen. Dies geschieht lediglich implizit durch sprachliche Feinheiten, welche aufzeigen, dass vom Evaluator keine abschließende Bewertung vorgenommen wurde. So wird von einem „fraglichen Kosten-Nutzenverhältnis“

gesprochen und dieses nicht abschließend als „gut“ oder „schlecht“ bewertet. Zudem schreiben sich die Autoren lediglich zu, aufgezeigt zu haben, dass die Einführung eines Videobeweises „möglich“ wäre und welcher Ansatz „vielversprechend“ ist. Somit kann sich zwar stichhaltig gegen Kritik i.S.v. Unterstellung eines naiven Realismus erwehrt werden, dennoch erscheint retrospektiv eine explizitere Einbeziehung der Stakeholder in die theoretische Einbettung als sinnvoll.

Hierzu muss gesagt werden, dass der Artikel zu Beginn der Promotion verfasst wurde bzw. zeitnah nach der Festlegung auf dieses Promotionsthema. Zwischen der Verfassung des Artikels und der Dissertationsschrift liegen somit mehr als eineinhalb Jahre intensive Auseinandersetzung mit der Evaluationsforschung im Allgemeinen sowie speziell im Hinblick auf den Untersuchungsgegenstand. Dementsprechend gut lässt sich an dieser Veröffentlichung die Entwicklung der evaluationstheoretischen Position und des damit verbundenen evaluationsstrategischen Grundgerüsts aufzeigen¹⁸. Dies betrifft vor allem die wissenschaftstheoretische Position und eben die Bewertung der Rolle von Stakeholdern bei der Bestimmung des Wertes des Evaluanden.

Die Bedeutung des Meinungsbildes der Stakeholder wird auch durch die aktuellen Entwicklungen bzgl. der Torlinientechnologie in der Praxis unterstützt. Wie erwähnt, wurde die Studie im Auftrag der DFL durchgeführt, mit der Intention die Ergebnisse den stimmberechtigten Vereinsdelegierten als Informationsmaterial, im Vorfeld einer Abstimmung über die Torlinientechnologie, zur Verfügung zu stellen. Bei dieser Abstimmung am 24.03.2014 sprachen sich die Delegierten gegen eine Einführung der Torlinientechnologie in den Bundesligen aus.

Die öffentlichen Reaktionen fielen eher negativ aus, was exemplarisch an den Reaktionen führender Printmedien aufgezeigt werden kann. Der Kicker, die auflagenstärkste Sportzeitschrift mit dem Schwerpunkt Fußball, wirft der Bundesliga vor, „am falschen Ende zu sparen“. Die Süddeutsche Zeitung, Deutschlands auflagestärkste überregionale Qualitätszeitung, sprach gar von „einem tristen Montag für Fairness und

¹⁸ Ebenso aufzeigen lässt sich, anhand eines sprachlichen Vergleiches dieses Artikels mit den anderen beiden Veröffentlichungen, der verbesserte Umgang mit der englischen Sprache. Dass dies an diesem ersten Artikel so besonders deutlich wird, ist aber auch den unorthodoxen redaktionellen Prozessen des entsprechenden Publikationsorgans geschuldet.

Gerechtigkeit“. Diese negativen Reaktionen führten in Verbindung mit einem prominenten torlinien-kritischen Ereignis¹⁹ zu einer erneuten Abstimmung im Dezember, diesmal mit positivem Votum bzgl. des Einsatzes in der ersten Bundesliga (mit 15 von 18 Stimmen). In dieser kurzen Spanne lagen keine grundlegenden Änderungen des Wettkampfgeschehens oder der technischen Möglichkeiten vor. Es kann deshalb angenommen werden, dass das veränderte Abstimmungsverhalten eine Reaktion auf das (vermutlich unsystematisch anhand der Pressereaktionen erfasste) Meinungsbild relevanter Stakeholder darstellte. Dies deckt sich nicht nur mit der Einordnung der öffentlichen Meinung bzw. der Presse als wichtigem Faktor für (sport-) politische Entscheidungsträger (Weiss, 1982), sondern unterstreicht zudem die Bedeutung der Stakeholder für die Bewertung des Evaluanden.

Interessant gestaltet sich auch der Blick auf die weiteren Entwicklungen in der Praxis bezüglich des Videobeweises. So genehmigte das für Regeländerungen im Fußball zuständige Gremium, das International Football Association Board (IFAB), auf der Generalversammlung am 05.03.2016 eine Testphase zur Einführung des Videobeweises. Dabei fällt auf, wie nahe die Überlegungen des vorliegenden Artikels bezüglich der Anwendung am nun entwickelten Konzept der IFAB sind. Im Artikel wurde eine Anwendung ähnlich dem sogenannten *Booth Review* aus der NFL (American Football) als besonders vielversprechend für die Bedürfnisse im Fußball identifiziert. Dort werden sämtliche *Turnover* und *Scores*²⁰ sowie sämtliche überprüfbaren Szenen in den letzten beiden Minuten einer Halbzeit obligatorisch von einem Videoschiedsrichter überprüft. In der hier eingereichten Publikation wird eine solche obligatorische Überprüfung im Fußball für Elfmeter- und Torentscheidungen vorgeschlagen. Dabei wird auch auf die aufgrund der von diesen Ereignissen bereits ohne Videobeweis hervorgerufenen Unterbrechungen hingewiesen, welche im Schnitt mit knapp unter einer

¹⁹ Im Pokalfinale am 17.05.2014 klärte der Münchener Dante einen Kopfball des Dortmunder Mats Hummels – laut offizieller Schiedsrichterentscheidung – auf der Torlinie. Die Schiedsrichterentscheidung konnte jedoch nicht durch die TV-Bilder bestätigt werden. Der Antrag des FC Bayern München auf eine erneute Abstimmung wurde vier Tage später, am 21.05.2014, eingereicht. Als explizite Gründe führte der Antragsteller auf, dass der ursprüngliche Antrag speziell von Mitgliedern der 2. Bundesliga aus Kostengründen abgelehnt wurde.

²⁰ Die Struktur eines Football-Spiels ist unterteilt in klar voneinander abgegrenzte sogenannte *Downs*, was ins Deutsche mit Spielzug im Sinne eines Zuges im Schach übersetzt werden könnte. Werden mit so einem Spielzug Punkte irgendeiner Art erzielt, spricht man von einem *Score*. Gelangt die gegnerische Mannschaft innerhalb oder durch einen Spielzug unbeabsichtigt, aus Sicht des zuvor ballbesitzenden Teams, in Ballbesitz, spricht man von einem *Turnover*.

Minute (Siegle & Lames, 2012) für eine Überprüfung via Videoschiedsrichter ausreichen könnten. Das Konzept der IFAB sieht nun genau so eine obligatorische Form der Überprüfung für kritische Tor- und Elfmeterentscheidungen vor, erweitert um direkte Platzverweise und (vermeintliche) Spielerverwechslungen bei Verwarnungen.

Sollte der Videobeweis tatsächlich dauerhaft eingeführt werden, bringt dies direkte Konsequenzen für die Torlinientechnologie mit sich, da dieser potenziell ebenfalls in der Lage ist kritische Torlinienentscheidungen aufzulösen. Der entsprechende positive Einfluss der Torlinientechnologie auf das Wettkampfgeschehen würde abnehmen oder ggf. komplett verschwinden. Das Hilfsmittel zur Quasi-Substituierung des Schiedsrichters (Torlinientechnologie) könnte in Zukunft also selbst durch ein anderes Hilfsmittel (Videobeweis) substituiert werden.

4.1.2 The use of vanishing spray reduces the extent of rule violations in soccer

Bei der Untersuchung des Freistoßsprayeinsatzes sind die in dieser Dissertation dargelegten evaluationstheoretischen Überlegungen offensichtlicher. Zwar fokussiert sich die Evaluation auf das Wettkampfgeschehen, jedoch wird auch auf die anderen beiden Ströme explizit eingegangen. Speziell hervorgehoben wird die Bedeutung der Stakeholder bei der Bewertung des Evaluanden, auch im Hinblick auf die Parallelen zur Torlinientechnologie. Beide Technologien kreieren bezüglich des Wettkampfgeschehens nur einen kleinen Vorteil, dieser wird jedoch scheinbar von den Entscheidungsträgern – unter Berücksichtigung weiterer Einflussfaktoren – als jeweils ausreichend angesehen, um den Einsatz des technischen Hilfsmittels zu rechtfertigen.

Bei der isolierten Betrachtung des Freistoßsprays fällt auf, dass die entsprechenden Verbände als Vorteil des Freistoßsprays vor allem den Rückgang an dementsprechenden Gelben Karten und Freistoßwiederholungen sehen. Ein derartiger Effekt konnte für die vorliegende Stichprobe nicht nachgewiesen werden, was daran lag, dass weder in der Kontroll- noch in der Treatmentgruppe Regelverstöße dementsprechend geahndet wurden. Es konnte aber aufgezeigt werden, dass entsprechende Regelverstöße dennoch vorlagen. Als wahrscheinlichsten Grund für die ausbleibende Ahndung wird die Existenz sogenannter *Unwritten Rules* (D'Agostino, 1995) gesehen, da es sich um ein vergleichsweise übersichtliches Setting für den Schiedsrichter handelt. Es wird also von einer unausgesprochenen, eventuell sogar

unbewussten Übereinkunft zwischen den Akteuren auf dem Spielfeld ausgegangen, wonach die im Regelwerk vorgesehene Bestrafung nicht angemessen ist. Die durchgeführte Studie war damit in der Lage, Informationen über den eigentlichen Untersuchungsgegenstand hinaus zu generieren, in diesem Fall zur Anwendung der zugrundeliegenden Regel. Dies gilt auch für die Ergebnisse bezüglich der Abschätzung des Mindestabstands durch den Schiedsrichter, welche im Durchschnitt um über 65 cm von der im Regelwerk geforderten Distanz abweicht. Zusammenfassend kann also festgehalten werden, dass ein technisches Hilfsmittel zur Unterstützung der Durchsetzung einer Regel eingeführt wurde, welche aber selbst nicht vollständig dem Regelwerk entsprechend angewendet wird.

Nimmt man an, dass die Hauptintention der Verbände bzgl. des Wettkampfgeschehens die Gewährung des regelkonformen Mindestabstands war, so müsste das Freistoßsray durch weitere Maßnahmen unterstützt werden. Eine solche Maßnahme, nämlich laserunterstützte Messgeräte zur Bestimmung des korrekten Mindestabstands, wurde bereits 2014 von der IFAB diskutiert, jedoch (noch) nicht eingeführt. Flankiert werden sollte die Ausstattung mit technischen Hilfsmittel(n) aber zudem von Maßnahmen, welche an der Problematik bzgl. der Ahndung von Mindestabstandsverletzungen ansetzen. Eine Möglichkeit wäre, das Bewusstsein für die korrekte Regelanwendung in solchen Fällen nochmals zu schärfen. Teilt man die Annahme, dass die Diskrepanz zwischen dem Regelwerk und der Anwendung auf dem Feld ein Fall von *Unwritten Rules* ist, wäre alternativ auch eine dementsprechende Anpassung des Regelwerks möglich.

Die Ergebnisse der Studie zeigen dementsprechend auch auf, wie sinnvoll Evaluationsforschung bereits vor der Einführung von innovativen technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern in Sportsportarten wäre. Die besprochenen Phänomene der Mindestabstandsschätzung und Regelanwendung hätten bereits im Zuge einer Bewertung der Notwendigkeit des technischen Hilfsmittels aufgedeckt werden können. Dies betrifft nicht nur eine Betrachtung des Wettkampfgeschehens, sondern auch eine Untersuchung des Meinungsbilds der Stakeholder, auch bezüglich der erlangten Erkenntnisse. Gegebenenfalls hätte man dadurch direkt weitere geeignete Hilfsmittel, zum Beispiel zur Bestimmung des genauen

Abstands auf dem Feld, mit in die Interventionsstrategie einbeziehen können. Alternativ wäre sogar denkbar gewesen, der Empfehlung Borooahs (2016) zu folgen und anstatt in technische Hilfsmittel vermehrt in die Schiedsrichterausbildung bzw. –fortbildung zu investieren. Die Studie zum Freistoßspray unterstützt dementsprechend die Annahmen des idealtypischen Grundgerüsts für entsprechende Evaluationsstrategien.

4.1.3 Technological Officiating Aids

Wie oben beschrieben, stellen technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Sportsportarten einen sehr vielschichtigen Forschungsgegenstand dar. Dies liegt zum einen an der Komplexität der Einflüsse auf und Auswirkungen von Interventionen in sozialen Settings, zum anderen schlicht an dem Fakt, dass eine Reihe derart verschiedener Settings abgebildet wird. Ziel des systematischen Reviews war deshalb die Untersuchung der relevanten Literatur zur Extraktion system- bzw. sportartenübergreifender Problemfelder im Zusammenhang mit diesem Forschungsgegenstand. Darüber hinaus liefert eine derartige Literaturarbeit auch ein detailliertes Bild des Forschungsstandes. Hervorzuheben sei hierbei nochmals die Neuartigkeit und Aktualität des Themas, mit der ersten Veröffentlichung im Jahr 2000, sowie das steigende Interesse an diesem Phänomen mit einer großen Mehrheit der Studien mit einem Veröffentlichungsdatum in dieser Dekade. Zudem zeigte sich, dass diese Thematik Disziplinen über die Sportwissenschaft hinaus bedient bzw. durch diese bedient werden kann, wie beispielsweise die rechtstheoretischen Beiträge zeigen.

Die Extraktion der bisher identifizierten übergreifenden Problemfelder führte zu zwei universellen Erkenntnissen zu technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern. Erstens zeigt das Spektrum der Problemfelder wie auch das Spektrum innerhalb dieser Felder die Vielschichtigkeit der vorliegenden Thematik. Fragestellungen beziehen sich beispielsweise auf Themen wie Genauigkeit (*accuracy*), aber auch auf das Ausmaß an auf das Hilfsmittel übertragene Autorität (*authority granted*). Der exemplarische Blick in das Problemfeld *accuracy* zeigt darüber hinaus grundlegend unterschiedliche Fragestellungen bezüglich der Genauigkeit. So untersucht Cross (2014) empirisch die Exaktheit des virtuellen Ballabdrucks, während Collins (2010) und Royce (2012) sich über die Rolle der Genauigkeit für die Gerechtigkeit von Schiedsrichterentscheidungen streiten.

Zweitens lassen sich anhand der behandelten Felder die Forschungslücken aufzeigen. Wie bereits diskutiert, da Ausgangspunkt dieser Arbeit, trifft dies universell für die Betrachtung technischer Hilfsmittel zur Unterstützung von Schiedsrichtern als Intervention und somit als Gegenstand von evaluativer Forschung zu. Unterstellt man den vorhandenen empirischen Arbeiten eine implizite Evaluationsabsicht, so fehlen Darstellung des theoretischen Hintergrunds und besonders der Einbezug des Meinungsbildes der Stakeholder.

An dieser Stelle sollte außerdem explizit auf die Studie zum Freistoßsray im Hinblick auf das Review eingegangen werden. Der Artikel zur Torlinientechnologie wurde bereits vor der Datenbankrecherche publiziert und konnte dementsprechend mit aufgenommen werden. Das Paper zur Evaluation des Freistoßsprayeinsatzes wurde aber erst nach dem Zeitintervall für die Datenbankrecherche und erst kurz vor der Einreichung des Reviews vom entsprechenden Publikationsorgan zur Veröffentlichung akzeptiert. Dementsprechend konnte diese Veröffentlichung nicht mit aufgenommen werden, weshalb sie an dieser Stelle thematisch im Hinblick auf das Review verortet werden soll. Wie der Artikel zur Torlinientechnologie, behandelt auch dieser Artikel das zugrundeliegende Phänomen (*underlying phenomena*), also die Regelverstöße und das damit zusammenhängende Schiedsrichterverhalten. Zudem wird die Häufigkeit des Einsatzes dargestellt (*usage rate*) und Auswirkungen auf den Verlauf der Freistöße über die Regelverstöße hinaus beschrieben (*nature of the game*).

