Avugos, Raab, MacMahon, Bar-Eli PSE (accepted March 2021)

Inter-individual Differences in Sport Refereeing: A Review of Theory and Practice

Declaration of Conflicting Interests

The authors declare that they had no conflicts of interest with respect to the authorship or the publication of this article.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Abstract

Most studies on sport officials ignore the variance of the data in terms of individual differences. Here we highlight the need to focus on these differences in refereeing as the main point and goal of the current review. We argue that the study of individual differences is important for referee selection, development and performance evaluation. We present the available research on differences related to referee communication styles, types of interactions, and game management. We regard expertise and experience level as a key contributor to individual differences. Given the gap in research around the in-group variance, we discuss new directions for research and further recommendations for the field. We suggest that the areas in which more research and practice will inform our selection, training, and evaluation processes are communication and personal communication style, while having more flexibility in applying different refereeing styles to facilitate performance in different contexts.

Keywords: sports official, individual difference, decision making, contextual judgment, training

Introduction

Most studies on sport officials report the means between conditions or groups, but ignore the variance of the data in terms of individual differences. Therefore, the core goal of this review is to highlight that we do not know enough about intra-individual and inter-individual differences in referees. We argue that the study of these differences is important for referees' selection, development and performance evaluation. For example, Raab et al. (2020) argue that the existing models and descriptions of the trade-off in making accurate and adequate decisions are not sensitive to intra- or inter-individual differences, and that the awareness of these differences would help to develop personalized and context-specific diagnostic and intervention profiles for referees. An area where individual differences can make a difference to performance outcomes would be, for example, through understanding that the performance of officials improves the sport and reduces injury or injury risk (e.g., Gilis et al., 2006). The study of individual differences in research on referees exists in communication styles (e.g., conflict management styles; Mascarenhas et al., 2008), beliefs (e.g., referees' efficacy; Guillén & Feltz, 2011), and is described in specific observable referee behavior (e.g., number of yellow cards; Unkelbach & Memmert, 2008), interviews and qualitative observations (e.g., types of player management; Cunningham et al., 2014) or self-reports (e.g., self-efficacy questionnaire; Guillén & Feltz, 2011). However, the knowledge is still far from being broad or applicable. In the current review, we point to the gap in research and propose new directions for investigation, as well as recommendations for the field, and, in particular, applications to the selection and training of referees.

Sport officials have always attracted less attention by sport science compared to athletes and coaches, in spite of the often-crucial impact that their decisions may have

on athletes and the outcome of competitions (MacMahon et al., 2015). However, this has been increasingly changing in the last two decades. More research has been conducted around the physical (e.g., Krustrup et al., 2009; Weston et al., 2010) and psychological (e.g., Philippe et al., 2009; Voight, 2009) aspects of refereeing, the communication between and interaction with other sports actors (e.g., other officials, competitors, and coaches; see Cunningham et al., 2014; Mellick et al., 2007), the psychological and physiological aspects linked to referee decision making (e.g., Helsen & Bultynck, 2004; Mascarenhas et al., 2009), as well as the introduction of technology into sports officiating (e.g., Royce, 2012; Spitz et al., 2020).

The early officiating research was focused on bias, personality factors, and different sources of stress experienced by officials (e.g., fear of physical harm, time pressure, worry about making errors). The focus then shifted to training activities, expert-novice differences, physiological functioning, and perceptual-cognitive abilities. More recently, research has covered issues related to communication, vision and decision making, self-efficacy, gender, injury and career development. Several review articles (e.g., García-Santos et al., 2020; Plessner & Haar, 2006; Samuel et al., 2020), book chapters (e.g., Bar-Eli et al., 2011; Dosseville et al., 2014; MacMahon & Plessner, 2013), and books (Livingston et al., 2020; MacMahon et al., 2015) have been published during the last several years. However, most empirical studies approach understanding officials and officiating using the means between conditions or groups, while not much investigation has been done on within-group differences (for recent exceptional case-study type approaches around inter-individual differences see Larkin et al., 2018; O'Connor et al., 2018). With this in mind, we advocate returning to a focus on individual differences in officiating, the core of this paper.

Approach

The importance of individual differences is illustrated in that they are proposed to be linked to on-the-ground officiating choices. While early and arguably more blunt analyses of referee skills focused on personality as a unitary, central factor, our narrative does not hold a personality-like sport psychology perspective. Rather, we use existing inter- and intra-individual differences in refereeing that need explanation and have practical consequences. In addition, we regard expertise and experience level as a key contributor to individual differences. Thus, we use a systematic literature review to show below what we currently know about individual differences in communication/social interactions and player management, and highlight the areas suitable for continued exploration that may help inform the key areas of improving referee practice in selection, training, and evaluation.

Current State of the Research

We used the PRISMA approach (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Moher et al., 2009) to understand the current state of knowledge on sport refereeing, and in particular, individual differences. This systematic literature search revealed that indeed research on referees, sport officials and umpires is a growing and important field for studies. In SPORTDiscus (SD) and Web of Science (WoS) databases, we found in single-search-words, a large number of entries for sport referee(s) (SD 13,071; WoS 590), sport official(s) (SD 164; WoS 45), sport judge(s) (SD 3,085; WoS 646), and sport umpire(s) (SD 2,463; WoS 114) (retrieved 10th of December 2020). However, the gap in the literature and the need for more empirical work on individual differences were evident when we combined the search with "individual difference", and only 24 papers were found in the databases. After removal of duplicates, 19 papers remained, from which only 6 were deemed of a standard to be included in a quantitative synthesis such as a meta-analysis. Scanning

the recent reviews (e.g., García-Santos et al., 2020; Samuel et al., 2020) and books (Livingston et al., 2020; MacMahon et al., 2015), it is clear that the topic of individual differences has been discussed, and a more detailed analysis of papers does reveal that both from a quantitative approach (e.g., standard deviations, subgroup analyses) or qualitative approach (individual variance in themes in interviews or responses) data exist, but not much attention is paid to this emphasis, despite a general acceptance that individual approaches and styles influence performance (Kittel et al., 2019; Raab et al., 2020).

To illustrate, a paper by Balmer et al. (2007) tested how crowd noise influences refereeing consistency in soccer (a hot topic during the COVID-19 restrictions in 2020, given games with and without spectators). Referees made decisions in response to videos in a silent mode (only video without noise) and a noise condition (video with crowd noise) as compared to a base line condition. Although the ratings for self-confidence in these conditions did not differ (16.8, 16.9, 16.8), there were large standard deviations (between 7.6 and 8.4), showing individual differences. In the almost 100 citations of Balmer et al., individual differences are barely discussed. This is surprising given that it may be important to understand whether high or low self-confidence may change the way information from players, video-replay or assistant referees is used for a decision, or how self-confidence changes the way game-management is applied.

Given the above analyses, a systematic meta-analytical review cannot be completed, and as indicated as the goal of this paper, a focus on individual differences will provide a leverage point for future research.

Individual differences in communication

The critical nature of communication in refereeing and in the role of the referee is illustrated by the fact that even poor decisions are often received more positively, if they are communicated effectively (MacMahon et al., 2015). Moreover, players rate referees fairer and more correct when they communicate decisions calmly or provide a short interpretation of the decision (Simmons, 2010). Knowing which decisions need selling and which will be easily accepted is a skill that officials develop with experience (Mellick et al., 2005).

So far, there has been remarkably little research investigating the important area of communication within sports officials. The available research is mainly focused on the nature of referee communications when the penalty for players' offenses can have a greater impact upon the game (contextual judgment). The evidence suggests that referees use 'preventive communication' to guide players away from committing fouls (Mascarenhas et al., 2005a). Another distinct aspect of referee communication, with a different goal, is aimed at persuading the players to accept that they have infringed the rules. However, the same decision could be communicated to players in very different ways (i.e., a 'person-centered' style; Burleson, 2007). A study in Australian sport bodies such as soccer (football), rugby union, rugby league, Australian Rules Football, netball, basketball, and field hockey revealed that interpretive (e.g., interpreting players' behavior and context) and interactive (e.g., adapting communication to the context) communication skills were perceived to be the most important and challenging theme for officiating-development managers and coaches (Cunningham et al., 2014). This highlights that referee communication is an individual difference variable that is critical, yet underexplored.

Referee skilled interactions

Previous reviews of talent development in sport officials have broken down referees' skilled interaction into communication and player management. From this conceptualization, we will focus here on officiating communication (Cunningham et al., 2014).

According to Constructivism Theory (Burleson, 2007), and its application to sport officiating, there are three layers to the communication process: social perception (e.g., the expectations of players and fans), message production (i.e., a message sent by the referee verbally and/or non-verbally) and message reception – sometimes by multiple audiences (i.e., players, coaches, media, spectators, the refereeing body). At the micro level, communication is with a single player, while the communication in the meso level is directed to all players on the field, and the macro level includes a large group of people such as the media or the public at home (Cunningham, 2012). For example, the increased use of open microphones in several top-level sports allows the audience both in the stadium and at home to hear the assistant referees when reporting foul play to the head referee.

Creating communication includes the verbal and non-verbal components. In addition to the spoken language (the words), the vocal messages also include characteristics such as volume, articulation, pitch, emphasis and speech rate – sometimes using the acronym VAPER (Nelson-Jones, 2012). This may help the officials ensure that their message is delivered in the appropriate and effective way. Under stressful game-play conditions, each of these VAPER characteristics is likely to increase, so that referees might speak loader and faster, with greater articulation and a higher pitch and emphasis. Communication becomes even more critical during stressful moments. Yet, the non-verbal skills (e.g., facial expression, appearance, eye contact, body language, use of whistle) are also very important when communicating

both the official's 'call' and 'no-call' decisions. In particular, 'presence' (e.g., physically fit body, strong posture, eye expression) was found to be the most dominant attribute when describing the elite official profile (Mascarenhas, 2006). For example, the acronyms RAC (Relaxed, Assured, Confident) – for body language and CAC (Clearly, Accurately, Concisely) – for message description, are used by touch judges in rugby to guide them in reporting foul play to referees.

It should be noted that because many of these variables in communication differ between male and female bodies, and the dynamics between male and female communicators differ, this is an area ripe for investigation, particularly where sports are now promoting more: a) professional female competitions, and b) female elite officials who may officiate males. So, combinations of interactions are also unexplored, but may shed light on key performance factors: male/female official and male/female competitors (Coulomb-Cabagno et al., 2005; Souchon et al., 2013).

Officials should recognize that there are multiple refereeing styles available to them in order to regulate and manage any conflict that may arise between themselves and the players (Mascarenhas et al., 2008). For example, a referee who is low in assertiveness and low in cooperativeness would probably say nothing in response to a player who challenges his/her decision (i.e., an *avoiding* style), while another referee who is high in both dimensions would assertively ask the coach to calm down his players (i.e., a *collaborating* style). Other conflict management styles include the *forcing* style (high assertiveness, low cooperativeness), the *accommodating* style (low assertiveness, high cooperativeness), and the *compromising* style (medium assertiveness, medium cooperativeness). Referees who are most effective at managing conflict tend to be those who are able to switch between different conflict management styles.

Underscoring the importance of communication and style is the work of Slack and colleagues (2013), which identified that referees' communication skills and their ability to establish player and manager respect are key components of successful game management in the English premiership soccer league. Other main interaction components include conveying positive body language (67% of respondents), building player trust (47%), talking to players and managers (40%), and showing empathy for players (40%). What has not been explored, however, is whether different styles may also be appropriate for different persons/situations in the game. Therefore, referees should develop familiarity with their own style that they are likely to use when the pressure is on and be flexible enough to change it according to the characteristics of the situation encountered. The current understanding of this area of practice is lacking, and may provide greater understanding of referee performance.

Types of referee player management

A study of personal qualities among officiating managers and coaches in seven peak national Australian sport bodies identified several positive qualities associated with good referees (e.g., respectful, empathetic, approachable, professional, calm), and a number of negative qualities (e.g., dictatorial behavior, over-controlling, domineering) representing poor refereeing (Cunningham et al, 2014). A 'person-centered' approach (Burleson, 2007) suggests that a referee should be empathetic to the player's position, rather than being judgmental and threatening. Moreover, an effective interaction requires referees to develop not only communication skills, but also humanistic judgment skills that can help them to be more responsive to fluctuations in player temperaments.

Players usually are aware of referees' different styles and they know what to expect from each referee (e.g., which referee they can talk to). However, individual

differences in refereeing communication styles, personality, and their impact on game management have not yet been explored. There is also little research to identify the best practice for referee decision communication. Future research exploring players' and coaches' preferences for different refereeing styles to differentiate patterns between different sports, levels of play, cultures and nationalities will inform progressions of officials between levels (e.g., national to international/world championship levels) and facilitate performance in different contexts. This highlights that not only is communication an individual difference that may be influential, but flexibility of style may be more important and a facilitating or limiting factor that can be explored (e.g., a referee who cannot adjust to another context) for different sports.

Individual differences in perception, decision making, and behavior

Visual perception

Perception refers to the ability to pick up the relevant visual information from the environment for decisions and actions. Possible influences on perception are therefore of great importance for later judgments. So far, there has been little research about how vision and perception influence decisions in officiating. In particular, although we know that positioning is critical (e.g., Ghasemi, et al., 2009; Hüttermann et al., 2018; Spitz et al., 2016; Ste-Marie, 2003a), not much is known about the ability of referees to position themselves on the field for accurate decision making.

Perceptual skills develop through different learning processes (e.g., deliberate effort, training), however most officials acquire them through experience in officiating actual games or through their experience as former players (Renden et al., 2014; MacMahon et al., 2009). Few studies have examined expertise differences in refereeing, however the evidence shows that the more experienced referees stand out in their visual performance (Pizzera & Raab, 2012a; Ste-Marie, 2003a). For example,

referees in the Bundesliga were generally better in their visual performance than the lower-class referees in handball (Jendrusch et al., 1993), and differences in gaze behaviors were also found between higher- and lower-level ice hockey referees (Hancock & Ste-Marie, 2013).

Early study with professional gymnastics judges showed high abilities by the judges with regard to visual search (Bard et al., 1980). Studies also showed that more experienced referees develop more effective perceptual strategies (e.g., using episodic memory; MacMahon & Ste-Marie, 2002; Paull & Glencross, 1997; Ste-Marie, 1999, 2000), and that expert soccer referees have better basic visual skills (e.g., peripheral vision, eye movements, speed of shape recognition) than novices or laypeople (Ghasemi et al., 2009; Ghasemi et al., 2011; Pizzera et al., 2018). In general, experienced referees (e.g., football; Gilis et al., 2006; MacMahon et al., 2007. ice hockey; Hancock & Ste-Marie, 2013) are better at picking-up and processing the more relevant and useful information for their decision making than lower-level referees, despite the fact that no differences were found between groups in gaze behaviors.

Differences in visual attention were also found among artistic gymnastics judges, with a lower number of fixations and significantly more targeted eye fixations among the expert judges. Moreover, novice judges detected only half the gymnast's errors detected by the expert judges (Bard et al., 1980). Studies showed that the simple choice of visual perspective could also affect the accuracy of perception. In two studies in baseball (Ford et al., 1995; Ford et al., 1997) it was found that viewing more altitude and distance cues than the traditional umpire position could lead to fewer mistakes in the ball-strike decision. Thus, individual umpires have worked out the optimal position for themselves, and researchers understanding this can lead to better instruction and exploration for positions that other umpires can explore. This

points to the idea that we need to study individual differences and have some options for exploration to help practice. Moreover, if we have differences between groups based on expertise, it is also worthwhile to examine differences *within* these groups to form even more sensitive analyses.