Abschließend soll nochmals angemerkt werden, dass die Themenkategorien induktiv formuliert wurden. Dies impliziert– auch bei systematischen – Vorgehen einen hohen subjektiven Anteil. Andere, objektivere Verfahren (vgl. Mayring, 2008) der qualitativen Inhaltsanalyse erschienen nicht geeignet, sowohl für einen derartig breiten Untersuchungsgegenstand als auch für ein Review aus derartig unterschiedlichen wissenschaftlichen Richtungen.

4.2 Evaluationstheoretische Position des vorliegenden Grundgerüsts

Die evaluationstheoretische Einordnung des erarbeiteten Forschungsansatzes wird anhand der Klassifikation von Cook und Shadish (1986) vorgenommen, was die Notwendigkeit einiger Vorbemerkungen mit sich bringt. Die beiden Autoren nutzten diese Klassifikation zur Systematisierung von Theorien zur Evaluationsforschung (ein

Begriff, den sie austauschbar mit Programmevaluation verwenden), wobei sie darunter exklusiv den Kontext der Bewertung sozialer Programme i.S.v. politischer Programme - z.B.: zur Reduzierung der Arbeitslosigkeit - verstehen. Um soziale Interventionen von solch gesellschaftlicher Bedeutung handelt es sich bei innovativen technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern nicht, allerdings, wie bereits in der Methodik erwähnt, um Interventionen im sozialen Setting der Spielsportarten. Dementsprechend eignet sich die Verwendung der obigen Systematisierung nach Cook und Shadish (1986), allerdings muss diese unter dem Bewusstsein dieses Unterschieds vorgenommen werden. Im Zuge der Verortung des evaluationsstrategischen Grundgerüsts wird darüber hinaus explizit auf daraus resultierende Abweichungen hingewiesen.

Die Klassifikation von Cook und Shadish (1986) enthält fünf Kriterien zur Systematisierung der Evaluationstheorien, welche von Lames (2000) folgendermaßen ins Deutsche übersetzt wurden: *Aspekte der Theorie sozialer Programme*, *Verwendung von Wissen aus Evaluationen*, *Theorie der Bewertung sozialer Interventionen*, *Wissenschaftstheoretische Position* und *praktische Umsetzung von Evaluationen*. Diese ergeben sich mit Ausnahme der *wissenschaftstheoretischen Position* („Knowledge base about constructing valid knowledge“) wortwörtlich aus den Originalbezeichnungen von Cook und Shadish (1986), jedoch entfernt sich Lames strukturell von dem Klassifikationsschema. So wird die *Verwendung von Wissen aus Evaluationen* unter *Aspekte der Theorie sozialer Programme* abgearbeitet und die *praktische Umsetzung von Evaluationen* implizit abgehandelt. In dieser Arbeit wird sich wieder näher an der ursprünglichen Klassifikation orientiert und werden zudem die Bezeichnungen, welche in dieser Diskussion auch als Kapitelüberschriften dienen, für diesen Kontext angepasst. Die praktische Umsetzung für den Kontext innovativer technischer Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten wird jeweils im Zuge der Diskussion der anderen vier Kriterien aufgezeigt.

4.2.1 Natur von Interventionen im sozialen Setting Spielsportarten

Wie bereits erwähnt, handelt es sich beim vorliegenden Evaluanden nicht um ein soziales Programm nach dem Verständnis von Cook und Shadish (1986), was zu der entsprechenden Anpassung in der Bezeichnung (Interventionen im sozialen Setting

Spielsportarten) führt. Bezüglich der systematischen Stellung sind diese Interventionen aber den Programmen zumindest ähnlich, da es sich bei der Einführung technischer Hilfsmittel zur Unterstützung von Schiedsrichtern um eine Maßnahme bzw. ein Bündel an Maßnahmen handelt, welche/s in eine Handlungsstrategie in Bezug auf ein bestimmtes Thema oder Problemfeld eingebettet ist. Dabei bleibt jedoch unklar, auch aufgrund der in der Methodik bereits angeschnittenen Kommunikationspolitik der Verbände, wie explizit diese Handlungsstrategie ist. In Bezug auf die bereits eingeführten und zukünftigen technischen Hilfsmittel zur Unterstützung von Schiedsrichtern im Fußball spricht beispielsweise der ehemalige FIFA Präsident Blatter (2014) diesen den Beitrag zu, die Sportart Fußball voran zu bringen. Neben derart globalen oder weichen Strategien wird vom Autor den Verbänden zumindest die implizite Handlungsstrategie unterstellt, die Anwendung der entsprechenden Regelwerke verbessern zu wollen.

Die Auffassungen über den Kontext dieser Strategien soll nun mit dem der sozialen Programme verglichen werden, was anhand der von Cook und Shadish (1986) (für soziale Programme) abgelehnten rationalen Modelle zur Lösung von Problemen vorgenommen wird²¹. Diese Modelle gehen aus von der Annahme, dass ein klar definiertes Problem vorliegt. Für die fehlerhafte Anwendung von Regeln kann dies bezüglich des Wettkampfgeschehens verhalten optimistisch gesehen werden. Sofern geeignete Methoden vorliegen, können zumindest Wahrscheinlichkeiten aufgezeigt werden, wie häufig (bzw. ob und wann) fehlerhafte Anwendungen vorliegen. Allerdings können auch diese Probleme, wie bzgl. sozialer Programme (Bryk, 1983), von verschiedenen Interessensgruppen unterschiedlich bewertet werden.

Weiter gehen simple entscheidungstheoretische Modelle davon aus, dass eine Reihe verschiedener und geeigneter Lösungsvorschläge generiert wird. Auch das kann für die Einführung der meisten technischen Hilfsmittel mindestens theoretisch beansprucht werden. Handelt es sich bei dem Problemfeld zum Beispiel um Fehler bei der Bestimmung der genauen Position des Spielgeräts, so gibt es eine Reihe von Verfahren, diese zu bestimmen. In der Praxis hat die FIFA bereits sogar

²¹ Diese ist ähnlich der im gleichen Werk aufgeführten Suchman'schen (1967) Annahmen, allerdings detaillierter aufgeteilt, noch nicht für den Kontext sozialer Programme spezifiziert und eignet sich deshalb besser zur Diskussion des vorliegenden Kontexts.

unterschiedliche solche Verfahren für die Überprüfung torlinien-kritischer Ereignisse akkreditiert, nämlich sowohl Bildererkennungs- als auch magnetfeldbasierte Systeme. Neben der Wahlmöglichkeit zwischen verschiedenen Systemen, können verschiedene Auswahlmöglichkeiten auch unterschiedliche Ausmaße und Muster der Implementierung sein, also wie oft und vor allem in welchen Situationen technische Hilfsmittel eingesetzt werden dürfen.

Anschließend werden als geeignet erachtete Maßnahmen angewendet und deren Anwendung evaluiert. Theoretisch betrachtet ist auch dies zumindest teilweise möglich, unter Berücksichtigung der in der Methodik gestellten Forderung an das Setting. Unter rein experimentellen Gesichtspunkten betrachtet wäre das die Einführung des Hilfsmittels in dem tatsächlich für die Verwendung vorgesehenen Wettbewerb. Dieser hat aber sehr wahrscheinlich die Eigenschaft, dass er von den beteiligten Stakeholdern als bedeutsam angesehen wird. Dementsprechend schwierig dürfte es sein, den Einsatz von Hilfsmitteln zu Testzwecken vor den Stakeholdern zu vertreten. Aus diesem Grund werden für erstmalige Anwendungen oftmals andere, als unbedeutender angesehene Wettbewerbe gewählt. Ein weiteres Problem, welches in der Praxis auftreten kann, ist, dass nachdem ein Hilfsmittel erstmal eingeführt wurde nur noch geringes Interesse der Entscheidungsträger an Evaluationen vorliegt, speziell wenn diese die Einführung als (sport-)politischen Erfolg ansehen.

Als ein Problem bezüglich der Evaluierung bei sozialen Programmen wird von Shadish (1984) aufgeführt, dass die zu erwartenden Effekte oftmals nur gering sind. Dies ist für die hier besprochenen Interventionen nicht zwingend der Fall, kann jedoch bezüglich des Wettkampfgeschehens durchaus auftreten, wie in dieser Arbeit bereits aufgezeigt wurde. Sowohl das Freistoßsplay als auch die Torlinientechnologie generieren einen eher geringen (positiven) Effekt auf das Wettkampfgeschehen. Beide technischen Hilfsmittel haben zudem jedoch gemeinsam, dass sie in immer mehr Wettbewerben zum Einsatz kommen. So wurde das Freistoßsplay im vergangenen Jahrzehnt exklusiv im süd- und mittelamerikanischen Raum eingesetzt, seit 2013 aber auch bei FIFA- und UEFA-Turnieren, sowie in mehreren nationalen Ligen in unter anderem Europa und Asien. Die Torlinientechnologie wurde noch 2014 lediglich in FIFA-Wettbewerben und der höchsten englischen Spielklasse (Premier League) genutzt,

hinzukamen seitdem unter anderem die Bundesliga sowie die UEFA Champions League.

Zusätzlich sind dem Autor keine Wettbewerbe bekannt, die eines der beiden Hilfsmittel nach einer Einführung wieder abgeschafft haben, mit Ausnahme der temporären Aussetzung des Videobeweises in der National Football League. Der geringe Effekt im Wettkampfgeschehen wird also in der Regel als groß genug angesehen, was wiederum die Bedeutung von Stakeholdern bei der Bestimmung des Wertes aufzeigt. Jeder Stakeholder betrachtet eben nicht nur die Einflüsse auf das Wettkampfgeschehen alleine, sondern stellt diese mindestens implizit den Kosten bzw. negativen Effekten gegenüber, welche auch – ebenso wie die Mehrwerte – immaterieller Art sein können. Aufgrund des Zweifels, sämtliche dieser Aspekte und deren Wechselwirkungen abbilden zu können, speziell im Hinblick auf die Subjektivität der letztendlichen Bewertung, sieht der Autor die Anwendung solch naiver Modelle zur Problemlösung auch für diesen Kontext als nicht geeignet an.

Als nächsten Schritt sehen solche Modelle laut Cook und Shadish (1986) vor, dass das aus den Evaluationen gewonnene Wissen verbreitet werden soll und die Entscheidungsträger basierend auf diesem Wissen eine Entscheidung fällen. Die Verbreitung der Evaluationsergebnisse kann dabei durchaus heikel sein, was bei Verbänden zu dem Bedürfnis führen kann, gewisse Ergebnisse nicht öffentlich zu machen, zum Beispiel um die Schiedsrichter zu schützen, wenn die Evaluationsergebnisse fehlerhaftes Schiedsrichterverhalten aufdecken. Ebenso können politische Motive der Entscheidungsträger zu einem geringen Interesse an Transparenz führen. Bezüglich der Rationalität der am Ende gefällten Entscheidung ist die Stimmung bzgl. sozialer Programme pessimistisch. So spricht Weiss (1993) den Ergebnissen von Evaluationen eher eine untergeordnete Rolle zu und sieht die Hauptgrundlage der Entscheidung in der politischen Position des Entscheidungsträgers. Hinzu kommt die angesprochene Bedeutung der öffentlichen Meinung (Weiss, 1986), welche aber keineswegs ein zwingendes Ausschlusskriterium für Rationalität darstellt. Wird diese Meinung systematisch und mit wissenschaftlichen Methoden erfasst, kann sie durchaus als eine Entscheidungsgrundlage bzw. eine der Entscheidungsgrundlagen herangezogen werden. Rationalen Entscheidungstheorien folgt dagegen nicht eine

unsystematische Orientierung an den Medien. Legt man Evaluationsforschung von technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern das erarbeitete Grundgerüst zugrunde, könnte eine rationale Entscheidung in dem Sinn beansprucht werden, dass die öffentliche Meinung als ein Ergebnis der Evaluation mit herangezogen wird. Für den Einfluss der politischen Position des Entscheidungsträgers kann hingegen nicht kontrolliert werden.

4.2.2 Verwendung von Wissen aus Evaluationen

Direkt verbunden mit dem letzten Punkt im vorangehenden Kapitel zur Natur von Interventionen ist die Frage nach der Verwendung des Wissens aus Evaluationen²². Inzwischen sehr pessimistisch gesehen wird die Annahme, dass aus Evaluationsforschung generiertes Wissen zu bedeutenden Änderungen der evaluierten sozialen Interventionen führt. Für Cook und Shadish (1986) besteht sogar Konsens unter Evaluationsforschern, dass Evaluationen nur selten zu Anpassungen des Budgets für Interventionen führen. Fälle, in denen soziale Programme aufgrund von Evaluationen abgebrochen worden sind, sind ihnen gänzlich unbekannt und Evaluationen lediglich „[...] another political act that occurs in a context where power, ideology, and interests are more powerful determinants of decision making than feedback about programs“ (1986, S. 200). Für die hier untersuchten Interventionen im sozialen Setting Sportsportarten zeigen die Erfahrungen aus und um die beiden durchgeführten Studien, dass Evaluationen in diesem Kontext keine Ausnahme darstellen. Dies wäre bei einem gesellschaftlichen Teilsystem, welches nicht autonom vom gesamten gesellschaftlichen System operiert (Schimank, 2005), auch eine sehr überraschende Erkenntnis.

Einen Grund für die schwache Rolle von Evaluationsergebnissen in Entscheidungsprozessen sehen Cook und Shadish (1986) auch in der oben angeschnittenen Annahme, dass die Ergebnisse von Evaluationen oftmals nicht zwingend genug sind. Zwar gab es in der Sportwelt bereits Fälle, in denen technische Hilfsmittel wieder abgeschafft, ausgesetzt oder erheblich verändert wurden – so schaffte beispielsweise die National Football League Anfang der Neunziger vorübergehend das dortige Videobeweis-System (Instant Replay) ab – es bleibt aber unklar, inwieweit

²² Für Lames (2000) ist diese Verbindung sogar so unmittelbar, dass er die beiden Klassen unter seinem Unterpunkt „Aspekte der Theorie sozialer Programme“ zusammenfasst.

Evaluationen diese Entscheidung beeinflusst haben. Die NFL (2016) nennt zwar als Gründe der Aussetzung die mangelnde Genauigkeit des Systems sowie dessen geringen Einfluss auf das Wettkampfgeschehen, allerdings wurde die letztendliche Entscheidung von den Besitzern der damals 28 Teams getroffen, in deren jeweiligen Entscheidungsprozesse kein Einblick besteht. Dies ist ähnlich der oben beschriebenen Situation bezüglich der Torlinientechnologie.

Wenn also evaluative Forschung Entscheidungen bezüglich des Evaluanden nur bedingt beeinflusst, welche Funktion kann sie dann bzw. darüber hinaus noch haben. Weiss (1977) schreibt Evaluationsforschung die Funktion „enlightenment“ zu, also die fundierte Aufklärung über das zugrundeliegende Problem und die Auswirkungen der Intervention. Diese Informationen sollen dabei nicht nur die Entscheidungsträger aufklären, sondern auch interessierten anderen Stakeholdern zugänglich gemacht werden (Guba & Lincoln, 1989). Dieser diffuse Anspruch kann für den Kontext technischer Hilfsmittel zur Unterstützung von Schiedsrichtern beansprucht werden, allerdings mit den oben genannten Schwierigkeiten, falls Sportverbände als Auftraggeber für Evaluationsforschung auftreten.