Being in an optimal position or being able to move to the best position is necessary for picking up specific and/or contextual information (Gilis et al., 2009; Plessner et al., 2009). Changes in viewing perspective occur due to either locomotion (run, sprint, jog, walk), or eye/head movements while scanning the environment, or both. For example, Mallo and his colleagues (2012) found that accurate decision making in the FIFA Confederation Cup was optimal when the referee was at distances of 11-15 meters from the scene, but distance did not have a significant influence on the quality of decisions for Brazilian soccer referees (de Oliveira et al., 2011).

A group of researchers (Oudejans et al., 2000) has found that errors in judging offside were, at least partly, due to incorrect positioning of the referees relative to the actual offside line and the corresponding viewing angle. In most cases, top-level assistant referees in the European First League were positioned about 1 meter away from the ideal position to judge offside (Catteeuw et al., 2010b; Oudejans et al., 2005). Similarly, assistant referees in FIFA World Cup matches were frequently not on the offside line (between 76% and 87% of cases) when making offside judgments (Catteeuw et al., 2010c). Other researchers (e.g., Helsen et al., 2006. see Oudejans et al., 2007 for a rebuttal) suggest that an optical error associated with the flash-lag effect (Baldo et al., 2002) could explain the overall bias in offside decisions (Nijhawan, 2001). However, what we do not know is whether there may be individual differences within this. For example, individuals may drive their behavior within their individual constraints (e.g., Newell, 1986) and adjust their position based on their

particular movement speed or height, or even depth perception. Thus, more detailed understanding of individual choices, or a 'tool box' of strategies that can fit for both individuals and situations, can be explored.

Errors due to the human perceptual system occur even if the optimal positioning of the officials is relatively more stationary and prescribed by the rules of the sport, such as in baseball, tennis, and volleyball. For example, Plessner and Schallies (2005) found that gymnastics judges err about whether the arms of athletes deviated from horizontal when holding a cross on the rings, and these errors in judgment increased with an increase in deviation (0, 30 or 60 degrees) from a frontal view. Relating this to our main idea, we suggest that all of this knowledge of the differences feeds into an argument for the need to examine on individual levels.

Decision making and the refereeing context

Through their development to become experts, officials acquire both the declarative knowledge of their respective sport (the rules), and the procedural knowledge of how to correctly use it. For example, FIFA referees were found to be better in making decisions for football incidents than national level referees, and national referees again were better than players (Gilis et al., 2006; MacMahon et al., 2007). Different aspects of physical demands in sports officiating have been examined (e.g., for soccer refereeing see Castagna et al., 2007; Reilly & Gregson, 2006), however the relationship between the physical demands (e.g., fatigue) and decision making has been exclusively studied through the use of indirect analysis methods. In particular, a few studies have investigated the effect that distance from the play has on the accuracy of judgments (e.g., Mallo et al. 2012; de Oliveira et al., 2011), with inconclusive results. Other studies (e.g., Krustrup et al., 2009; Krustrup et al., 2002) found no relationship between the overall distance covered and the speed that referees

were moving during a game (walk/jog/sprint) and the quality of their decisions (Mascarenhas et al., 2009).

Human judgment and decision making are assumed to arise from the interaction of two different cognitive systems, Type 1 and Type 2 (Kahneman & Frederick, 2002). For sport officiating, the accuracy of decisions made by referees is likely to rely on an intuitive type of judgment (Type 1), whereas adequate decisions rely on both intuitive and deliberate processing (Type 2), in which the default mode of information processing is system 1. In other words, deliberate and slower processing is triggered when contextual cues (e.g., prior decisions) cast doubts on the initially triggered decisions (Helsen et al., 2019). Biases are more expected in high-conflict situations (e.g., second penalty for the same soccer team; Plessner & Betsch, 2001) or in ambiguous situations (e.g., umpires' leg-before-wicket decisions in cricket; Adie et al., 2020).

Decision processes differ between the types of officials in terms of the different levels of cognition that are involved (MacMahon & Plessner, 2013; Plessner & Haar, 2006). For example, having to decide about a contact situation in soccer is clearly more complex than deciding whether a ball in tennis has crossed the line. Also, within a sport, different decisions are relatively more or less complex. Therefore, we need to have a more detailed look at officiating, and not assume that every decision works the same way for every official (see MacMahon et al., 2015 for officiating in different sports). The modeling of cognitive processes leading from the stimulus event (i.e., contact between players) to a subsequent decision (e.g., awarding yellow card) include a sequence of steps as suggested by social cognition theory, and applied to referees by Plessner and Raab (1999) and later by Plessner and Haar (2006). Sequences of steps can also be explained by information-processing models such as

the conceptual model of sequential decisions among soccer referees (Samuel et al., 2020), or by Decision Field Theory through accumulation of evidence for ball games (Raab et al., 2020).

Empirical evidence shows that referees often err in their judgments, either because social information that should not have any meaning is used to form judgments (e.g., the judgments of other judges; Scheer et al., 1983), or because the relevant available information is not systematically considered, but rather integrated in a shortened, heuristic way (e.g., judgments based on an overall impression; Plessner, 1997, 1999). Offside decisions by the assistant referees in soccer are a good example. It is often impossible for the referees to consider all the information needed to make immediate decisions, and therefore, it may make sense to leave out information and rely on shortened judgment strategies or even single cues, which in most cases allow a correct judgment to be made.

Errors at different stages of information processing can lead to incorrect decisions. For example, there can be misperception of physical contact between players, or a false memory that a player previously violated the rules and thus should be strongly punished. Biased judgments have also been detected due to the influence of various types of prior knowledge, such as the reputation of players or teams (e.g., Jones et al., 2002; Plessner, 1999), stereotypes about race (e.g., Stone et al., 1997), skin tone (Silberzahn et al., 2018), physical appearance (of the player: Van Quaquebeke & Giessner, 2010; of the referee: McCarrick et al., 2020) and gender (e.g., Souchon et al., 2009). So, these are key individual difference variables that can be discussed and considered.

Biases could also be the result of memory influences of prior information about an athlete's performance, as demonstrated in a series of experiments by Ste-Marie and

her colleagues (e.g., Ste-Marie, 2003b; Ste-Marie, & Valiquette, 1996). Research in soccer has documented biases related to the recent history in the game (e.g., Plessner & Betsch, 2001), visual errors (e.g., Gilis et al., 2008; Maruenda, 2005), crowd noise (e.g., Balmer et al., 2007; Nevill et al., 2002; Page & Page, 2010; Unkelbach & Memmert, 2010), home team advantage (e.g., Boyko et al., 2007; Sutter & Kocher, 2004), favoritism towards successful teams (Erikstad & Johansen, 2019), and uniform color (e.g., Greenlees et al., 2008). For example, when players were dressed in black, experienced soccer referees judged their behavior as more aggressive, compared to when players, with the same behavior, were dressed in white (Frank & Gilovich, 1988). However, this effect does not seem to be the case for all cultures. In a study with Turkish football referees, Tiryaki (2005) found no comparable influences of black uniforms. On the other hand, Hagemann et al. (2008) found that taekwondo competitors were favored by the referees when wearing red rather than blue protective gear. Motivational influences might also contribute to judgment errors. For example, judges at the Los Angeles Olympics gave relatively fewer points to gymnasts when they came from teams that were either just ahead of or just behind the respective judge's team (Ansorge & Scheer, 1988). Again, looking for the individual difference variables, it might be that nationality plays a role in biased behavior.

Studies have also shown that referees are influenced by their own prior decisions (e.g., Plessner & Betsch, 2001, see Mascarenhas et al., 2002 for a rebuttal), possibly because they need to demonstrate consistency in applying the rules (Unkelbach & Memmert, 2008). For example, data on successive penalty decisions in the Bundesliga from 1963-2006 reflected a referee tendency towards compensation, decreasing the chances of the same team having another penalty, while increasing the chances of a

penalty for the other team (Schwarz, 2011). In this case, individual effects might arise in a desire for consistency.

To summarize, biases in judgment are likely to occur among all three major types of officials, as classified by MacMahon and Plessner (2013). These errors of judgment may be the result of both internal (competitors and referee-related) and external (situational) error sources (Plessner & Raab, 1999), and should be further explored for individual difference effects. Among the documented internal problems are those related to the different cognitive stages of information processing (e.g., repetition effects, perspective, heuristic judgment, social influence), whereas the external problems are those related to the capability of referees to regulate their processing (e.g., overload of information), the selection of competitors and referees (e.g., "international bias"), and the competition site (e.g., the home advantage). Rules with a high margin of judgment for the referees (e.g., "dangerous play" in hockey) or overload of information processing (e.g., part recognition on the rings in artistic gymnastics) add to the potential external errors of judgment. Despite the influence of both internal and external factors, sport associations mostly attempt to reduce errors by addressing the situational, external factors, such as FIFA's decision to use two referees in football games.

Various possible solutions could well eliminate the internal source of errors, including appropriate gaze strategies and repositioning of umpires and referees for optimal viewing angles (Ford et al., 1997; Ford et al., 1995). The inclusion of an individual differences perspective in an approach such as this can amplify its effectiveness. Additional solutions are warm-up of athletes without judges, isolation of competitors and referees, control of known cues, and the use of external memory aids. Similar suggestions include the use of a systematic error analysis (e.g.,

systematic categorization of errors by artistic gymnastics judges) followed by practice and training of diverse scenarios using this information, for example, through the use of training videos (Niebuhr et al., 1998). Again, an individual differences perspective will accelerate improvements with such an approach. For referees in ball games, skills such as attention-steering, selection and weighting of information for judgments and decisions should be trained separately, due to the large number of sources of information to pick-up and process from the environment (Plessner & Raab, 1999). The individual differences here might be around learning approaches, and how to integrate individual differences into learning design. For example, referees should be asked to explore and discover the cues that work best for them, rather than having uniform and dictated cues.

Possible solutions for the external problems include the use of technical aids (e.g., video; Spitz et al., 2020), partitioning the referee's tasks, equal participation of law courts, and the choice to run playoffs on neutral ground. In ball games, there are three possible strategies for improving referees' decision processes. First, providing referees with extra information and direct help for their decisions (e.g., goal line technology that provides information on the ball relative to the goal in soccer), or even to reduce error to the greatest possible extent (e.g., the "hawk eye" in tennis that directly provides location information with a high level of precision). Second, using a delayed replay of the corresponding situation (e.g., fouls) that enables referees to correct erroneous decisions with review by another referee or judge, as in rugby's use of the Television Match Official, or match officials in international cricket matches. Finally, having an authority that can overrule decisions, such as the chair umpire in tennis, who can overrule a decision by a line judge (Hartmann, 1991).

In light of the increasing debates around the role of technology in sports officiating (e.g., Collins, 2010; Johnson & Jason, 2016), and the explicit improved accuracy over unassisted human officiating, there have been calls lately (e.g., Bordner, 2019) to revise the current criteria by which accuracy is measured in modern variants of sports games (e.g., touching the ball with the middle finger bone in beach volleyball, or dribbling after at least 3 m in basketball). Thus, improving referees' decisions should be based on psychologically meaningful partitions of the referee's tasks, as well as the corresponding rule changes.

With regard to measurement and analytics, we suggest that sophisticated analytics can contribute as feedback for an official. For example, analytics may show that the referee as an individual makes errors in a specific situation, or tends to act in a certain way, which is individual to this particular official. Such analytics should help evaluations and pick up on things that are subtle but significant, and might not be processed otherwise.

Interaction and game management

Referees' choices in ball games have been described as a tradeoff situation between having accurate versus adequate decisions (Schweizer & Plessner, 2016). A preventive refereeing approach would suggest penalizing even the minor violations early in a game (MacMahon & Mildenhall, 2012). However, in officiating games adequately, referees are required to adapt their refereeing to the particular dynamics of each game (Brand et al., 2008). For example, elite basketball referees under a laboratory condition with no sequential context awarded more rigorous sanctions than their colleagues (Brand et al., 2006). The decision about "what is right for the game" is mostly based on the referees' previous experiences and gut feelings about the consequences of their decisions (Brand & Neß, 2004), and often follow a socially

driven thinking process (MacMahon & Plessner, 2013). This elasticity in applying the rules seems to represent good practice, but it forces the referees to invest high cognitive and psycho-physiological effort. We suggest that there are different 'styles' that can be individual, and may be part of a referee profile. In any case, sport associations (e.g., FIBA) do not expect referees to make perfectly accurate decisions, but rather to avoid incorrect decisions in critical "game-changing" situations (e.g., goals, penalty decisions, direct red cards incidents).

Investigating decision patterns, Unkelbach and Memmert (2008) detected a tendency among referees in the first German soccer league to avoid severe calls early in the game. The data showed that only 606 yellow cards were awarded during the first 15 minutes of the games compared to 1505 during the last 76-90 minutes interval. Similar patterns were detected in the analysis of red cards during 41 seasons of the same soccer league (Bar-Eli et al., 2006). Unkelbach and Memmert suggested that referees need enough time to calibrate their judgment scale, and therefore they avoid using sanctions at the extreme end of the scale (i.e., yellow cards) right at the beginning of the game. If they start with a yellow card early, they might end up with many additional cards to stay consistent. Considering individual differences, there might also be soccer referees who value the best opportunity to act early in a disciplinary manner, in order to indicate the ‘disciplinary line’ and standards to the players, and thus avoid a game with too many yellow cards.

While the game management approach is essentially a conscious and deliberate strategy to decision making (i.e., based on planned considerations), calibration is assumed to be an unconscious, automatic process, which functions as a pre-condition for game management. Yet, experiments aimed at validating the role of these two mechanisms provided inconclusive results (Memmert et al., 2008). Additional follow-

up studies (Fasold et al., 2012; Unkelbach & Memmert, 2014; Unkelbach et al., 2012) compared the calibration explanation with other accounts of serial position effects, however this was based on lab experiments conducted in areas outside the sport domain (e.g., academic examinations, pricing decisions). Other studies showed that some of the offences detected by the VAR-system (Video Assistant Referee – a system that reviews on-field decisions made by the head referee with the use of video footage and headsets for communication) are not being called (Dawson et al., 2007; Stride et al., 2011), either because the referees deliberately ignored them, or were unaware of their occurrence. Therefore, it was suggested that is would be reasonable to consider that the behavior of referees reflect a game management approach, even if calibration contributes to some extent to the effect.

Several process models were proposed to describe decision contexts based for example, on social cognition and cue-based learning theory (for an overview see Goldstein, 2004), and dual process theories (e.g., see Diederich & Trueblood, 2018 for risky choices). For example, the idea of cue-based learning and heuristic judgment was applied to the problem of designing a training tool for improving the quality of decisions made by soccer referees (Brand et al., 2009). The models are able to describe how situations are judged considering the relevant information and processing, but do not explain how referees actually *shift* from rule application (accurate) to game management (adequate decisions) during game play. Moreover, the descriptions are not sensitive to individual differences that would help to personalize and specify the selection and development of referees.