Verbunden mit der Frage nach der Verwendung des Wissens ist auch die nach der Rolle des Evaluators gegenüber den Entscheidungsträgern und weiteren Stakeholdern. Geht man davon aus, dass es immer einen Auftraggeber für die Evaluationen gibt, – irgendjemand muss den Evaluator am Ende schließlich bezahlen – kommt man nicht an Wholey (1983) Auffassungen vorbei. Dieser sieht die hauptsächliche Aufgabe in der Beschaffung von Informationen über Kontext und Programm zur Unterstützung der Entscheidungsträger. Idealtypisch würde man sich darüber hinaus, i.S.v. „enlightenment“, eine Funktion als Lehrer sämtlicher Stakeholder inkl. der Öffentlichkeit wünschen, wie von Cronbach und Shapiro (1982) gefordert. Dieser fordert darüber hinaus, dass der Evaluator dabei nicht nur Informationen weitergibt, sondern auch seine daraus generierten Schlussfolgerungen. Cook und Shadish (1986, S. 208) interpretieren die vorherrschende Meinung unter Evaluationstheoretikern sogar folgendermaßen: „Evaluators should do all they can to get publicity for their findings in the mass media and professional outlets.“ Diese idealtypische Umsetzung der Aufklärungsfunktion erscheint in der Praxis eher schwierig,

da er Sportverbände abschrecken könnte, ihre sozialen Interventionen zu evaluieren. Für Evaluationsforscher aus der Wissenschaft ist die Veröffentlichung von Erkenntnissen aber in der Regel ein Bedürfnis. Dies kann zu nicht zu unterschätzenden Interessens- und Gewissenskonflikten bei wissenschaftlichen Evaluationen von technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern führen.

Eine weitere Funktion des Evaluators, die unter anderem von Cousins und Earl (1995) vertreten wird, könnte hingegen das Interesse von Sportverbänden für die Beauftragung von Evaluationsforschern steigern. So sehen Cousins und Earl (1995) den Evaluator in ihrem „participatory“-Modell als Ausbilder der Mitarbeiter in dem Sinn, dass diese weitestgehend eigenständig in der Lage sind, Evaluationsforschung durchzuführen. Die Rolle des Evaluators verschiebt sich dabei idealtypisch immer mehr in Richtung eines Beraters. In dieser Rolle sieht den Evaluator auch Fetterman (1998), allerdings als Berater für sämtliche Stakeholder, was in der Realität aber nur schwer umsetzbar erscheint. Darüber hinaus zeigt Fetterman (1998) aber auf, warum solch schulende Modelle besonders geeignet sind. Evaluationsforschung ist ein fortlaufender Prozess, da der Wert von Programmen oder Interventionen nicht statisch ist. Dementsprechend vorteilhaft ist die Fähigkeit zu evaluieren, verbunden mit einem Klima, welches Evaluationsforschung gegenüber aufgeschlossen ist. Genau diese Auffassung über Wertedynamik und Evaluationsforschung als fortlaufende Prozesse wird auch im entwickelten Grundgerüst der Evaluationsstrategien abgebildet. Die einbeziehenden Modelle erscheinen deshalb als besonders geeignet zur Bewertung technischer Hilfsmittel.

4.2.3 Bewertung von Interventionen

Wie in Cook und Shadish (1986) sowie Lames (2000) ist dieser Absatz nach den Stufen von Scriven's (1980) „logic of evaluation“ strukturiert. Dieser unterteilt den Bewertungsprozess in vier Stufen: Erstens müssen angemessene Kriterien bestimmt werden, anhand derer die Leistung des Evaluanden bestimmt werden kann. Zweites werden Standards i.S.v. Richtwerten fixiert, die Evaluanden erreichen sollten. Der dritte Schritt ist die Erfassung der Leistung anhand der Kriterien. Final sollen alle erlangten Ergebnisse zu einem Statement über den Wert des Evaluanden zusammengetragen

werden. Diese Stufen sollen nun in Hinblick auf den hier vorliegenden Kontext diskutiert werden.

Bezüglich der Bestimmung von Kriterien bieten sich intuitiv die damit verbundenen Zielstellungen an, welche sich global aus der Art des eingesetzten Hilfsmittels ergeben. Handelt es sich um eine Technik, die den Schiedsrichter für eine bestimmte Entscheidung ersetzt, muss diese für diesen Typ von Entscheidungen eine höhere Rate korrekt entschiedener Fälle aufweisen. Unterstützt das Hilfsmittel den Entscheidungsprozess, so sollte dieser Entscheidungsprozess fehlerfreier sein als ein Entscheidungsprozess ohne Hilfsmittel. Wird durch die Innovation die Durchsetzung von Regeln unterstützt, dann werden weniger diesbezügliche Regelverstöße erwartet. Allerdings wird die exklusive Orientierung an den Zielstellungen kritisch gesehen, unter anderem da Zielstellungen oftmals nur vage (Weiss, 1977) oder politisch motiviert formuliert sind (Cook und Shadish, 1986). Für den Einsatz technischer Hilfsmittel zur Unterstützung von Schiedsrichtern wären, zumindest bezogen auf das Wettkampfgeschehen, klare Zielvorstellungen denkbar. Zum Beispiel ein Rückgang der Mindestabstandsverletzungen bei Freistößen im Fußball um einen gewissen Prozentsatz durch die Nutzung des Freistoßsprays. Allerdings waren die entsprechenden Ziele in der Praxis von den Verbänden vager formuliert (siehe oben).

Hinzu kommt, dass Interventionen auch unbeabsichtigte Nebenwirkungen hervorrufen können, beispielsweise ein verändertes Schiedsrichterverhalten oder mehr und längere Spielunterbrechungen. Ein damit verbundenes Problem ist die Identifizierung von Quellen solcher Nebenwirkungen, also die Frage: Wo könnten welche weiteren Effekte auftreten? Dies erfordert vom Evaluator in diesem Kontext ein vertieftes Verständnis der Sportart, welches nach dem hier vorliegenden Modell a priori durch die Erfassung des Meinungsbildes der Stakeholder unterstützt werden soll. Es kann dabei aber nicht per se davon ausgegangen werden, dass dadurch sämtliche möglichen Nebenwirkungen im Vorfeld bewusst gemacht werden können. Dementsprechend kommen Cook und Shadish (1986) zu dem Schluss, dass keine unfehlbare Methode existiert, um a priori vollständig alle Quellen für mögliche ungeplante Effekte zu identifizieren.

Der zweite Schritt in Scriven's Logik, das Setzen von Standards, kann absolut oder relativ erfolgen, wobei letzteres laut Cook und Shadish (1986) in der Praxis dominiert. Es wird also nicht nur geprüft, ob ein Evaluand bestimmte Kriterien erfüllt, sondern ob er sie besser erfüllt als alternative Lösungen. Für den hier vorliegenden Kontext werden auch relative Standards bevorzugt. Es soll mindestens Wissen darüber generiert werden, ob die Intervention zu einer Linderung des zugrundeliegenden Problems führt. Gegebenenfalls gibt es sogar die Möglichkeit, mit anderen Interventionen zu vergleichen. Diese müssen nicht zwangsweise ein anderes technisches Hilfsmittel darstellen, sondern können auch nur in der Anwendung des identischen Hilfsmittels variieren. So kann zum Beispiel der Einsatz des Videobeweises für bestimmte Situationen mit dem Einsatz für weniger, mehr oder andere Situationen verglichen werden.

Eine Problematik, die diesbezüglich unter Evaluationstheoretikern diskutiert wird, ist die Identifizierung geeigneter Vergleichs- bzw. Kontrollgruppen. Für die Studie Freistoßsprint wurde ein Vorher-Nachher-Vergleich durchgeführt, angelehnt an den Vergleich von verschiedenen Stufen eines Programms nach Cronbach (1963). Dieses Vorgehen wird für die Evaluation technischer Hilfsmittel als besonders geeignet angesehen, denn es birgt den Vorteil, dass genau in dem Setting bewertet wird, in dem auch interveniert werden soll. Ein Testlauf in einem anderen Wettkampf oder eine Implementierung einer Alternative in einem anderen Wettkampf bringen das Problem der Vergleichbarkeit mit sich. So können sich die Wettkämpfe bzgl. der Einflüsse und Auswirkungen unterscheiden, was beispielsweise die unterschiedlichen Ergebnisse bzgl. der Notwendigkeit von Torlinientechnologie in der Bundesliga verglichen mit der 2. Bundesliga andeuten.

Als dritten Schritt sieht Scriven (1980) die Erfassung der Leistung des oder der Evaluanden bezüglich der aufgestellten Kriterien, worunter er ausschließlich diesen operativen Schritt versteht. Diese Erfassung soll lediglich den finalen Schritt unterstützen, die Zusammenfassung aller Ergebnisse zu einem Evaluanden und die Bildung einer abschließenden Bewertung. Dies beinhaltet für Scriven (1980) auch die Transformation aller verschiedenen Ergebnisse in eine Einheit, sowie die Gewichtung der verschiedenen Kriterien untereinander. Für die vorliegenden Interventionen wird

dieses Vorgehen als ungeeignet betrachtet. Einen Grund stellt die Schwierigkeit dar, die verschiedenen Kriterien untereinander zu gewichten. Nach welchem Vorgehen sollen beispielsweise allein unterschiedliche Einflüsse auf das Wettkampfgeschehen gewichtet werden, zum Beispiel Mindestabstandsverletzungen gegenüber Erfolg von Freistößen nach Einführung des Freistoßsprays. Noch komplexer wird dies unter Einbezug von Kriterien, die die Auswirkungen auf Stakeholder beschreiben sollen.

Selbst wenn sich alle Kriterien in ein metrisches Maß umwandeln ließen, wäre die zusammenfassende Bewertung immer noch ein subjektiver Prozess bezüglich der Gewichtung. So kann man beispielsweise naturgemäß annehmen, dass Athleten den Einfluss auf das Image des Schiedsrichters geringer gewichten als Schiedsrichter selbst. Trotzdem muss sich der Evaluator nicht zwangsweise auf die Rolle des Informationsbeschaffers beschränken, sondern kann zumindest Schlussfolgerungen bzgl. der einzelnen Kriterien ziehen und diese auch den Stakeholdern berichten. Anhand der subjektiven Abschlussbewertungen könnte entweder ein repräsentatives Stakeholder-Panel einen Konsens bzgl. des Wertes des Evaluanden finden oder die Entscheidungsträger anhand dieser Bewertungen eine Entscheidung treffen.

4.2.4 Wissenschaftstheoretische Auffassungen

In diesem Kapitel sollen das entwickelte evaluationsstrategische Grundgerüst sowie die Annahmen über Evaluationsforschung für innovative technische Hilfsmittel zur Unterstützung von Schiedsrichtern in Spielsportarten hinsichtlich der wissenschaftstheoretischen Position diskutiert werden. Viele Annahmen ähneln denen des konstruktivistischen Paradigmas nach Guba und Lincoln (1989), allerdings wird dieses nicht in allen Punkten vollends geteilt. Legt man die historisch-konzeptionelle Einteilung von Lames (2000) zugrunde, handelt es sich aus Sicht des Autors deshalb am ehesten um eine pluralistische Position. Diese Ansicht soll im Folgenden anhand der Annahmen über Ontologie, Epistemologie und Methodologie begründet werden, auch indem speziell die Abweichungen vom konstruktivistischen Paradigma hervorgehoben werden.

Es wird nicht die Auffassung vertreten, dass eine fehlerhafte Anwendung des Regelwerks, kombiniert mit dem Vorhandensein eines geeigneten Hilfsmittels, unabhängig vom betrachtenden Subjekt zur Notwendigkeit einer Intervention führt. Dies

zeigten die Entwicklungen nach der Studie zur Notwendigkeit zur Torlinientechnologie. Gleiches gilt für die Bewertung der Intervention. Man kann dementsprechend Wahrheit über die Notwendigkeit und Anwendung von Interventionen als Konstruktion sehen. Es wird aber die Annahme vertreten, dass darüber hinaus zumindest eine objektiv erfassbare Abweichung von einer Regel vorliegt sowie ebenso messbare Effekte auf mindestens das Wettkampfgeschehen. Ontologisch ist die Regelanwendung zudem Teil eines komplexen Systems und dementsprechend können Interventionen nicht nur im Hinblick auf die Effekte hinsichtlich dieser Regelanwendung betrachtet werden. Interventionen generieren zusätzlich Effekte auf Stakeholder - z.B.: Athleten oder Zuschauer - sowie Effekte auf das Wettkampfgeschehen über die Zielstellung der Maßnahme hinaus. Dementsprechend schwer abzuschätzen sind nicht nur die Folgen der Einführung eines technischen Hilfsmittels, sondern auch Kausalzuschreibungen generell. Führt ein Freistoßspray zu weniger Regelverstößen aufgrund der visualisierten Markierung oder aufgrund einer erhöhten Sensibilisierung für diese Thematik?

Konsequenzen dieser ontologischen Betrachtung für die Epistemologie sind, dass nicht von einem Objekt-Subjekt-Dualismus ausgegangen werden kann und die erlangten Erkenntnisse höchstens als bestmögliche Annäherung an Wahrheit aufgefasst werden können. Die Auswirkungen einer fehlerhaften Regelanwendung, ebenso wie die der diesbezüglich unternommenen Interventionen, können weder vollständig erfasst noch objektiv bewertet werden. Generalisierbarkeit von Wissen aus den durchgeführten Evaluationsforschungen ist dementsprechend nur möglich, wenn die Ergebnisse unter expliziter Darlegung des Kontexts präsentiert werden. Dieser Kontext beinhaltet mindestens eine detaillierte Beschreibung des Settings und des Evaluanden sowie das theoretische Modell der Funktionsweise der Intervention. Möchte beispielsweise ein Fußballverband einen Videobeweis einführen und dafür Ergebnisse aus dem American Football nutzen, benötigt er unter anderem Wissen über die Spielstruktur dieser Sportart und das Klima unter den dortigen Stakeholdern gegenüber solchen Innovationen.

Methodologisch bringt dies mit sich, dass sich nicht ausschließlich auf quantitatives Vorgehen beschränkt werden kann. Qualitatives Vorgehen kommt sowohl bei der Konzeption als auch bei der Erfassung des Meinungsbilds der Stakeholder und bei der Ableitung der Schlussfolgerungen zum Einsatz. Dabei spielen in diesem Prozess

idealtypisch die Auffassungen von Stakeholdern sowohl bei der Konzeptionierung als auch bei der Bewertung wesentliche Rollen. Dementsprechend müssen Verfahren verwendet werden, die in der Lage sind, diese Aspekte so zu erfassen, dass sie an den jeweiligen Stellen einen Beitrag leisten können. Wie im Kapitel „Methoden“ beschrieben, wird die Bewertung als iterativer Prozess unter Berücksichtigung der Meinung sämtlicher Stakeholder, bzw. eines repräsentativen Panels, gesehen. Dies ist durchaus orientiert am Modell von Guba und Lincoln (1989), allerdings wird als wahrscheinlicherer Abschluss eines Evaluationsprozesses nicht der Konsens aller Teilnehmer, sondern eine Entscheidung der Entscheidungsträger unter Berücksichtigung der Bewertungen der einzelnen Interessensgruppen gesehen.

Das quantitative Vorgehen betreffend ist streng experimentelles Vorgehen nur in wenigen Ausnahmefällen möglich, zum Beispiel bei Labortests über Eigenschaften eines technischen Hilfsmittels. Es können für den hier vorliegenden Kontext darüber hinaus praktisch keine randomisierten Stichproben gezogen werden, welche dann als Treatment- oder Vergleichsgruppen jeglicher Art dienen können. Als geeigneter wird quasi-experimentelles Vorgehen im Cronbach'schen Sinn des Vergleichs von verschiedenen Stufen des Programms gesehen (1963). Dies trägt auch den oben beschriebenen pessimistischen Auffassungen Rechnung, überhaupt geeignete Vergleichsgruppen zu finden. Und zwar bezogen auf die Vergleichbarkeit des Settings wie auch aufgrund der besprochenen politischen Gegebenheiten.