Recently, Raab et al. (2020) proposed a dynamic type of threshold model that is based on concepts derived from Decision Field Theory (DFT; Busemeyer & Townsend, 1993). The model suggests that differences in decisions between soccer

referees can be explained by game management shifts related to individual thresholds and changing context of play. The model has been described as a threshold model with two thresholds of game management (high/low) and each event (foul/not foul) as getting closer to one of these thresholds. Thresholds can be explained by individual differences in personality traits, accumulated experiences and preferences. The threshold is an important parameter for controlling speed-accuracy tradeoffs.

For example, when considering inter-individual differences for two individuals with different referee styles, the model predicts that a referee with a high preference for intuition and lower threshold for game management ('let the game flow' type of referee) could include more game management as compared to a 'by the book' referee, who is high on preference for deliberation. The model thus highlights that decision preferences might be a key individual difference variable. Further, the model predicts that situations of changing context require shifts in game management. Thus, it is expected that under a 'no aggression' scenario, the referees would apply the rules as they are, because they are both under their threshold of applying game management. Context is overlooked, and thus accuracy is increased. But if the game is one with constant high aggression, then communication and sanctions would systematically go up. In particular, it is expected that a low threshold referee would call more fouls (e.g., more yellow cards) in order to stay consistent with his/her previous decisions. Thus, the level of accuracy and adequacy of decisions is influenced by the principle of consistency.

Finally, if the level of aggression fluctuates all through the game, as is more typical in most ball games, the model predicts more communication and sanctions that are less consistent with previous calls. In this case, the accuracy and adequacy of decisions are influenced by each referee's subjective threshold, and depend on the

specific context at the time that the offense occurs. In addition, the model assumes an interaction of both factors: inter-individual differences and context, however, part of this context is the individual referee's unique history and past experiences. This would mean that in games with varied levels of aggression, a referee with a high threshold for switching to game management would make a switch *later* in a game that has increased in aggression, compared to the referee with a lower threshold.

Discussion

In sports, selecting or developing stakeholders is often focused on athletes or coaches. The core goal of this review is to highlight that we do not know enough about intra-individual and inter-individual differences in referees to allow us to mirror talent systems that have been developed for other sport stakeholders (e.g., athletes). Nevertheless, the basic principles of talent systems in sport can be transferred to referees as well. Most structures would focus on talent selection and talent development, as well as performance evaluation, so we will apply these to the road of expertise in refereeing as well.

The Foundations–Talent–Elite–Mastery (FTEM) framework is used to examine development systems for sports, and it has also been adapted to officials (FTEM-O; MacMahon et al., 2015). Within current developments of this model, there is a particular acknowledgement that individual circumstances, such as an official's role outside the sport (e.g., skills developed in their profession), has an influence on their performance and skills inside the sport. This signals the importance of individual variables.

We suggest that these three activities in refereeing and officiating (i.e., selection, development, and performance evaluation) for sports organizations should be aided,

for example, by: a) contributing to identifying individual areas for development within a selection process, b) attending to individual needs in the training process, and c) acknowledging and allowing for individual approaches in the evaluation process.

Selection

Most sports have difficulties finding enough sports officials at all levels, which is why it is relatively simple to become a sport official. In many cases, the only requirements are being at least 14 years of age and a member of a club. Officials usually do not even have to be active as athletes in the sport they would like to officiate. However, would-be officials need to pass theoretical and practical tests, and in some sports (e.g., soccer, American football) additional fitness tests are required on a regular basis. Even when officials are very talented, it usually takes years to reach a high level (e.g., an average of 16 years for soccer referees) due to the demand of gaining experience and undergoing years of close monitoring and evaluation, particularly when it comes to referees who are not former elite players (MacMahon et al., 2007). It is important to note that the pay structure for many sports means it is not a viable career option, and thus referees usually have additional priorities.

An additional and related acute problem is the dropout rate of sports officials, which is often caused by both organizational (e.g., inadequate resources and facilities) and personal reasons (e.g., poor feedback, abuse by coaches and spectators). However, research has shown that despite abuse and other stress factors facing sports officials (e.g., time pressure, performance concerns), they seem to use social interaction in a way that helps them continue to officiate (Kellett & Shilbury, 2007). More research is needed on individual differences around the negative and positive experiences of sport officials, which can facilitate the development of better strategies

to retain them in their job, and explore whether related variables also influence performance.

Development

Officials' basic training includes the study of the written rules of the sport (i.e., the declarative knowledge). The implementation of the rules (i.e., the procedural knowledge) is possible through the development of training tools that focus on the demands (e.g., physical fitness), key decisions or typical errors in officiating. For example, research has indicated that improvement in aerobic metabolism seems to help soccer referees to keep up with play and achieve better viewing position to make accurate judgments of foul situations (Krustrup & Bangsbo, 2001). Although few studies exist that examine the effect of fatigue on cognitive judgments (e.g., Elsworthy et al., 2014; Kittel et al., 2019; Paradis et al., 2016), more research is needed in order to extend our understanding in this particular area.

If taking an individual differences lens and allowing for multiple ways to achieve a goal in a task, then you can allow for the fact that one referee can compensate for a lack of skills in one area with a higher level of skills in another area (e.g., she may make a few more errors of law, but is an excellent communicator). So, while we can point out the need to, and keep working on, the deficits, two referees having the same overall performance can be selected to particular games or tournaments, given where their deficits lie. Beyond having individual training, and not just group training, some officials may be more self-regulated than others, so they will seek out their own training, while others may need a different approach, but both achieve the same level of structure proficiency. Thus, programmers need to be sensitive to these differences and to accommodate them.

Independent of physical fitness, decision making performance can be trained through the use of referee-specific training models and video-based training programs, as well as training on the field or in the gym by having referees watch scenarios in real-life environments (Pizzera & Raab, 2012b). For example, video training methods combined with appropriate feedback have been shown to enhance offside decisions (e.g., Catteeuw et al., 2010a) and decision making in potential foul situations in soccer (e.g., Schweizer et al., 2011). Schweizer et al. developed training programs that specifically focus on better decisions in soccer referees called SET (SET is German for Referee Decision Training). Similarly, a six-week period of training using videotapes of different sets of tackles significantly improved the correct decisions made by rugby referees (Mascarenhas et al., 2005a).

Training for sports officials should also include different aspects of coping with stressful events, while considering the individual differences in referees' personal style (e.g., avoidance coping, approach coping) as suggested by Anshel and Weinberg (1999). Consideration of individual differences in such an approach would train different adaptive coping strategies that the referee can then use according to his/her preference. Investigating of the efficacy of such training interventions will advance the field in both research and practice of refereeing.

Finally, recent work in athlete development (Burns et al., 2019) shows the importance of lifestyle practices and psychological skills in highly successful athletes. Applying a similar lens to understand how performers in refereeing account for their particular circumstances is a much-needed area of research that can directly inform training and development.

Performance evaluation

Most sport associations apply evaluations for the performance of their sports officials. However, those evaluation programs have not always been validated, and the criteria for judging performance are not necessarily clear or obvious (MacMahon et al., 2015, see Chapter 9 on evaluations in different sports). Physical fitness (e.g., endurance, speed) is regularly evaluated, especially in physically demanding officiating sports such as ball games. Task-specific performance in these sports is measured with expert referee coaches who regularly watch referees during their games and judge them on different aspects of good referee performance. This includes mastery of the game rules (e.g., fouls/violations, handling of no calls) and the personal interactions (e.g., player/coach communication, teamwork, leadership). For example, an observational rating instrument for basketball referees (BARS-BR) evaluates their competencies on thirteen behavioral categories (Anshel, 1995). The resulting protocols are then used for feedback and also for scheduling referees for important games.