4.3 Einschränkungen der Dissertation

Die Einschränkungen der einzelnen Studien wurden bereits in der Methodik und der Diskussion der Einzelbeiträge besprochen. Dementsprechend fokussiert sich dieser Abschnitt auf Limitierungen der bisher getätigten Forschungen auf globaler Ebene. Die überragende Bedeutung der Stakeholder für Evaluationsforschung von innovativen technischen Hilfsmitteln zur Unterstützung von Schiedsrichtern in Sportsportarten wurde ausführlich dargelegt, allerdings in den beiden Anwendungen nur in Ansätzen mit einbezogen. In die Konzeption der Studien flossen neben dem Know-How des Evaluators nur die Auffassungen eines Stakeholders ein, nämlich die des Auftraggebers. Weitere Meinungsbilder von Stakeholdern wurden nicht erfasst, was natürlich auch den wirtschaftlichen Gegebenheiten der akademischen Realität geschuldet ist. Die

dankenswerterweise bereitgestellten Mittel für beide Studien ermöglichten die durchgeführte Erfassung des Wettkampfgeschehens, aber nicht die Umsetzung eines umfassenden Evaluationsprozesses nach dem beschriebenen Grundgerüst.

Des Weiteren ist unklar, ob sich die vollständige Evaluationsstrategie in der Praxis durchsetzen ließe. Ein Grund ist der Aufwand an Ressourcen, ein weiterer das sportpolitische Klima. Sind die Sportverbände interessiert an einer systematischen Erfassung von Meinungsbildern verschiedener Stakeholder, stützen sie sich auf die Meinung von Massenmedien oder suchen sie sich ggf. das Meinungsbild sogar aus den Massenmedien so zusammen, dass es ihren politischen Zwecken dient? Begrüßen sie eine Verbreitung der Ergebnisse von Evaluationen an andere Stakeholder? Und ganz grundsätzlich stellt sich sogar die Frage, ob es bei Entscheidungsträgern überhaupt ein Bedürfnis für evaluative Überprüfungen von Maßnahmen, welche von ihnen selbst eingeführt wurden, gibt?

5 Fazit

Bereits im Artikel „Technological Officiating Aids“ wurde das Fazit gezogen, dort vor allem im Hinblick auf den Umfang allumfassender Evaluationsforschungen, dass nicht jede evaluative Maßnahme alle relevanten Problemfelder bedienen kann. Verbunden mit den in der Diskussion beschriebenen Gegebenheiten der sportpolitischen Realität, kann diese These in dem Sinne erweitert werden, dass nicht jede Evaluation alle Anforderungen an die Evaluationsstrategie erfüllen kann. Allerdings sollte jede Evaluation unter dem Bewusstsein über die Komplexität des Untersuchungsgegenstands an sich, des umrahmenden Settings sowie der Forschungsstrategie Evaluationsforschung durchgeführt werden. Der Evaluator sollte dieses Bewusstsein von seiner Herangehensweise bis zu den Schlussfolgerungen deutlich machen, speziell im Hinblick auf die Limitierungen der jeweiligen evaluativen Maßnahme.

Finanziell potente Sportverbände oder Auftraggeber generell sollten aber motiviert werden, die Einführung innovativer technischer Hilfsmittel mit einer umfassenden Evaluationsstrategie zu begleiten. Die beiden durchgeführten Studien zur Torlinientechnologie und dem Freistoßspray verdeutlichen Notwendigkeit und Nutzen dieser Maßnahmen. Neben den expliziten Aussagen zum Wettkampfgeschehen und den potenziellen technischen Mitteln geschieht dies auch durch das – teilweise implizite – Aufzeigen der Bedeutung des Meinungsbildes relevanter Stakeholder. Zudem konnte die Berechtigung des idealtypischen chronologischen Ablaufes unterstrichen werden.

Alle drei Studien zeigen außerdem auf, welchen Beitrag die Methoden und Werkzeuge der Wettkampfdiagnostik liefern können. Dabei dominierte für die exemplarischen Studien die Methode der systematischen Spielbeobachtung, allerdings sind ähnlich umfangreiche Einsatzmöglichkeiten für andere Methoden der Wettkampfdiagnostik vorstellbar. So könnte zum Beispiel die Positionserfassung sowohl für entsprechende Fragestellungen bezüglich des Wettkampfgeschehens als auch der Eignung oder des Entwicklungsstandes der technischen Hilfsmittel eingesetzt werden. Im Umkehrschluss möchte sich der Autor deshalb auch dafür aussprechen, der Thematik innovativer technischer Hilfsmittel zur Unterstützung für Schiedsrichter mehr

Beachtung in wettkampfdiagnostischer Forschung und Lehre zu schenken, als dies bisher der Fall ist.

Die Gültigkeit dieser Forderung lässt sich auf die gesamte Sportwissenschaft ausdehnen. Die wissenschaftliche Evaluation dieser Interventionen erfordert Wissen und Methoden aus mehreren wissenschaftlichen Disziplinen, zum Beispiel dem Sportmanagement oder der Sportpsychologie, und bietet diesen umgekehrt somit auch Beitragsmöglichkeiten. Evaluative Forschungen zu diesen Hilfsmitteln eignen sich dementsprechend auch besonders für interdisziplinäre Projekte.

Abschließend bleibt festzuhalten, dass es sich bei innovativen technischen Hilfsmitteln zur Unterstützung von Schiedsrichter in Spielsportarten um ein relativ junges und an Bedeutung zunehmendes Phänomen der sportlichen Praxis handelt. Diesem wurde bisher wenig, aber aktuell ansteigende, Beachtung in der Sportwissenschaft geschenkt. Neben dem Phänomen an sich betrifft dies auch die evaluative Forschung zu diesem, sowohl in der wissenschaftlichen Welt als auch in der Sportwelt an sich. Die Hoffnung des Autors der vorliegenden Dissertation ist es, Sportwissenschaft und Sportpraxis für die Notwendigkeit der Evaluation innovativer technischer Hilfsmittel zur Unterstützung von Schiedsrichtern sensibilisieren zu können und zudem eine Basis für dementsprechende zukünftige Maßnahmen schaffen zu können. Damit verbunden ist zudem die Hoffnung, einen Beitrag zu leisten, sportliche Wettkämpfe *fairer* und *gerechter* zu machen.

6 Literaturverzeichnis

- 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
- Alkin, M. C. (2013). Comparing Evaluation Points of View. In M. C. Alkin (Hrsg.), *Evaluation roots: A wider perspective of theorists' views and influences* (S. 3-11). Thousand Oaks, CA: Sage.
- Alkin, M. C., & Christie, C. A. (2004). An evaluation theory tree. In M. C. Alkin (Hrsg.), *Evaluation roots: Tracing theorists' views and influences* (S. 12-65). Thousand Oaks, CA: Sage.
- 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
- Blatter, S. (2014). No need to fear video. *The FIFA Weekly*, 23.
- 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.
- Bryk, A. S. (1983). *Stakeholder-based evaluation*. San Francisco, CA: Jossey-Bass.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Hrsg.), *Handbook of research on teaching* (S. 171-246). Chicago, IL: Rand McNally.
- 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.
- Christie, C., & Alkin, M. (2013). An evaluation theory tree. In M. C. Alkin (Hrsg.), *Evaluation roots: A wider perspective of theorists' views and influences* (S.11-58). Thousand Oaks, CA: Sage.
- 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.
- Cook, T. D., & Shadish, W. R. (1986). Program evaluation: The worldly science. *Annual review of psychology*, 37(1), 193-232.
- Cousins, J. B., & Earl, L. M. (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer Press.
- Cronbach, L. J. (1963). Course Improvement Through Evaluation. *Teachers College Record*, 64, 672-683.

- Cronbach, L. J., & Shapiro, K. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass.
- Cross, R. (2014). The footprint of a tennis ball. *Sports Engineering*, 17(4), 239-247.
- D'Agostino, F. (1995). The Ethos of the Game. In W. J. Moran & K. V. Meier (Hrsg.), *Philosophic Inquiry in Sport* (S. 48-49). Champaign, IL: Human Kinetics.
- Dohmen, T., & Sauermann, J. (2015). Referee Bias. *Journal of Economic Surveys*, 30(4), 679-695. doi:10.1111/joes.12106
- Duden online. (2016a). Stichwort: Innovation. Zugriff unter <http://www.duden.de/rechtschreibung/Innovation#Bedeutung1a>
- Duden online. (2016b). Stichwort: Technik. Zugriff unter <http://www.duden.de/rechtschreibung/Technik#Bedeutung1>
- Dyer, B. (2015). The controversy of sports technology: a systematic review. *SpringerPlus*, 4(1), 1-12. doi:10.1186/s40064-015-1331-x
- Fetterman, D. M. (1998). Empowerment Evaluation and Accreditation in Higher Education. In E. Chelmsky & W. R. Shadish (Hrsg.), *Evaluation for the 21st Century* (S. 381-395). Thousand Oaks, CA: Sage.
- FIFA. (2014). FIFA Quality Programme for Goal-Line Technology. TESTING MANUAL 2014. Version 2.0. Zugriff unter <http://quality.fifa.com/contentassets/148295d585194b1fb6030bbcd890b652/testing-manual-2014---fifa-quality-programme-for-goal-line-technology-.pdf>
- Ford, G. G., Goodwin, F., & Richardson, J. W. (1996). Perceptual factors affecting the accuracy of ball and strike judgments from the traditional American League and National League umpiring perspectives. *International Journal of Sport Psychology*, 27(1), 50-58.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Hansen, G., & Lames, M. (2001). Die Qualitative Spielbeobachtung. Eine Beobachtungsvariante zur Trainings-und Wettkampfsteuerung im Spitzensport. *Leistungssport*, 31(1), 63-71.
- 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. *J Sports Sci*, 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. *J Sports Sci*, 25(9), 991-994. doi:10.1080/02640410601150488
- Hohmann, A., Lames, M., & Letzelter, M. (2014). *Einführung in die Trainingswissenschaft* (6. unveränderte Auflage). Wiebelsheim: Limpert Verlag.
- ITF. (2015). Automated Line-Calling Systems: ITF Evaluation. Zugriff unter <http://www.itftennis.com/media/248075/248075.pdf>
- 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. & Lames, M. (2017). Scientific approaches to technological officiating aids in game sports. *Current Issues in Sport Science*, 2:001, 1-10. doi: 10.15203/CISS_2017.001

- Kolbinger, O. & Link, D. (2016). The use of vanishing spray reduces the extent of rule violations in soccer. *SpringerPlus*, 5:1572, 1-7. doi: 10.1186/s40064-016-3274-2
- 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. In Cabri, J., Barreiros, J. & Pezarat-Correia, P. (Hrsg.), *Sports Science Research and Technology Support* (S. 107 - 118). Schweiz: Springer International Publishing. doi: 10.1007/978-3-319-25249-0_8
- Lames, M. (1991). *Leistungsdiagnostik durch Computersimulation*. Frankfurt am Main: Verlag Harri Deutsch.
- Lames, M. (1994). *Systematische Spielbeobachtung*. Münster: Philippka.
- Lames, M. (2000). Evaluationsforschung in historischer und zeitgenössischer Betrachtung. In V. Bank & M. Lames (Hrsg.), *Über Evaluation* (2. unveränderte Auflage, S. 3-47): bajOsch - Hein Verlag für Berufs- und Wirtschaftspädagogik.
- Lames, M. (2006). *Glücksspiel Fußball – Zufallseinflüsse beim Zustandekommen von Toren*. Paper presented at the Elf Freunde sollt ihr sein!? 38. Jahrestagung der Arbeitsgemeinschaft Sportpsychologie.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J. & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Med*, 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
- Mayring, P. (2008). *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim, Germany: Beltz.
- Mertens, D. M., & Wilson, A. T. (2012). *Program Evaluation Theory and Practice: A Comprehensive Guide*. New York: The Guilford Press.
- Mills, B. M. (2014). Social pressure at the plate: Inequality aversion, status, and mere exposure. *Managerial and Decision Economics*, 35(6), 387-403.
- 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. *Marq Sports L Rev*, 15, 13-27.
- National Football League. (2016). HISTORY OF INSTANT REPLAY. Zugriff unter <http://operations.nfl.com/the-game/history-of-instant-replay/>
- Oudejans, R. R., Bakker, F. C., & Beek, P. J. (2007). Helsen, Gilis and Weston (2006) err in testing the optical error hypothesis. *J Sports Sci*, 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.

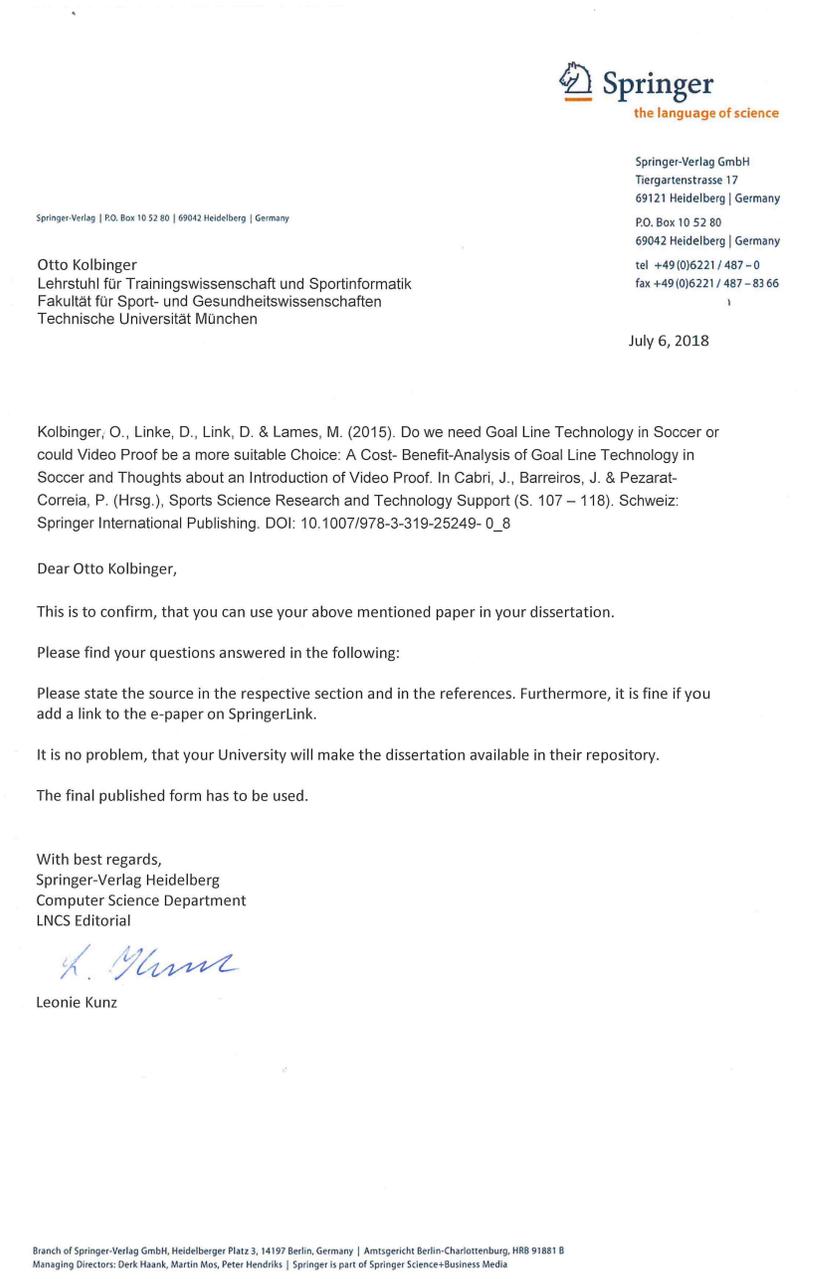
- Oudejans, R. R., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbruckner, M., & Beek, P. J. (2000). Errors in judging 'offside' in football. *Nature*, 404(6773), 33. doi:10.1038/35003639
- Patton, M. Q. (1997). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- Pope, B. R., & Pope, N. G. (2015). Own-Nationality Bias: Evidence from Uefa Champions League Football Referees. *Economic Inquiry*, 53(2), 1292-1304. doi:10.1111/ecin.12180
- Reid, H. L. (2012). *Introduction to the Philosophy of Sport*. Plymouth, United Kingdom: Rowman & Littlefield Publishers, Inc.
- Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach*. Newbury Park, CA: Sage Publications.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2004). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.
- Royce, R. (2012). Refereeing and technology—Reflections on Collins' proposals. *Journal of the Philosophy of Sport*, 39(1), 53-64.
- Scheller, D. (2016). Team- und Schiedsrichterspezifisches Verhalten bei Freistößen der Bundesliga (Unveröffentlichte Bachelorarbeit). Technische Universität München, München, Deutschland.
- Schimank, U. (2005). Die Entwicklung des Sports zum gesellschaftlichen Teilsystem. In U. Schimank (Hrsg.), *Differenzierung und Integration der modernen Gesellschaft* (S. 111-142). Wiesbaden: Springer.
- Schmidt, G., & Bloch, M. (1980). Warum gibt es ein Schiedsrichterproblem? Bericht über eine Fallstudie. *Leistungssport*, 10(4), 290-299.
- Scriven, M. (1980). *The logic of evaluation*. Inverness, CA: Edgepress.
- Shadish, W. R. (1984). Policy research: Lessons from the implementation of deinstitutionalization. *American Psychologist*, 39(7), 725-738.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.
- Siegle, M., & Lames, M. (2012). Game interruptions in elite soccer. *J Sports Sci*, 30(7), 619-624. doi:10.1080/02640414.2012.667877
- Simon, R. (2007). Deserving to be lucky: Reflections on the role of luck and desert in sports. *Journal of the Philosophy of Sport*, 34(1), 13-25.
- Stadtman, G., Weigand, J., & Wissmann, M. (2004). Borussia Dortmund in der Kritik ‚Wichtig ist nicht nur auf dem Platz‘. *Sport und Gesellschaft*, 1(2), 145-164.
- 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.
- Stufflebeam, D. L. (2001). Interdisciplinary Ph. D. programming in evaluation. *The American Journal of Evaluation*, 22(3), 445-455.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.
- Suchman, E. A. (1967). *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*. New York: Russell Sage Foundation.
- 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.

- Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy analysis*, 3(4), 531-545.
- Weiss, C. H. (1982). Policy research in the context of diffuse decision making. *The Journal of Higher Education*, 53(6), 619-639.
- Weiss, C. H. (1986). The Circuitry of Enlightenment: Diffusion of Social Science Research to Policymakers. *Scientific Communication*, 8(2), 274-281.
doi:doi:10.1177/107554708600800213
- Weiss, C. H. (1993). Where Politics and Evaluation Research Meet. *American Journal of Evaluation*, 14(1), 93-106.
- Wholey, J. S. (1983). *Evaluation and effective public management*. Boston, MA: Little, Brown.

Anhang A Publikationen

A.1 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

A.1.1 Erlaubnis zur Einbindung



A.1.2 Originalveröffentlichung

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

Otto Kolbinger^(✉), Daniel Linke, Daniel Link, and Martin Lames

Department for Sport and Health Sciences, TU München, Munich, Germany
{otto.kolbinger,daniel.linke,
daniel.link,martin.lames}@tum.de

Abstract. The aim of this study was to investigate the necessity of goal line technology for top level soccer and compare it to video proof. 1167 games of the 1st and 2nd German Bundesliga were screened for critical goal line decisions as well as other critical goal decisions including penalty calls. 16.8 critical goal line decisions were found per season, of which 76.6 % could directly be resolved by tv-review. On average, 5.0 and 2.8 cases were found respectively in the 1st and 2nd Bundesliga that could justify goal line technology. Furthermore, just 5.0 % of all critical calls concerned goal-line decisions, while 84.3 % involve offside, representing the main reason for critical goal calls. Based on these findings, in terms of a cost-benefit-relation, a video proof could be more suitable than goal line technology. Consequently, we discuss the obstacles and preconditions for the introduction of a video proof in soccer.

Keywords: Goal line technology · Video proof · Soccer

1 Introduction

For scoring a goal in Soccer the whole of the ball has to pass the goal line, according to the official laws of the game [1] of the Fédération Internationale de Football Association (FIFA). History shows that this easy fact could be arguable in seldom cases. Perhaps the oldest instance was the fifth goal of the 1966 World Cup final between England and Germany. With the game tied in extra time, a shot of the English striker Geoff Hurst hit the underside of the crossbar, bounced down towards the line and hit the ground near the goal line. The referee and his linesman respectively decided that the ball had passed the goal line completely. Whether this call was right (or wrong) was solved only 1996, when a study of Reid and Zisserman [2] concluded that the ball didn't cross the line entirely by at least six centimeters. In Germany, this famous incident led to the expression "Wembley Goal" for disputed goal calls, after the ball hitting the underside of the crossbar.

© Springer International Publishing Switzerland 2015
J. Cabri et al. (Eds.): icSPORTS 2014, CCIS 556, pp. 107–118, 2015.
DOI: 10.1007/978-3-319-25249-0_8

In recent years technologies became available which are able to track the position of the ball and therefore can decide if the ball crossed the goal line. Thus, the International Football Association Board (IFAB) decided to permit goal line technology in July 2012. To get accredited as FIFA licensed goal line technology providers, a company has to pass a series of certain tests, including laboratory tests as well as field tests [3]. The two main requirements are an accuracy of 1.5 cm (3.0 cm before 2014) and the indication of whether a goal has been scored must be communicated to the referee in less than one second.

So far, the FIFA has officially accredited four systems, representing two different technologies. The camera based systems use several high speed cameras which are located around the pitch, continuously capturing the ball's position in three dimensions when it is close to the goal. The Hawk-Eye system is already used e.g. in the English Premier League, GoalControl was the official provider for the 2014 FIFA World Cup™. For these systems no modifications or adaptations of the ball are necessary. Magnetic-field-based systems need the implication of receptors in the ball, for enabling the system to calculate the position of the ball. The accredited systems using this technology were run by the Fraunhofer Institute for Integrated Circuits and by ChyronHego.

The study was commissioned and funded by Deutsche Fußball Liga (DFL, trans. German Football League) [5] in the run up of a general meeting on March 24th 2014. On a vote at this meeting, delegates of the clubs of German 1st and 2nd Bundesliga decided to not introduce goal line technology in these two leagues. The discussion about the introduction of a goal line technology didn't stop in the aftermath of this meeting. Thus, in a second vote, at the general meeting on December 4th 2014, the delegates approved the use of goal line technology at least in the 1st Bundesliga.

Scientific research concerning GLT so far mostly centers on technological issues [4]. Up to now, there is no study investigating in how far the implementation of this new technology is necessary in terms of a cost-benefit-relation. For example the costs for the installation and maintenances of Hawk-Eye systems can be set for round about 2.30 million euro per season (for a league with 18 stadiums). Therefore this study examines the frequency of critical goal-line situations in order to put these costs in perspective with these frequency. In addition, it was obtained how many of these critical incidents could be solved with videos provided by the usual TV-footage. Since the very majority of professional soccer matches is recorded by several cameras, this method should illustrate a procedure of checking critical goal-line decisions without (or just with low) extra expenses.

The method of scoring, according to the official laws of the game, also requires "that no infringement of the Laws of the Game has been committed previously by the team scoring the goal" [1]. So the number of critical goal-line situations was compared to other sources of disputed decisions, i.e. offside, handling the ball, foul and penalty calls. These situations were also checked with the TV-footage. So far, the IFAB didn't approve any kind of video replay. The concerns of changing the nature of the game outweigh the possible advantages, but the topic shall still be discussed at the IFAB's Annual General Meetings. Thus, this study also investigated the necessity and enlightenment rate of a video replay.

Due to the nature of soccer there are also other issues concerning the introduction of new technologies. The use of video proof would need rules for the video proof itself as well as its embedding in the current laws of the game. Based on sports that already use video proof we discuss its possible regulations and the necessary (rule) preconditions for its introduction in soccer.

2 Methods

2.1 Database

The sample contained matches from day 18 of the 2011/2012 season to day 20 of the 2013/2014 season of 1st and 2nd German Bundesliga. The DFL provided footage of 1167 of these matches via their database Deutsches Fußball Archiv. Even since this footage is rather an offer for the media than for performance analysis or scientific researches, it included all available camera angles as well as slow-motion replays. The number of available camera angles varied from seven (e.g. low priority 2nd Bundesliga matchup) to 22 cameras (e.g. nationwide broadcasted matches). The results section will include numbers per match day (or round) and per season which were extrapolated based on the sample of the 1167 matches.

2.2 Operationalization

Critical Goal Line Decisions. It was assumed that the frequency of goal line technology uses would be determined by the frequency of scenes that arouse the suspicion that a goal line decision could be critical. This includes situations where a goal was called as well as when the referee decided that the ball didn't cross the goal line. This led to the following two definitions:

1. A goal is called, but the ball doesn't touch the net
2. The ball approaches the goal line by less than one diameter, before or after it gets deflected, with players of the offensive team complaining because of the goal-line decision

The second definition includes a positional precondition as well as a certain behavior of the offensive players. There are several scenes in a soccer match with the ball approaching the goal line by less than one diameter but it is obvious for all the involved people that no goal was scored. So the second definition includes protests of players of the offensive team.

Critical Goal Decisions. To quantify the rate of critical goal line decisions of all critical goal calls, critical goals due to other sources were obtained. A goal was called critical by fulfilling at least one of the following three rule infringements by the scoring team:

- Offside: Involved player (of the offensive team) is offside in a certain period of time, from 0.25 s before until 0.25 s after the last touch of a teammate
- Foul: Defensive player loses body control due to a contact with an involved offensive player
- Handling the ball: Involved offensive player touches the ball with his hand, arm or shoulder

Besides these kinds of critical goals all penalties were captured.

2.3 Video Replay

All critical goal line decisions as well as critical calls were evaluated and double-checked by two experts using the provided TV-footage. After the review the call was rated as “correct”, “false” or “not to decide clearly”. When a critical goal line decision was “not to decide clearly” by video replay, it was concluded that this scene could only be solved by using goal line technology. Penalty calls were just rated as “doubtless” or “doubtful”, due to influences like the interpretation of the laws of the game by the referee. Thereby “doubtless” included just scenes in which the referee’s call was obviously right, assuming that the other calls could arouse suspicion that the penalty call was disputable. Due to the nature of this study, as it was commissioned by the DFL, all these evaluations had to be double-checked by both experts.

2.4 Reliability

The reliability of the observation method was investigated with an inter-rater reliability study. Therefore the matches of three randomly selected game days (i.e. 27 matches) were investigated independently by both experts. The reliability was measured using Cohen’s Kappa for the identification of critical goals and the source of critical goal decisions. As mentioned above, it wasn’t possible to check the inter-rater-agreement of the evaluations, since the evaluations had to be double-checked anyway. Both variables showed a very good strength of agreement (according to the suggested limits of Altmann [5]), with k-values of 0.95 and 1.0 for the identification and the source respectively. Even by suggesting more strict limits for easier variables, or almost obvious variables like the source of critical goal decisions, the classification of the inter-rater-agreement can be kept up.

3 Results

3.1 Critical Goal Line Decisions

For the investigated 613 matches of the 1st German Bundesliga 39 scenes were identified that arouse the suspicion that a goal line decision could be critical. As shown in Table 1, that means that per game day 0.58 and per season 19.6 critical goal line

decisions appear. In the 2nd German Bundesliga these rates were lower. 25 critical goal line decisions were found in 554 matches, meaning 0.41 per round and 13.9 per season.

Table 1. Scenes in which goal line technology would be mandatory and critical goal line decisions (in brackets) of the German Bundesliga and 2nd Bundesliga.

	Bundesliga		2 nd Bundesliga	
total	10	(39)	5	(25)
per match	0.02	(0.06)	0.01	(0.05)
per round	0.15	(0.58)	0.08	(0.41)
per season	5.02	(19.6)	2.78	(13.9)

As mentioned above, critical goal line decisions which were rated as “not to decide clearly” by video reviews were classified as scenes that could only have been solved by goal line technology. For the 1st German Bundesliga 5.02 such scenes were found per seasons, meaning 0.15 per round and 0.02 per match. For the 2nd German Bundesliga these numbers are lower, too. 2.78 scenes per season had been identified, meaning 0.08 per game day and 0.01 per match. Regarding the detection rate of critical goal line decisions by video proof, the results showed that 76.6 % of these situations could be solved with the available TV-footage.

3.2 Critical Goal Decisions

Prevalence. The rates of critical goals are listed in Table 2. The table includes critical goals due to critical goal line calls as well as possible infringements of the law of the game by the offensive team. The results show that 4.72 % and 5.59 % of all critical goals were critical because of goal line calls in the 1st and 2nd Bundesliga respectively. Offside calls were the major source for critical goal decisions, with 85.0 % in the 1st Bundesliga and 83.2 % in the 2nd Bundesliga. Foul (5.91 %) and handling the ball (4.33 %) are in the same range as critical goal line decisions in the 1st Bundesliga. In the 2nd Bundesliga of the 554 screened matches of this league, critical goals due to possible fouls (9.79 %) have a slightly bigger rate, while only two goals in total were critical due to handling.

Table 2. Critical goal decisions in the 1st and 2nd Bundesliga by source of doubt and their rate of all the critical goals per league in percent (in brackets).

	Offside	Handling	Foul	Goal line
1 st Bundesliga	216 (85.0)	11 (4.33)	15 (5.91)	12 (4.72)
2 nd Bundesliga	119 (83.2)	2 (1.40)	14 (9.79)	8 (5.59)

Penalties. In total 310 penalties were called in the investigated 1167 matches, or rather 0.27 per game in the 1st Bundesliga and 0.26 per game in the 2nd Bundesliga. Fouls were a more frequent rule infringement leading to penalties with 81.7 % (1st Bundesliga) and 83.6 % (2nd Bundesliga).

20.6 % of all penalty calls (foul: 19.6 %, handling: 25.9 %) were rated as “doubtful” by TV-review and therefore as critical goal calls. Regarding the success rates of 81.7 % (1st Bundesliga) and 74.0 % (2nd Bundesliga), this means that there were 0.39 critical goals per round due to their preceding penalty calls.

Video Replay. After thorough video inspections, all critical goal calls were rated as “correct”, “false” or “not to decide clearly”. As mentioned above, the detection rate for critical goal line decisions is 76.6 %. Figure 1 illustrates the detection rates for all sources of critical goals as well as the rate of all investigated critical decisions. The highest values were reached for offside decisions with 89.3 % and fouls with 86.2 %. The lowest detection rate showed handling calls (69.2 %). In total 86.6 % of all the investigated critical calls could be solved with video-replays based on the available TV-footage.

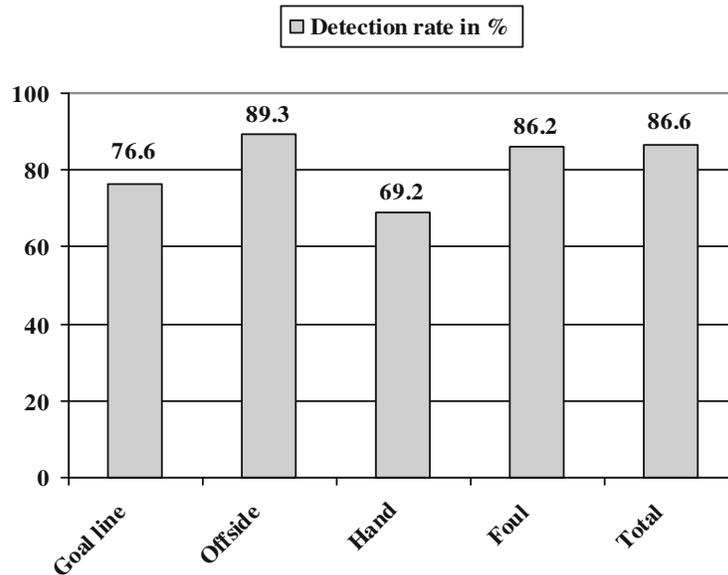


Fig. 1. Detection rate of critical goal line calls (also including scenes where no goal was called) and critical goal calls sorted by category and in total.

4 Discussion

4.1 Prevalence of Critical Goal Line Decisions and Other Critical Goal Calls

The results of this investigation show that merely 5.0 % of all critical goal decisions concern the question whether the ball completely passed the goal line or not. Furthermore, the frequency of these decisions is very low with 0.02 per match in the 1st German Bundesliga and 0.01 in the 2nd Bundesliga respectively. This means, using Bernoulli distributions, that per stadium per season in the 1st Bundesliga the probability of having at least one case were the use of goal line technology would be non-negotiable is just 0.24. This probability is even lower in the 2nd Bundesliga (0.14). Figure 2 also shows that these probabilities are over 95 % only after eleven full seasons in the 1st Bundesliga and 20 seasons in the 2nd Bundesliga. Regarding the 135.00 € per stadium per season, there is a potential chance of spending this money without getting any kind of benefit.

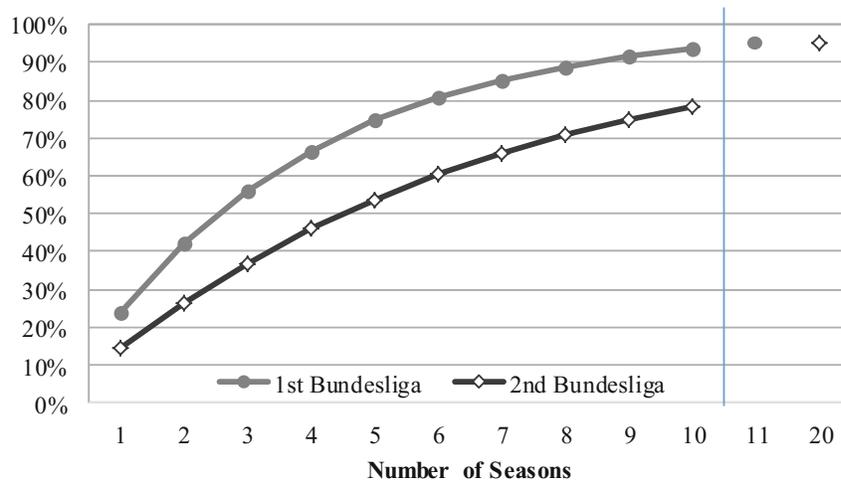


Fig. 2. Bernoulli probabilities for at least one non-negotiable use of goal line technology in one stadium for 1st and 2nd Bundesliga from one to ten seasons and (after the vertical line) the number of seasons until this probability reaches 95 %

Other sources for critical goal calls outnumber critical goal calls by far. Especially offside, which represents the major reason for disputable goal decisions with approximately 85 %. In the 1st Bundesliga appeared one critical goal – excluding critical goal line decisions – at least every three games (0.39 per game), and in the 2nd Bundesliga almost in one game out of four. These numbers exclude critical goals because of disputable penalty calls, which adds another 0.39 such goals per round.

All these critical goal calls couldn't be solved with goal line technology. Video proof could be a possible solution to solve them. Since every match in the two German Bundesligen is recorded by at least seven (up to 22) cameras this technology would have a much better cost-benefit-relation. And this technology could solve such calls as well as critical goal line decisions.

4.2 Rate of Enlightenment

86.6 % of all the investigated critical calls could be clearly solved with the provided TV-footage, including the solved 76.6 % of critical goal lines decisions. The highest detection rate was reached for critical offside decisions (89.3 %), followed by disputable foul calls and no-calls respectively (86.2 %). The rate of enlightenment for handling the ball was the lowest with 69.2 %. This could be a consequence of the rule, which interpretation is partly subjective based on vague definitions in the official laws of the game [1]. For example, the referee has to decide whether the ball moves towards the hand or the hand moves towards the ball. Usually both move and due to the nature of the game move towards each other not exactly seldom. The referee also has to take in consideration if the distance between the handling player and the ball was far enough so that it isn't an "unexpected ball" for the player anymore. Furthermore he has to assess the position of the hand. Taking all this information into account it can be stated that the rules for handling the ball are partly subjective.

We want to point out that the used video-footage was not recorded specifically for a use as video proof. Furthermore, it wasn't possible to stream the videos in high-definition-quality. We would expect even higher detection rates, especially for objective decisions like offside, with better or rather spoken more specific video-footage (which is partly already in use). For example, in some matches there were cameras on level with the goal line which made the solving of critical goal line decisions in this matches much easier. In addition, tools could be created to support the decision-making with video-proofs, e.g. an offside-plane. All this new footage or tools could also be used in the TV-broadcasts and other media, which would improve the overall cost-benefit ratio.

4.3 Obstacles and Pre-considerations for the Introduction of Video-Replay

Obstacles. There are generally two main objections to the introduction of a video proof in soccer. One is the nostalgic point of view, that disputable calls are a part of the game and could be a reason for the popularity of soccer. Of course, there are instances of games that were especially famous because of critical goal calls, e.g. the above-mentioned World Cup final of 1966. But the introduction of goal line technology shows that this traditional mindset is already given up by the FIFA and the IFAB.

The other main doubt is the change of the nature of the game. It is supposed that the introduction of a video replay would lead to more and/or longer interruptions of the game. For verifying this thesis two different topics have to be taken into account:

the possible duration of a video replay and the already given pattern of interruptions in soccer matches. On average there are well over 100 interruptions in elite men soccer matches and the total duration of all interruptions is 30 to 40 min, meaning approximately a third of the complete match time [7, 8]. The interruptions after goals and before penalties are on average already over 50 s, which could be enough for video replays. Free kicks merely take twenty seconds but last significant longer in the offensive areas of the pitch [8].

A current pilot scheme of the Koninklijke Nederlandse Voetbal Bond (KNVB, tans. Royal Dutch Soccer Association) suggests that video replay interventions should last 15 s at maximum [9]. But this project focuses only on the avoidance of obvious mistakes and there are no results published yet. However, there are no empirical values showing how long video replays could last in soccer. Consequently, the experiences of other sports, that already use video proof, should be taken as reference. The Major League Baseball expanded their replay rules in beginning of the 2014 season, so for example tag plays and force plays became reviewable [10]. As different as soccer and baseball are, replays of these two types of call could deliver valuable hints for the duration of replays in soccer. Both types of decisions are based on very objective rules. For critical force play calls the referee has to check whether the runner touches the base before the defensive player (touching the same base) catches the ball or not. For tag plays it has to be verified if the runner was touched with the ball (or the glove that holds the ball).

The official time data of all such 823 replays from 1st of May 2014 until the end of the regular season 2014 were investigated. The information used here was obtained free of charge from and is copyrighted by Retrosheet [10]. The average duration was 01:43 min with the median at 01:35. As shown in Fig. 3, none of the replays was

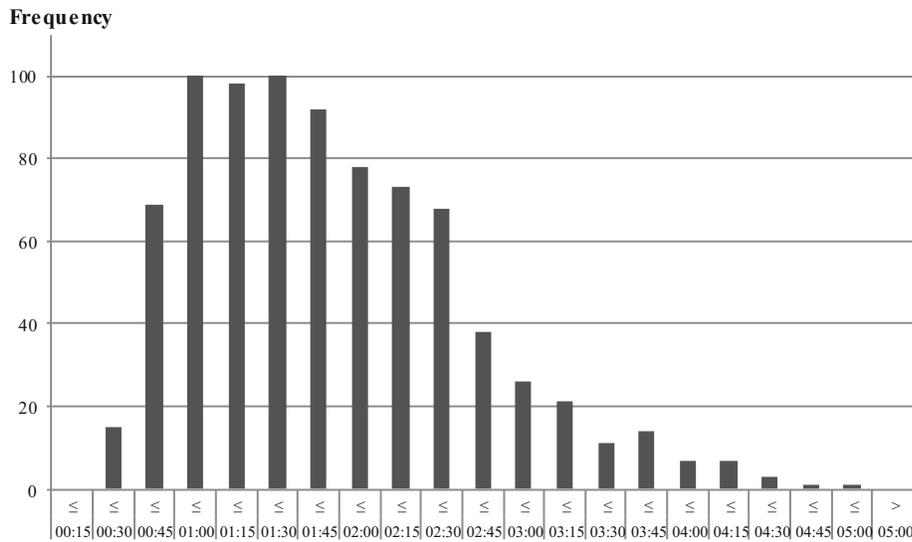


Fig. 3. Frequencies of the durations of force and tag plays in the MLB 2014 regular season, excluding games in April. The range of values are set as 15 s.

shorter than 15 s and only 1.82 % were shorter than 30 s. Half of the replays (from percentile rank 25 to 75) lasted between 1:04 and 02:14, while there were also durations of up to 4:46 detected.

Even considering the different circumstances in baseball, where the time of interruptions isn't a factor (yet), the set 15 s of the KNVB seem to be unrealistic. On the other hand, regarding the average durations of interruptions for goals and penalties, it at least doesn't seem impossible to evaluate the specific calls during the period of these interruptions.

Pre-considerations. For an introduction of video replays new regulations for its use would be necessary. Starting with the consideration, based on the thoughts above, which calls should be reviewable and when. In this context, "when" means in which state of the game. When the referee evaluates a critical situation he just interrupts the game if he decides that there was an infringement of the laws of the game (or the ball passed the goal line for critical goal line decisions). In this case a replay could be used without large-scale rule changes, but how could calls be reviewed where the referee doesn't interrupt the game? This question does not occur in sports that are interrupted after every play like baseball or American football. Hockey is also an invasion game (like soccer) and already uses video replays. Unfortunately, this problem isn't satisfyingly solved in the official hockey regulations [12]. It is determined that the referee can stop the game, but not the restart-scenario for video replays that result in no-calls, i.e. situations in which the game wouldn't be interrupted. Possibilities could be a drop ball or a free kick for the team that was in possession before the game was interrupted for the replay. The main problem will be to re-establish a game situation, when one team had an advantage at the moment of the interruption, at worst some kind of scoring opportunity.

This problem could be avoided by the selection of reviewable calls. In Hockey the video umpire should especially (but explicitly not exclusive) assist for goals, penalty strokes and penalty corners. A similar use in soccer would include goals, penalties and eventually free kicks in highly promising positions. In all these situations there would be interruptions of decent intervals as shown above. Possible scoring opportunities that were taken away by the referee due to wrong calls by the referee would remain unresolved. To overcome this source of disputes it could be conceivable encouraging referees to just interrupt the game in such situations if there is clear evidence of a rule infringement. This would be an indirect intervention in the laws of the game and would especially change the nature of refereeing.

Another fact that has to be considered, directly connected with the aforementioned facts, is the regulation of the number of replays per game. American football [13], baseball and hockey each run replay systems in which the teams have a certain number of so-called challenges. In addition, the referees or umpires can ask for video replays as often as (they think it is) necessary. In American football and baseball only the coaches are entitled to challenge, while in hockey any player on the pitch can request for video proof. A good fit for soccer could be the implementation of parts of the replay system in American football. In so called booth-reviews all scoring plays are automatically reviewed. This could be implemented for soccer without direct changes of the nature of

the game, even by including penalty calls. Of course, this could still lead to indirect changes, for example due to adapted refereeing styles.

Furthermore, there are different possibilities for the process of video replays. In the MLB umpires in the league's headquarter evaluate the calls and give the requesting umpires feedback. In hockey the referees refer decisions to video umpires that sit in a kind of broadcast unit in the direct proximity of the pitch or stadium respectively. The project of the KNVB works with this process. It might also be conceivable to give the forth official the task of doing the video replays.

5 Conclusion

The results of this study show that there are several disputable calls in soccer matches and previous studies proofed that limits of the human perception are a reason for it. As one source of error, for example, the flash-lag effect was identified uniformly [14, 15] while another source is still controversially discussed (optical error hypothesis) [16–18]. The IFAB and the FIFA already accommodated this issue and approved the use of goal line technologies. The results of this study showed that this technology has a disputable cost-benefit-relation as only 5.0 % of all critical goals concern goal line decisions.

The Introduction of a video proof seems to be a more suitable choice as 86.6 % of all critical goal calls, including critical goal line decisions, could be resolved by using only the usual TV-footage. Based on knowledge about interruptions in soccer and experiences with video proofs in other sports, we could show that the inclusion of a video replay could be possible without large-scaled changes of the nature of the game. The introduction of a booth-review, as in American football, reviewing all scoring plays and penalty calls, might be promising.

References

1. Fédération Internationale de Football Association. http://www.fifa.com/mm/document/footballdevelopment/refereeing/81/42/36/log2013en_neutral.pdf
2. Reid, I., Zisserman, A.: Goal-directed video metrology. In: Cipolla, R., Buxton, B. (eds.) *Computer Vision – ECCV'96*. LNCS, vol. 2, pp. 647–658. Springer, Heidelberg (1996)
3. Fédération Internationale de Football Association. http://www.fifa.com/mm/document/affederation/administration/02/35/96/73/fifa-background-paper_glt_june2014_en_neutral.pdf
4. Spagnolo, P., Leo, M., Mazzeo, P.L., Nitti, M., Stella, E., Distante, A.: Non-Invasive soccer goal line technology: a real case study. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1011–1018. Conference Publishing Services, Los Alamitos (2013)
5. Kolbinger, O., Linke, D., Link, D., Lames, M.: *Zur Notwendigkeit von Torlinientechnologie in den Fußball-Bundesligen*. Deutsche Fußball Liga GmbH, Frankfurt (2014)
6. Altman, D.G.: *Practical Statistics for Medical Research*. Chapman and Hall, London (1991)

7. Augste, C., Lames, M.: Differenzierte betrachtung von taktischem verhalten und belastungsstrukturen auf der basis von spielunterbrechungen im fussball. In: Woll, A., Klöckner, W., Reichmann, M., Schlag, M. (eds.), Sportspielkulturen erfolgreich gestalten, pp. 113–116. Czwalina, Hamburg (2008)
8. Siegle, M., Lames, M.: Game interruptions in elite soccer. *J. Sports Sci.* **30**(7), 619–624 (2012)
9. Koninklijke Nederlandse Voetbal Bond. <http://english.knvb.nl/content/485/refereeing-20>
10. Major League Baseball. http://mlb.mlb.com/mlb/official_info/official_rules/replay_review.jsp
11. Retrosheet. <http://www.retrosheet.org/replays.htm>
12. International Hockey Federation. <http://www.fieldhockey.ca/files/Officials%20Resources/FIH%20Tournament%20Regulations%20Outdoor%20-%20January%202013.pdf>
13. National Football League. <https://www.nfl.info/download/2012mediaguides/%20nfl%20rule%20book.pdf>
14. Gillis, B., Helsen, W., Catteeuw, P., Van Roie, E., Wagemans, J.: Interpretation and application of the offside law by expert assistant referees: perception of spatial positions in complex dynamic events on and off the field. *J. Sports Sci.* **27**(6), 551–563 (2009)
15. Helsen, W., Gillis, B., Weston, M.: Errors in judging “offside” in football: test of the optical error versus the perceptual flash-lag hypothesis. *J. Sports Sci.* **24**(5), 521–528 (2006)
16. Oudejans, R.R.D., Verheijen, R., Bakker, F.C., Gerrits, J.C., Steinbrückner, M., Beek, P.J.: Errors in judging “offside” in football. *Nature* **404**, 33 (2000)
17. Oudejans, R.R.D., Bakker, F.C., Beek, P.J.: Helsen, gillis and weston (2006) err in testing the optical error hypothesis. *J. Sport Sci.* **25**(9), 987–990 (2007)
18. Helsen, W., Gillis, B., Weston, M.: Helsen, gillis and weston (2006) do not err in questioning the optical error hypothesis as the only major account for explaining offside decision-making errors. *J. Sport Sci.* **25**(9), 991–994 (2007)

A.2 The use of vanishing spray reduces the extent of rule violations in soccer

A.2.1 Erlaubnis zur Einbindung

Friday, July 6, 2018 at 3:15:18 PM Central European Summer Time

Betreff: 00888435 RE: Questions about copyright

Datum: Freitag, 15. Juni 2018 um 12:31:26 Mitteleuropäische Sommerzeit

Von: [REDACTED]

An: Kolbinger, Otto

SPRINGER NATURE

Dear Dr Kolbinger,

Thank you for contacting Springer Nature.