While evaluating, it is important to understand why an official made either an incorrect or correct decision and the factors that had an impact on this. This is where individual differences are particularly relevant, and we advocate for conversations and knowledge of individuals between the official and the evaluator, where this enhances the evaluation outcomes without bias (Raab et al., 2020). Therefore, flexibility is needed in the evaluation process for officials (MacMahon et al., 2015, see Referees calls in Chapter 9). This may involve assessments and decisions made *in context* for a particular official, wherein evaluations are made relative to the most recent performance or the stage in the season, with particular key influential factors of a performance accounted for (e.g., crowd behaviors). This may even consider key development factors, as mentioned previously, around lifestyle and psychological

skills, to account for progress of individual referees and link development to performance.

Performance measures in technical sports (e.g., gymnastics) are often based on theoretical tests, which include questions on rule-based knowledge and the analysis of videotaped athletic routines, similar to the demands at real competitions. The results are used for the evaluation of performance and, consequently, for the decision about whether a referee can move up to a higher judging level.

In conclusion, this review of the existing research on referees highlights that individual differences have been implied, but not directly addressed. It points out that individual differences are evident in skilled interactions, player management, and perceptual and decision making behavior – including decision process preference, positioning, and notably in the experiences and prior knowledge that influence choices and performance. These individual differences create a toolbox of strategies and constraints that can be selected and utilized flexibly, in order to lead to acceptable performance outcomes. Referees are responsible for managing contests as fair, entertaining vehicles to showcase athletic talents and abilities. We accept that any two different athletes can possess a different combination of skills, strengths, and approaches to performance, yet be equally valued for their capabilities and performance level. We argue for applying the same perspective to refereeing, despite the long-held view that there is one correct set of behaviors, and an emphasis on one accuracy measure of performance. This more nuanced view advocates that individual approaches, for example in communication style, can lead to similarly acceptable performance levels.

Applying this philosophy to the behavior of referees calls for a broader research focus on individual differences, which on a theoretical level shed light on variations in

behavior. This perspective also acknowledges more than just physical fitness or the detection of fouls – it emphasizes experience, interaction, communication, and positioning and the interplay between all of these variables and decision making. Methodologically, this point of view will enhance mean-oriented data descriptions to include individual difference measures, and on a practical level, advocates selecting, training and evaluating referees based on their individual differences.

The areas in which more research and practice will inform our selection, training, and evaluation are communication and personal communication style, with more flexibility in applying different refereeing styles to facilitate performance in different contexts. The main barriers to research in these areas are related to both the researchers (e.g. the emphasis on statistical significance in academic models) and the practitioners (officiating managers and/or officials themselves, e.g., primary interest in short-term outcomes). These barriers contribute to the smaller volume of work around individual differences in refereeing.

Future research should go beyond the current emphasis on descriptive, empirical studies that ignore intra- and inter-individual differences and support general recommendations for practice. We think that in order to overcome the research-practice gap, researchers and practitioners should ask questions about how to improve training considering the individual, and research may need to design studies that more specifically answer these questions. For example, if practitioners want to know if they need team training instead of current individual referee training, a study testing two groups of referees after training in teams or individually would require predictions about responders and non-responders and amount of skill development via different training programs and sufficient measures of performance. At the end of these bidirectional research-practice enterprises we may find even more delight when

referees are part of a fair competition, with the athletic skills and competition on show, together with their own refereeing skills. Allowing for, understanding, and enhancing the nuance of individual differences in officiating, to optimize overall performance, as is accepted in the realm of athlete research, will enhance the sports we all love to watch. Who knows? An individual difference approach may even enhance the public's appreciation for individual officials, their particular skill sets, and their critical role in competition, helping officials receive the support they deserve.

References

Adie, J. M., Renshaw, I., Polman, R., Thompson, M. B., & Mann, D. L. (2020). When in doubt, it's not out: Match format is associated with differences in elite-level cricket umpires' leg-before-wicket decisions. *Psychology of Sport and Exercise, 51,* Article 101760. Ahead of print. Published online 28 July 2020. doi: 10.1016/j.psychsport.2020.101760

Anshel, M. H. (1995). Developing of a rating scale for determining competence in basketball referees: Implications for sport psychology. *The Sport Psychologist, 9*(1), 4-28. doi: 10.1123/tsp.9.1.4

Anshel, M. H., & Weinberg, R. S. (1999). Re-examining coping among basketball referees following stressful events: Implications for coping interventions. *Journal of Sport Behavior, 22*(2), 141-161.

Ansorge, C. J., & Scheer, J. K. (1988). International bias detected in judging gymnastic competition at the 1984 Olympic games. *Research Quarterly for Exercise and Sport, 59*(2), 103-107. doi: 10.1080/02701367.1988.10605486

Baldo, M. V. C., Ranvaud, R. D., & Morya, E. (2002). Flag errors in soccer games: The flash-lag effect brought to real life. *Perception, 31*(10), 1205-1210. doi: 10.1068/p3422

Balmer, N. J., Nevill, A. M., Lane, A. M., Ward, P., Williams, A. M., & Fairclough, S. (2007). Influence of crowd noise on soccer refereeing consistency in soccer. *Journal of Sport Behavior, 30*(2), 130-145.

Bard, C., Fleury, M., Carriére, L., & Hallé, M. (1980). Analysis of gymnastics judges' visual search. *Research Quarterly For Exercise and Sport, 51*(2), 267-273. doi: 10.1080/02701367.1980.10605195

Bar-Eli, M., Plessner, H., & Raab, M. (2011). *Judgment, decision Making and success in sport*. Wiley-Blackwell.

Bar-Eli, M., Tenenbaum, G., & Geister, S. (2006). Consequences of players' dismissal in professional soccer: A crisis-related analysis of group-size effects. *Journal of Sports Sciences, 24*(10), 1083-1094. doi: 10.1080/02640410500432599

Bordner, S. S. (2019). Why you don't have to choose between accuracy and human officiating (but you might want anyway). *Philosophies, 4*(2), 33-41. doi: 10.3390/philosophies4020033

Boyko, R. H., Boyko, A. R., & Boyko, M. G. (2007). Referee bias contributes to the home advantage in English Premiership football. *Journal of Sports Sciences, 25*(11), 1185-1194. doi: 10.1080/02640410601038576

Brand, R., & Neß, W. (2004). Regelanwendung und Game-Management: Qualifizierende Merkmale von Schiedsrichtern in Sportspielen [Rule administration and game management: Qualifying characteristics of referees in sport games]. *Zeitschift für Sportpsychologie* [*Journal for Sport Psychology*], *11*(4), 127-136. doi: 10.1026/1612-5010.11.4.127

Brand, R., Plessner, H., & Schweizer, G. (2009). Conceptual considerations about the development of a decision-making training method for expert soccer referees. In D. Araújo, H. Ripoll, & M. Raab (Eds.), *Perspectives on cognition and action in sport* (pp. 181-190). Nova Science Publishers, Inc.

Brand, R., Plessner, H., & Unkelbach, C. (2008). Basic psychological processes underlying referees' decision-Making. In P. Andersson, P. Ayton, & C. Schmidt (Eds.), *Myths and facts about football: The economics and psychology of the world's greatest sport* (pp.173-187). Cambridge Scholars Press.

Brand, R., Schmidt, G., & Schneeloch, Y. (2006). Sequential effects in elite basketball referees' foul decisions: An experimental study on the concept of game management. *Journal of Sport and Exercise Psychology, 28*(1), 93-99. doi: 10.1123/jsep.28.1.93

Burleson, B. R. (2007). Constructivism: A general theory of communication skill. In B. B. Whaley, & W. Samter (Eds.), *Exploring communication: Contemporary theories and exemplars* (pp. 105-128). Lawrence Erlbaum Associates.