The open access articles published in SpringerOpen's journals are made available under the Creative Commons Attribution (CC-BY) license, which means they are accessible online without any restrictions and can be re-used in any way, subject only to proper attribution (which, in an academic context, usually means citation).

The re-use rights enshrined in our license agreement (<http://www.springeropen.com/get-published/copyright/copyright-and-license-agreement>) include the right for anyone to produce printed copies themselves, without formal permission or payment of permission fees. As a courtesy, however, anyone wishing to reproduce large quantities of an open access article (250+) should inform the copyright holder and we suggest a contribution in support of open access publication.

If you have any questions, please do not hesitate to contact me.

With kind regards,

[REDACTED]
Global Open Research Support Executive
Global Open Research Support

Springer Nature

T +44 (0)203 192 2009

www.springernature.com

Springer Nature is a leading research, educational and professional publisher, providing quality content to our communities through a range of innovative platforms, products and services. Every day, around the globe, our imprints, books, journals and resources reach millions of people – helping researchers, students, teachers & professionals to discover, learn and achieve.

In the US: Springer Customer Service Center LLC, 233 Spring Street, New York, NY 10013

Registered Address: 2711 Centerville Road Wilmington, DE 19808 USA

State of Incorporation: Delaware, Reg. No. 4538065

Rest of World: Springer Customer Service Center GmbH, Tiergartenstraße 15 – 17, 69121 Heidelberg

Registered Office: Heidelberg | Amtsgericht Mannheim, HRB 336546

Managing Directors: Martin Mos, Dr. Ulrich Vest

A.2.2 Originalveröffentlichung

Kolbinger and Link *SpringerPlus* (2016)5:1572
DOI 10.1186/s40064-016-3274-2

 SpringerPlus

RESEARCH

Open Access



The use of vanishing spray reduces the extent of rule violations in soccer

Otto Kolbinger*  and Daniel Link

Abstract

Background: More and more sport associations introduce innovative devices to support referees and umpires respectively, affecting a strong need for the evaluation of these devices. This study evaluates the use of the new vanishing spray for free kicks in the German Bundesliga. In more detail, the aim of the study is to investigate if the spray reduces violations of the required minimum distance and consequently the respective punishments, if it reduces errors concerning the distance set by the referee and if it leads to a higher success rate of free kicks.

Methods: Therefore, 1833 free kicks of the 2013/2014 and 2014/2015 season of the German Bundesliga were screened using a self-designed observational system. For the statistical analysis two parallel samples were built of 299 free kicks each.

Results: The results showed no decrease of free kicks with distance violations but a significantly lower extent of these violations ($\chi^2 = 4.58$; $p < .05$). However, none of these violations were punished appropriately. Concerning the success of free kicks, no significant impact was found neither for shots nor for crosses. In addition, no influence on the distance set by the referee could be identified.

Conclusions: The main objective of the vanishing spray was basically realized, but the use didn't lead to any further positive (side) effects. Due to the lack of punishment, the authors raise concerns about the current application of the minimum distance rule.

Keywords: Vanishing spray, Evaluation research, Umpiring aids, Referees, Free kicks

Background

An increasing amount of innovative devices is introduced in several sports to support the umpires and referees respectively. In general these devices serve as another tool to ensure the legitimate outcome of sport competitions (Rock et al. 2013). The function of the tools is to compensate the limits of human perception that cause for example optical errors and flash lag effects (Oudejans et al. 2000; Helsen et al. 2006), as well as to eliminate bias that referees show towards the hosting team or players of their own race and the same country (Parsons et al. 2011; Pope and Pope 2015; Dohmen and Sauerermann 2015). The devices can be divided up into three different groups. Devices that support the referees in their

decision making, devices that replace the referees for certain decisions and devices that help them to enforce the rules of a sport. The new vanishing spray in soccer belongs to the latter category. Referees can use this spray to mark the required minimum distance of 9.15 m (10 yards) that players of the defending team have to obey before a free kick is taken. To mark the distance the referee draws a line between the ball and the goal that players of the defending team aren't allow to cross until the ball is touched by the offensive team. Its use was finally approved at the 126th annual general meeting of the International Football Association Board (IFAB) in 2012, but the spray was already introduced in several competitions across South America since 2000. The vanishing spray attracted worldwide attention with its appearance at the FIFA World Cup 2014 and was introduced in several European competitions, like the German Bundesliga, before and during the 2014/2015 season.

*Correspondence: otto.kolbinger@tum.de
Chair of Training Science and Sport Informatics, Technical University of Munich, Georg-Brauchle-Ring 60/62, 80992 Munich, Germany



© 2016 The Author(s). This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Currently, there is still a lack of evaluation research for this device, representing a common issue in sports, concerning these kinds of devices. Especially the impacts of umpiring aids are neglected. In this connection, the impact is not only the achievement of the objectives but also other (side) effects of the innovation, which can be positive as well as negative. This is quiet surprising, as investigating the merit or worth of interventions, should be a main goal due to most of the common definitions of evaluation (Scriven 1991; Stufflebeam and Shinkfield 2007). The evaluations of the respective associations focus mostly, sometimes even exclusively, on technical parameters that the devices have to fulfill. This technical aspect has the advantage that it can be investigated under laboratory conditions and it represents the most obvious precondition for the introduction of a device. Thus, taking goal line technology as an example, the FIFA established a series of technical tests that the systems have to pass to get credited as an official device, mainly focused on the accuracy of the systems and real time detection. The same applies to scholarly studies (Psiuk et al. 2014; D’Orazio et al. 2009). The FIFA as well as other soccer associations didn’t collect data about costs and benefits of the goal line technology. Kolbinger, Linke, Link and Lames (2015) showed a very low incidence of scenes that could be resolved exclusively by goal line technology. They found less than four such incidences per season per league and therefore raised concerns about the cost-benefit ratio, especially considering the costs of round about 2.4 million per year. For the use of the Hawk-Eye technology in Tennis even the standard of the technical evaluations is questioned. In two articles considering the effect of the presentation of the technology on the publics’ understanding of science, Collins and Evans (2008, 2012) denounced the test-design of the International Tennis Federation. Nevertheless some studies investigated the use and the impact respectively of this device. Mather (2008) as well as Abramitzky et al. (2012) both found that slightly under 40 % of the challenges were successful, meaning that the technology didn’t confirm the umpires’ call. Abramitzky et al. (2012) also showed that it is only used for a very low share of points, almost exclusively for balls within 100 mm of the line (Mather 2008), but that successful challenges can increase the winning probability by up to over 25 %.

No evaluation research was run for the use of the vanishing spray in soccer yet, so the aim of this study is to overcome this lack by investigating five hypotheses. In addition to those, the using patterns of the device were described. The spray was introduced to help the referee to enforce the minimum distance rule and stemming discussions about the spot of the ball or wall respectively. Thus, the first hypothesis to be tested is that the spray

leads to fewer violations (H 1.1)—or respectively a lower extent of violations (H 1.2)—of the required minimum distance by players of the defending team. According to the official laws of the game of the Deutscher Fußball Bund (DFB, trans. German Football Association), these violations should be punished with a yellow card and a repetition of the free kick. Taken this into account, it was checked if the spray affects fewer warnings and less free kicks that have to be retaken, to provide further information concerning the patterns of rule violations. Without reference of these targets, but as the innovation created more awareness of the required distance, a third hypothesis, that the spray reduces estimation errors of the set distance between the spot of the free kick and the wall, was investigated (H 2). In addition, we assumed that no violations of the required distance benefit the kicking team, resulting in a higher success rate for free kicks taken as crosses (H 3.1) as well as for those that were taken as shots (H 3.2).

Methods

Data recording

The data set consisted of all free kicks of the 2013/2014 and 2014/2015 season that were taken as shots or crosses with a distance less than 35.0 m of the goal line. By signing a contract of employment as a professional soccer player in the German Bundesliga, each players signs a statement of consent to being monitored during matches. The provided data included a match ID, the teams involved, event time, two dimensional coordinates of the spot of the free kick and whether it was taken as a shot or cross. Using a self-deigned observational system two specific trained experts collected data for further variables that are shown in Table 1. Therefore, specific videos of the free kicks were provided starting 90 s prior to the taking and ending 10 s after it.

The distance between the spot of the ball and the wall was obtained with a custom made analysis software. Using homography, this software enables the user to determine points on the field by transforming video coordinates into real world coordinates. Therefore this software requires not just positional data of a match, but also the respective specific tracking video. These data and videos respectively were only available to the authors for a subsample for which the distance set by the referee could be obtained. The respective variable |DistanceError| shows the absolute difference to the regulatory 9.15 m.

Reliability

The examination of the inter-rater-reliability showed excellent scores for most of the variables. Cohen’s Kappa reached a value of 1.00 for the use of the spray, .91 for the punishment of violations and 1.00 for the result of shots.

Table 1 The collected variables inclusive their respective categories and definitions

Variable	Categories and definition
Use of vanishing spray	<i>True or false</i>
Local category	<i>Left/right near:</i> On the sides of the penalty box <i>Left/right far:</i> On the sides of the virtually extended penalty box between 16.5 and 35 m distance of the goal line <i>Central near left/right:</i> Inside the virtually extended penalty box with not more than 26.5 m distance of the goal line (penalty box plus 10 m). Left and right are divided by a virtual line in the middle of the field drawn at right angle to the goal line <i>Central far:</i> Otherwise, but within 35 m distance to the goal line
Players in wall	Numbers of defensive players in the wall
Violation of the minimum distance	<i>True:</i> At least one player passes the referees mark with his entire foot <i>False:</i> otherwise
Massive violation of the min distance	<i>True:</i> More than one player commits a <i>violation</i> or a player reduces the distance by more than 1 m <i>False:</i> otherwise
Punishment for violations	<i>Yellow card:</i> A yellow card is awarded <i>Verbal cautions:</i> The referee corrects the players verbally <i>None:</i> No punishment
Free kick retaken	<i>True or false</i>
Success of shots	<i>OnTarget:</i> A goal is scored, the ball hits the goals border or the goalkeeper makes a save <i>Missed:</i> Ball misses the goal or is blocked by a player outside the wall <i>Wall:</i> Ball is blocked by the wall
Success of crosses	<i>Successful:</i> A player of the offensive team is able to perform a shot or pass with the first touch after the cross <i>Not successful:</i> Otherwise

The results of crosses and the identification of rule violations felt behind with .79 and .80 respectively but were still acceptable. The correlation coefficient (.98) as well as the relative observed agreement (92.6 %) for the numbers of players in the wall also showed a good agreement between the observers.

Statistical analysis

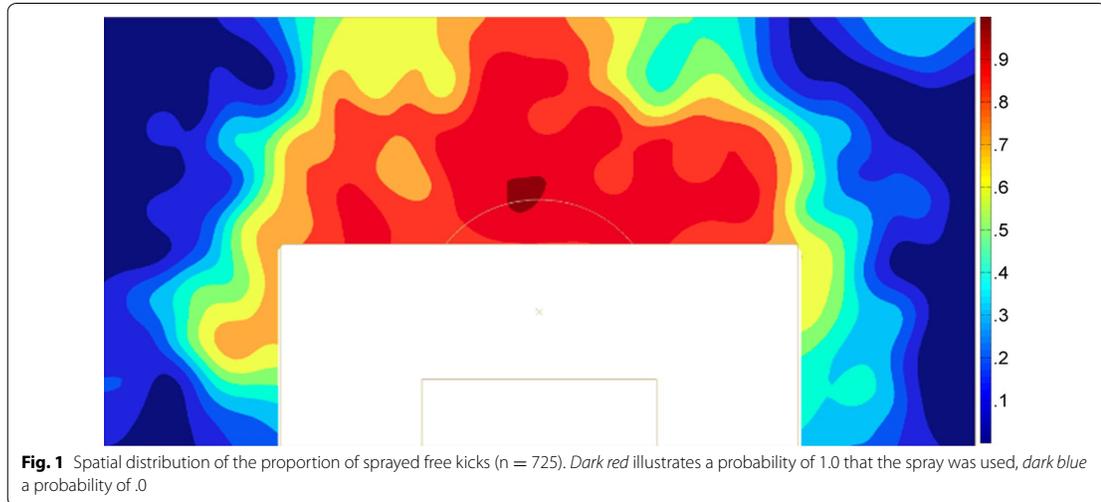
To investigate the influence of the vanishing spray, sprayed free kicks after the introduction were compared to those before, based on the idea of evaluating interventions by comparing respective variables on different time points of a program (Cronbach 1963). Thus, the mentioned variables were collected for 1833 free kicks in total. 1108 of these free kicks were taken prior to the introduction of the vanishing spray and 725 after its introduction on the 8th day of the 2014/2015 season. As the free kicks after the introduction showed no consistency considering the use of the spray, it was decided to run a parallel study design to investigate the influence of the new device. The parallelization was performed in three steps. First, the free kicks were grouped into local categories on the basis of its two-dimensional coordinates (see Table 1). The number of players in the wall served as the second criteria, representing the perceived risk of the defending team. At last, the free kicks were paired in these categories on the basis of the shortest

distance. Thus, two parallel samples were built with 299 free kicks each ($N_{\text{Spray}}/N_{\text{NoSpray}}$). 81 pairs of free kicks of these two groups represented the subsample for the investigation of the set distance by the referee.

The spatial distribution of the spray's use was visualized using the ISOPAR method (Stöckl et al. 2011). Due to the different styles of the obtained variables different statistical analysis were run, after verifying the assumptions of normality. On the one hand, a paired *t* test was calculated for the set distance between the ball and the wall. On the other hand, Chi square tests for the violations, the punishment of these violations, retakes and the success of free kicks. All statistical analysis were performed with SPSS (Version 23.0; Armonk, NY; IBM Corp.), except the respective effect sizes Cohen's *d* and Cramér's *V* that were calculated manually. The magnitudes of the effect sizes were evaluated based on the limits: .10 (small), .30 (medium) and .50 (large) for Cramér's *V* (Cramér 1946). The limits for Cohen's *d* were .20 (small), .50 (medium) and .80 (large) (Cohen 1992).

Results

For 308 of the 725 investigated free kicks after the introduction the referees decided to mark the regulatory distance with the vanishing spray. Figure 1 shows that the spray was used more likely for central free kicks, especially with decreasing goal distance. The spray was used



for all the investigated respective free kicks with six or more players in the defensive wall and for 88.9 and 89.0 % for free kicks with walls of four and five players respectively. This number decreases further for free kicks with three (70.9 %), two (34.7 %) or one player (7.8 %) participating in the wall.

The introduction of the vanishing spray showed no significant influence on violations of the required minimum distance by players of the defending team, as the share of free kicks with violations remains on a similar level (OR = 1.02; 95 % CI .69–1.49). However, the share of massive violations decreases significantly by 6 %, representing a trivial effect size but an odds ratio of .60 (95 % CI .36–.99). Despite these violations of the required minimum distance, none of the free kicks of the sample was retaken and no yellow cards were awarded for this reason. Six verbal cautions for such violations were recorded for the treatment group, showing a significant increase with small effect size compared to the zero verbal cautions of the control group ($\chi^2 = 6.06$; $p < .02$; $V = .10$; OR 132.7; 95 % CI 1.12– 8.56×10^3).

The comparison of the set distance between the ball and the wall by the referee showed no influence of the vanishing spray in any direction. The average distance was 9.24 m for each group ($SD_{NoSpray} = .87$ m, $SD_{Spray} = .93$ m). The average absolute difference from 9.15 m, the prescribed distance by the laws of the game, showed a non-significant mean difference of -5.12 cm (CI -23.2 to 12.9). Overall, 50.6 % of the estimations are within a $|DistanceError|$ of just .5 m, 38.6 % showed an absolute error of .5 to 1.5 m and 10.8 % were off by more than 1.5 m.

The influence on the success of free kicks was evaluated separately for crosses and shots. For neither category significant influences could be identified. Despite the slight increase of goals from 8.6 to 9.3 % after the introduction, fewer shots were recorded in the “OnTarget” category (OR .91; 95 % CI .60–1.38) and more shots in the category “Missed” (OR 1.10; 95 % CI .74–1.65). The number of shots that were blocked by players standing in the wall decreased slightly by .3 % (OR .99; 95 % CI .63–1.54). The success of free kicks that were taken as crosses was just .65 as high after the introduction of the vanishing spray (95 % CI .29–1.46). As stated in Table 2, none of these changes were significant.