Burns L, Weissensteiner J. R., & Cohen, M. (2019). Lifestyles and mindsets of Olympic, Paralympic and world champions: is an integrated approach the key to elite performance? *British Journal of Sports Medicine, 53*(13), 818-824. doi: 10.1136/bjsports-2018-099217

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432-459. doi: 10.1037/0033-295X.100.3.432

Castagna, C., Abt, G., & D'Ottavio, S. (2007). Physiological aspects of soccer refereeing performance and training. *Sport Medicine, 37*(7), 625-646. doi: 10.2165/00007256-200737070-00006

Catteeuw, P., Gilis, B., Jaspers, A., Wagemans, J., & Helsen, W. F. (2010a). Training of perceptual-cognitive skills in offside decision making. *Journal of Sport and Exercise Psychology, 32*(6), 845-861. doi: 10.1123/jsep.32.6.845

Catteeuw, P., Gilis, B., Wagemans, J., & Helsen, W. F. (2010b). Offside decision making of assistant referees in the English Premier League: Impact of physical and perceptual-cognitive factors on match performance. *Journal of Sports Sciences, 28*(5), 471-481. doi: 10.1080/02640410903518184

Catteeuw, P., Gilis, B., Garcia-Aranda, J. M., Tresaco, F., Wagemans, J., & Helsen, W. F. (2010c). Offside decision making in the 2002 and 2006 FIFA World Cups. *Journal of Sports Sciences, 28*(10), 1027-1032. doi: 10.1080/02640414.2010.491084

Collins, H. (2010). The philosophy of umpiring and the introduction of decision-aid technology. *Journal of the Philosophy of Sport, 37*(2), 135-146. doi: 10.1080/00948705.2010.9714772

Coulomb-Cabagno, G., Rascle, O., & Souchon, N. (2005). Players' gender and male referees' decisions about aggression in French soccer: A preliminary study. *Sex Roles, 52*(7/8), 547-553. doi: 10.1007/s11199-005-3720-z

Cunningham, I., Mellick, M., Mascarenhas, D. R. D., & Fleming, S. (2012). Decision making and decision communications in elite rugby union referees: An inductive investigation. *Sport and Exercise Psychology Review, 18*(2), 23-34.

Cunningham, I., Simmons, P., Mascarenhas, D. R. D., & Redhead, S. (2014). Skilled interaction: Concepts of communication and player management in the development of sports officials. *International Journal of Sport Communication, 7*(2), 166-187. doi: 10.1123/IJSC.2013-0098

Dawson, P., Dobson, S., Goddard, J., & Wilson, J. (2007). Are football referees really biased and inconsistent? evidence on the incidence of disciplinary sanction in the English Premier League. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(1), 231-250. doi: 10.1111/j.1467-985X.2006.00451.x

de Oliveira, M. C., Orbetelli, R., & de Barros Neto, T. L. (2011). Call accuracy and distance from the play: A study with Brazilian soccer referees. *International Journal of Exercise Science, 4*(1), 30-38.

Diederich, A., & Trueblood, J. (2018). A dynamic dual process model of risky decision making. *Psychological Review, 125*(2), 270-292. doi: 10.1037/rev0000087

Dosseville, F., Laborde, S., & Garncarzyk, C. (2014). Current research in sports officiating and decision-making. In C. Mohiyeddini (Ed.), *Contemporary topics and trends in the psychology of sports* (pp. 13-38). Nova Publishers.

Elsworthy, N., Burke, D., Scott B. R., Stevens, C., & Dascombe, B. (2014). Physical and decision-making demands of Australian football umpires during competitive matches. *Journal of Strength and Conditioning Research, 28*(12), 3502-3507. doi: 10.1519/JSC.0000000000000567

Erikstad, M. K., & Johansen, B. T. (2020). Referee bias in professional football: Favoritism toward successful teams in potential penalty situations. *Frontiers in. Sports and Active Living, 2*, Article 19. doi: 10.3389/fspor.2020.00019

Fasold, F., Memmert, D., & Unkelbach, C. (2012). Extreme judgments depend on the expectation of following judgments: A calibration analysis. *Psychology of Sport and Exercise*, *13*(2), 197-200. doi: 10.1016/j.psychsport.2011.11.004

Ford, G. G., Gallagher, S. H., Lacy, B. A., Bridwell, A. M., & Goodwin, F. (1997). Repositioning the home plate umpire to provide enhanced perceptual cues and more accurate ball-strike judgments. *Journal of Sport Behavior, 22*(1), 28-44.

Ford, G. G., Goodwin, F., & Richardson, J. W. (1995). Perceptual factors affecting the accuracy of ball and strike judgments from traditional American League and National League umpire perspectives. *International Journal of Sport Psychology, 27*(1), 50-58.

Frank, M. G., & Gilovich, T. (1988). The dark side of self- and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology, 54*(1), 74-85. doi: 10.1037/0022-3514.54.1.74

García-Santos, D., Gómez-Ruano, M. A., Vaquera, A., & Ibáñez, S. J. (2020). Systematic review of basketball referees' performances. *International Journal of Performance Analysis in Sport, 20*(3), 495-533. doi: 10.1080/24748668.2020.1758437

Ghasemi, A., Momeni, M., Rezaee, M., & Gholami, A. (2009). The difference in visual skills between experts versus novice soccer referees. *Journal of Human Kinetics, 22*(1), 15-20. doi: 10.2478/v10078-009-0018-1

Ghasemi, A., Momeni, M., Jafarzadehpur, E., Rezaee, M., & Taheri, H. (2011). Visual skills involved in decision making by expert referees. *Perceptual and Motor Skills, 112*(1), 161-171. doi: 10.2466/05.22.24.27.PMS.112.1.161-171

Gilis, B., Helsen, W., Catteeuw, P., Van Roie, E., & Wagemans, J. (2009). Interpretation and application of the offside law by expert assistant referees: Perception of spatial positions in complex dynamic events on and off the field. *Journal of Sports Sciences, 27*(6), 551-563. doi: 10.1080/02640410802702178

Gilis, B., Helsen, W., Catteeuw, P., & Wagemans, J. (2008). Offside decisions by expert assistant referee in association football: Perception and recall of spatial positions in complex dynamic events. *Journal of Experimental Psychology: Applied, 14*(1), 21-35. doi: 10.1037/1076-898X.14.1.21

Gilis, B., Weston, M., Helsen, W. F., Junge, A., & Dvorak, J. (2006). Interpretation and application of the Laws of the Game in football incidents leading to player injuries. *International Journal of Sport Psychology, 37*(2), 121-138.

Greenlees, I., Leyland, A., Thelwell, R., & Filby, W. (2008). Soccer penalty takers' uniform color and pre-penalty kick gaze affect the impressions formed of them by opposing goalkeepers. *Journal of Sports Sciences, 26*(6), 569-576. doi: 10.1080/02640410701744446

Goldstein, W. M. (2004). Social judgment theory: Applying and extending Brunswik's probabilistic functionalism. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision-making* (pp. 37-61). Blackwell Publishing.

Guillén, F, & Feltz, D. L. (2011). A conceptual model of referee efficacy. *Frontiers in Psychology, 2*, Article 25. doi: 10.3389/fpsyg.2011.00025

Hagemann, N, Strauss, B., & Leißing, J. (2008). When the referee sees red…. *Psychological Science, 19*(8), 769-771. doi: 10.1111/j.1467-9280.2008.02155.x

Hancock, D. J., & Ste-Marie, D. M. (2013). Gaze behaviors and decision making accuracy of higher- and lower-level ice hockey referees. *Psychology of Sport and Exercise, 14*(1), 66-71. doi: 10.1016/j.psychsport.2012.08.002

Hartmann, O. (1991). *Der Schiedsrichter im DTB* [*The referee in the DTB*]. Sindelfingen: Schmidt & Dreisilker.

Helsen, W., & Bultynck, J. B. (2004). Physical and perceptual-cognitive demands of top-class refereeing in association football. *Journal of Sports Sciences, 22*(2), 179-189. doi: 10.1080/02640410310001641502

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

Helsen, W., MacMahon, C., & Spitz, J. (2019). Decision making in match officials and judges. In M. Williams, & R. Jackson (Eds.), *Anticipation and Decision-Making in Sport* (pp. 250-266). Routledge.