Discussion

The aim of this study was the evaluation of the new vanishing spray in soccer, which is used for 42.5 % of the respective free kicks, especially for those in promising positions. Therefore, five hypotheses were tested of which just one, the lower extent of rule violations, could be supported. The four other hypotheses concerning the amount of violations, the distance set by the referee and the success of either crosses or shots were not supported by the results. After separately discussing the findings regarding each hypothesis, the authors point out the significance of these findings and null-findings respectively for the evaluation research process itself as well as the understanding of the underlying phenomena.

A reduction in violations of the minimum distance, could not be proofed by the share of free kicks with violations. One out of four free kicks is still affected by such an incident. Nevertheless, the number of massive

Table 2 Test statistics and effect sizes of the comparisons of the investigated variables before and after the introduction of the vanishing spray

Nominal variables	No spray	Spray	χ^2	p	V
Violations (H 1.1)	25.4 %	25.8 %	.01	.925	.00
Massive violations (H 1.2)	16.7 %	10.7 %	4.58 ^a	.032	.09
Successful crosses (H 3.1)	26.7 %	19.2 %	1.26	.261	.09
Success of shots (H 3.2)			.31	.857	.03
OnTarget	34.4 %	32.3 %	.23	.635	.02
Missed	38.8 %	41.2 %	.26	.611	.02
Wall	26.8 %	26.5 %	.00	.954	.00
Continuous variables			t		d
DistanceError (H 2)	.70 ± 0.64 m	.65 ± 0.58 m	.56	.577	.01

Continuous variables are stated as mean ± standard deviation

^a Significant differences

violation showed a significant decrease. Even though the respective effect size is just .09, the risk ratio shows a 1.56 times higher risk for massive violations for the control group. Thus, the spray had a positive effect on the extent of violations, and the main goal of the introduction was at least reached in curbing such incidences. In addition it is worthy to point out a limitation of this study. For sprayed free kicks, the observers could use the marked line as a reference, which (obviously) wasn't possible for other free kicks. Consequently, for the latter category, the observers were instructed to note only violations if they can see the defenders moving towards the ball. Thus, if the wall reduced the distance while weren't covered by the broadcast, this couldn't be identified by the observers. This step was necessary to reach a good inter-rater agreement, but likely affected an underestimation of violations for free kicks without spray due to methodological reasons.

Despite this frequent incidence, the results also showed that violations of the distance rule were punished neither before nor after the introduction of the vanishing spray. Just a few verbal cautions were awarded after the introduction, but these verbal cautions not fulfilling the required extent of punishment. Two reasons could lead to this lack of punishment for rule violations. The first reason simply is that the referee can't identify the violation. Since a free kick is a situation that has a comparatively clear structure in which the referees can focus more specific, the authors assume that the referees should be able to identify violations of the required minimum distance. In addition it should be even more possible for the referees with a marked line on the field, which now is available after introduction of the vanishing spray. Thus it is possible that some kind of unwritten rules come into play. According to D'Agostino (1995), unwritten rules can be

seen as unofficial, implicit conventions that determine how the rules of a game are to be applied in specific circumstances. Translated for this situation, we suspect some kind of agreement between players of the different teams and the referee that the distance rule should be enforced less rigorously for free kicks. Considering that the correct punishment of such violations is even stated one more time in an extra annotation in the official laws of the game, this could be seen as a discrepancy between the protagonists on and off the field.

For positioning the wall the referees have to estimate the distance to the ball without any external aids. In more than 50 % of the free kicks the referees were within a distance of half a meter, but also a decent amount of errors of more than 1.5 meter occurred. We think the mean absolute error of .70 m before the introduction was still acceptable, especially considering that the mean distance to the ball is just 9 cm to high. The vanishing spray can't support this process directly but led to more awareness of the distance in general. However, this didn't affect the quality of estimation by the referees as there was no significant decrease in the absolute error.

Round about a third of the shots in this study were on goal, slightly more than one out of four is blocked by the wall and the remaining round about 40 % missed the target or were blocked by another player outside the wall. The results are pretty similar to previous findings of Carling, Williams and Reilly (2005). In an investigation of 152 attempts of the World Cup 2002 they found a rate of 8 % for goals and round about 35 % on goal (combining their respective categories). They state that only seven percent of the free kicks were blocked by the wall but a surprisingly high amount of free kicks that went through the wall (15 %), which sums up to a similar rate compared to this study.

The overall rate of successful free kicks taken as crosses was short to 23 %. To compare the values to those of previous studies, it is necessary to consider that we used a different definition of "successful". Casal et al. (2014) for instance stated that of 21.8 % of indirect free kicks resulted in a shot in international competitions. As mentioned above in the current study a cross was rated successful if the ball reached a player of the same team in a way that this player could execute a controlled action with the ball. That's the only goal that the player who is taken the free kick can control and vice versa for what he has to pass the defensive wall. Thus, this definition of "successful" is more appropriate for this study and arguably also for other studies that focus rather on the taking player.

The use of vanishing spray should benefit the offensive team. Despite a slight increase of the goal rate, this could not be proved with the results of this study. No

significant differences were found for the success of free kicks, neither for free kicks taken as shots nor for crosses. The rate of successful crosses even decreased by 7.5 % which, however, wasn't a significant difference. Comparing the free kicks with massive violations of the distance rule indicate that this violations don't affect the outcome of the free kick, but there wasn't a large enough sample to run reliable analysis. Another reason could be that the curbing of rule violations is too small to create a significant benefit for the kicking team, especially considering the variations in the distance set by the referee.

Summed up, the introduction of the vanishing spray basically fulfilled its main goal, by reducing the extent of violations of the minimum distance rule, but didn't lead to further positive effects. Especially for the purpose of evaluation research, the respective null findings are as valuable as other findings to estimate the worth of an intervention. Rather spoken, this worth is estimated by the respective stakeholders, which must thoroughly consider all results of an evaluation for the decision making process (e.g. Stufflebeam 1983). This study illustrates an interesting example, as more and more competitions start to use the vanishing spray, despite an effect that seems to be rather small. Thus, in the eyes of the majority of the respective stakeholders the merit seems to be big enough to outweigh the disadvantages. A similar case was already made for another device in soccer before, the goal line technology (see Kolbinger et al. 2015).

In addition, evaluation researches create new knowledge about the underlying phenomena (Stufflebeam and Shinkfield 2007), which is especially true in this study for the distance set by the referees or the application of the respective set of rules. For the first time, it was shown that there is a discrepancy between the official set of rules for free kicks and its execution on the field, which is also true for the investigation of the referee's estimation of the minimum distance. The respective findings both raise questions concerning the use of the vanishing spray. Users of the device need to be aware that the device controls the compliance of a certain distance, which is in fact not the intended distance for most of the time (Oldfather and Fernholz (2009) describe a similar phenomenon concerning the first down marks in American Football). The respective associations also need to be aware, that there is a lack of punishment for minimum distance rule violations, which can't be solved solely by the use of vanishing spray.

Conclusion

The findings of this study point out the importance of evaluating innovative devices that support the referees in game sports. Based on the objective targets of the respective association the authors showed that the

vanishing spray fulfilled its main goal by decreasing the extent of violations at least in some extent. But in addition to that, these evaluations not just generate feedback concerning the new device but also on the underlying phenomena. In this study, the results also indicate a lack of application for the distance rule. Despite a frequent incidence of violations of the minimum distance, none of the investigated free kicks was retaken neither a yellow card was awarded due to this reason. The authors suggest two ways to overcome this discrepancy. On the one hand, the associations could try to increase the awareness of the appropriate punishment of violations of the distance rule. On the other hand, the rule itself could be adapted by changing the extent for the punishment.

Authors' contributions

OK conceived the study design, provided the theoretical framework, analysed the data and drafted the manuscript. DL conceived the study design, contributed to the statistical analysis and drafted the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

We want to gratefully thank Michael Stöckl (University of Vienna, Institute of Sport Science) for the visualization of the spatial distribution and Hendrik Weber (DFL, Head of Strategic Projects) for his contribution in study design.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Each professional soccer player signs a statement of consent to being monitored during matches by signing a contract of employment with the respective club of the German Bundesliga. All procedures performed in the study were in strict accordance with the Declaration of Helsinki as well as with the ethical standards of the Chair of Trainings Science and Sport Informatics of the Technical University of Munich.

Funding

The study was commissioned and funded by the DFL Deutsche Fußball Liga GmbH. This work was also supported by the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

Received: 6 June 2016 Accepted: 8 September 2016

Published online: 15 September 2016

References

- Abramitzky R, Einav L, Kolkowitz S, Mill R (2012) On the optimality of line call challenges in professional tennis. *Int Econ Rev* 53(3):939–964. doi:10.1111/j.1468-2354.2012.00706.x
- Carling C, Williams AM, Reilly T (2005) Handbook of soccer match analysis: a systematic approach to improving performance. Routledge, New York
- Casal CA, Maneiro R, Ardá T, Losada JL, Rial A (2014) Effectiveness of indirect free kicks in elite soccer. *Int J Perf Anal Spor* 14(3):744–760
- Cohen J (1992) A power primer. *Psychol Bull* 112(1):155
- Collins H, Evans R (2008) You cannot be serious! Public understanding of technology with special reference to "Hawk-Eye". *Public Underst Sci* 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 Underst Sci* 21(8):904–921

- Cramér H (1946) *Mathematical methods of statistics*, vol 9. Princeton University Press, NJ
- Cronbach L (1963) Course improvement through evaluation. *Teach Coll Rec* 64:672–683
- D'Agostino F (1995) The ethos of the game. In: Moran WJ, Meier KV (eds) *Philosophic inquiry in sport*. Human Kinetics, Champaign, pp 48–49
- Dohmen T, Sauerermann J (2015) Referee Bias. *J Econ Surv*. doi:10.1111/joes.12106
- D'Orazio T, Leo M, Spagnolo P, Nitti M, Mosca N, Distante A (2009) A visual system for real time detection of goal events during soccer matches. *Comput Vis Image Underst* 113(5):622–632. doi:10.1016/j.cviu.2008.01.010
- Helsen W, Gilijs B, Weston M (2006) Errors in judging "offside" in association football: test of the optical error versus the perceptual flash-lag hypothesis. *J Sports Sci* 24(5):521–528. doi:10.1080/02640410500298065
- 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, vol 556. Springer, Berlin, pp 107–118. doi:10.1007/978-3-319-25249-0_8
- Mather G (2008) Perceptual uncertainty and line-call challenges in professional tennis. *Proc Biol Sci* 275(1643):1645–1651. doi:10.1098/rspb.2008.0211
- Oldfather CM, Fernholz MM (2009) Comparative Procedure on a sunday afternoon: instant replay in the NFL as a process of appellate review. *Indiana Law Rev* 43(45):45–78
- Oudejans RR, Verheijen R, Bakker FC, Gerrits JC, Steinbruckner M, Beek PJ (2000) Errors in judging 'offside' in football. *Nature* 404(6773):33. doi:10.1038/35003639
- Parsons CA, Sulaeman J, Yates MC, Hamermesh DS (2011) Strike three: discrimination, incentives, and evaluation. *Am Econ Rev* 101(4):1410–1435. doi:10.1257/aer.101.4.1410
- Pope BR, Pope NG (2015) Own-nationality bias: evidence from Uefa champions league football referees. *Econ Inq* 53(2):1292–1304. doi:10.1111/ecin.12180
- Psiuk R, Seidl T, Strauß W, Bernhard J (2014) Analysis of goal line technology from the perspective of an electromagnetic field based approach. *Procedia Eng* 72:279–284. doi:10.1016/j.proeng.2014.06.050
- Rock R, Als A, Gibbs P, Hunte C (2013) The 5th umpire: automating crickets edge detection system. *J Syst Inform Cybern* 11(1):4–9
- Scriven M (1991) *Evaluation thesaurus*. Sage, Newbury Park
- Stöckl M, Lamb PF, Lames M (2011) The ISOPAR method: a new approach to performance analysis in golf. *J Quant Anal Sports*. doi:10.2202/1559-0410.1289
- Stufflebeam DL (1983) The CIPP model for program evaluation. In: Madaus GF, Scriven MS, Stufflebeam DL (eds) *Evaluation models: viewpoints on educational and human services evaluation*. Kluwer-Nijhoff, Boston
- Stufflebeam DL, Shinkfield AJ (2007) *Evaluation theory, models, and applications*. Jossey-Bass, San Francisco

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com

A.3 Scientific approaches to technological officiating aids in game sports

A.3.1 Erlaubnis zur Einbindung

Monday, July 9, 2018 at 11:50:00 AM Central European Summer Time

Betreff: AW: Questions about copyright

Datum: Montag, 11. Juni 2018 um 18:51:17 Mitteleuropäische Sommerzeit

Von: [REDACTED]

An: Kolbinger, Otto

Anlagen: image001.jpg

Dear Otto,
congrats to your Ph.D.-finalization.

To answer your questions:

- The dissertation, including the publication, would be available at the media and publications repository of the Technical University of Munich (mediatum.ub.tum.de). Does this cause any copyright issues?

No problem, because it's open access and we stated the following on our HP:

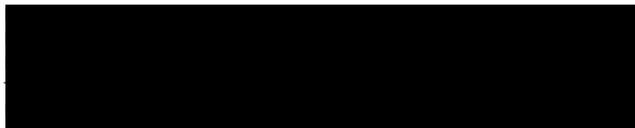
Copyright Notice

Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a Creative Commons Attribution License. CISS applies the Creative Commons Attribution License CC BY-NC to works we publish. This license was developed to facilitate open access – namely, free immediate access to, and unrestricted reuse of, original works of all types.

Under this license, authors agree to make articles legally available for reuse, without permission or fees, for virtually any purpose. Anyone may copy, distribute, or reuse these articles, as long as the author and original source are properly cited.

- Do we have to use the manuscript form or the published form?
How you prefer – you can use and cite the published version if convenient for you.

Best wishes, [REDACTED]



[Editor-in-Chief CISS](#)



A.3.2 Originalveröffentlichung

Current Issues in Sport Science 2 (2017)



Scientific approaches to technological officiating aids in game sports

Otto Kolbinger^{1,*} & Martin Lames¹

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:

Kolbinger, 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

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

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

(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 perceptual 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 perceptual 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, Gillis, 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, Gillis, & 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 perceptual 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 Saueremann (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. Petterson-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 Hamer-mesh (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 temporarily 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 systematic 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.

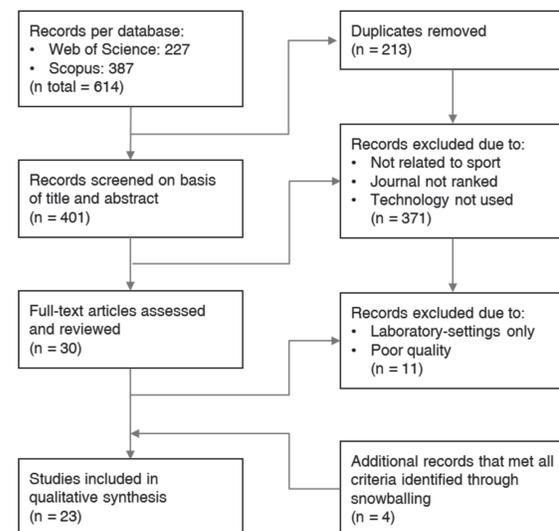


Figure 1: Flow diagram of the study selection

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 (Boroah, 2016), and the goal line technology in football (Kolblinger, 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 Model
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
Kolblinger, Linke, Link, and Lames (2015)	Goal line technology (Football)	Prevalence of critical goal line decisions
Borooh (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 pre-conditions 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, Kolblinger 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, Vajvodova, 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 Kolblinger 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 con-

cepts 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., & Saueremann, 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
- Kolblinger, 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., Steinbrückner, 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
- Petterson-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.