Hüttermann, S., Helsen, W. F., Put, K., & Memmert, D. (2018). Does visual attention impact on decision making in complex dynamic events? *Journal of Sport and Exercise Psychology, 40*(3), 163-166. doi: 10.1123/jsep.2017-0169

Jendrusch, G., Schmidt, O., Wilke, G., & Marées, H. de (1993). Zur visuellen Leistungsfähigkeit von Handball-Schiedrichtern [Visual performance of handball referees]. In H.-F. Voigt (Ed.), *Bewegungen lesen und antworten* [*Reading and responses to movements*] (pp. 73-87). Ahrensburg: Czwalina.

Johnson, C., & Jason, T. (2016). Rejecting technology: A normative defense of fallible officiating. *Sport, Ethics and Philosophy, 10*(2), 148-160. doi: 10.1080/17511321.2016.1152287

Jones, M. V., Paull, G. C., & Erskine, J. (2002). The impact of a team's aggressive reputation on the decisions of association football referees. *Journal of Sports Sciences, 20*(12), 991-1000. doi: 10.1080/026404102321011751

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment: In T. Gilovich, D. Griffin, & D. Kahneman

(Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge University Press.

Kellett, P., & Shilbury, D. (2007). Umpire participation: Is abuse really the issue? *Sport Management Review, 10*(3), 209-229. doi: 10.1016/S1441-3523(07)70012-8

Kittel, A., Elsworthy, N., & Spittle, M. (2019). Incorporating perceptual decision-making training into high-intensity interval training for Australian football umpires. *Journal of Sports Sciences 37*(1), 29-35. doi: 10.1080/02640414.2018.1480257

Kittel, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019). Identification of key performance characteristics of elite Australian football umpires. *International Journal of Sports Science & Coaching, 14*(4), 490-497. doi: 10.1177/1747954119845071

Krustrup, P., & Bangsbo, J., (2001). Physiological demands of top-class soccer refereeing in relation to physical capacity: effect of intermittent exercise training. *Journal of Sports Sciences, 19*(11), 881-891. doi: 10.1080/026404101753113831

Krustrup, P., Mohr, M., & Bangsbo, J. (2002). Activity profile and physiological demands of top-class soccer assistant refereeing in relation to training status. *Journal of Sports Sciences, 20*(11), 861-871. doi: 10.1080/026404102320761778

Krustrup, P., Helsen, W., Randers, M. B., Christensen, J. F., MacDonald, C., Rebelo, A. N., & Bangsbo, J. (2009). Activity profile and physical demands of football referees and assistant referees in international games. *Journal of Sports Sciences, 27*(11), 1167-1176. doi: 10.1080/02640410903220310

Larkin, P., Mesagno, C., Berry, J., & Spittle, M. (2018). Exploration of the perceptual-cognitive processes that contribute to in-game decision-making of

Australian football umpires. *International Journal of Sport and Exercise Psychology, 16*(2), 112-124. doi: 10.1080/1612197X.2016.1167760

Livingston, L., Forbes, S. L., Wattie, N., & Cunningham, I. (2020). *Sport officiating: Recruitment, development, and retention*. Routledge.

MacMahon, C., Helsen, W. F., Starkes, J. L., & Weston, M. (2007). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences, 25*(1), 65-78. doi: 10.1080/02640410600718640

MacMahon, C., Mascarenhas, D., Plessner, H., Pizzera, A., Oudejans, R. R. D., & Raab, M. (2015). *Sports officials and officiating: Science and practice*. Routledge.

MacMahon, C., & Mildenhall, B. (2012). A practical perspective on decision making influences in sports officiating. *International Journal of Sports Science & Coaching, 7*(1), 153-165. doi: 10.1260/1747-9541.7.1.153

MacMahon, C., & Plessner, H. (2013). The sports official in research and practice. In D. Farrow, J. Baker, & C. MacMahon (Eds.), *Developing sport expertise: Researchers and coaches put theory into practice* (2nd ed., pp. 71-95). Routledge.

MacMahon, C., Starkes, J. L., & Deakin, J. (2009). Differences in processing of game information in basketball players, coaches and referees. *International Journal of Sport Psychology, 40*(3), 403-423.

MacMahon, C., & Ste-Marie, D. M. (2002). Decision-making by experienced rugby referees: Use of perceptual information and episodic memory. *Perceptual and Motor Skills, 95*(2), 570-572. doi: 10.2466/pms.2002.95.2.570

Mallo, J., Frutosa, P. J., Juáreza, D., & Navarroa, E. (2012). Effect of positioning on the accuracy of decision making of association football top-class referees and

assistant referees during competitive matches. *Journal of Sport Sciences, 30*(13), 1437-1445. doi: 10.1080/02640414.2012.711485

Maruenda, B. E. (2005). Can the human eye detect an offside position during a football match? *British Medical Journal, 330*(7484), 188-190. doi: 10.1136/bmj.330.7484.188

Mascarenhas, D. R. D. (April 2006). The psychology of officiating. Southland NZFU Referees' Meeting. Southland, New Zealand.

Mascarenhas, D. R. D., Button, C., O'Hare, D., & Dicks, M. (2009). Physical performance and decision making in association football referees: A naturalistic study. *The Open Sports Sciences Journal, 2*(1), 1-9. doi: 10.2174/1875399X00902010001

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. W. (2002). The art of reason versus the exactness of science in elite refereeing: Comments on Plessner and Betsch (2001). *Journal of Sport and Exercise Psychology, 24*(3), 328-333. doi: 10.1123/jsep.24.3.328

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. W. (2005a). A naturalistic approach to training accurate and coherent decision making in rugby union referees. *The Sport Psychologist, 19*(2), 131-147. doi: 10.1123/tsp.19.2.131

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. W. (2005b). Elite refereeing performance: Developing a model for sport science support. *The Sport Psychologist, 19*(4), 364-379. doi: 10.1123/tsp.19.4.364

Mascarenhas, D. R. D., O'Hare, D., & Plessner, H. (2008). The psychological and performance demands of association football refereeing. *International Journal of Sport Psychology, 37*(2), 99-120.

McCarrick, D., Brewer, G., Lyons, M., Pollet, T. V., & Neave, N. (2020). Referee height influences decision making in British football leagues. *BMC Psychology, 8*(1), Article 4. doi: 10.1186/s40359-020-0370-4

Mellick, M. C., Bull, P. E., Laugharne, E. J., & Fleming, S. (2005). Identifying best practice for referee decision communication in association and rugby union football: A microanalytic approach. *Football Studies, 8*(1), 42-57.

Mellick, M. C., Fleming, S., & Davies, G. (2007). An interpretive analysis of interpersonal communication: A case study from elite rugby union match officiating. *International Journal of Performance Analysis in Sport, 7*(2), 92-105. doi: 10.1080/24748668.2007.11868399

Memmert, D., Unkelbach, C., Ertmer, J., & Rechner, M. (2008). Gelb oder kein Gelb? Persönliche Verwarnungen im Fußball als Kalibrierungsproblem [To award or not to award a yellow card? Personal warnings in soccer as a calibration problem]. *Zeitschift für Sportpsychologie* [*Journal for Sport Psychology*]*, 15*(1), 1-11. doi: 10.1026/1612-5010.15.1.1

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med, 6*(7): e1000097. doi: 10.1371/journal.pmed1000097

Nelson-Jones, R. (2012). *Introduction to counselling skills: Text and activities, 4th Edition*. Thousand Oaks, CA: Sage.

Nevill, A. M., Balmer, N. J., & Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise, 3*(4), 261-272. doi: 10.1016/S1469-0292(01)00033-4

Newell, K. M. (1986). Constraints on the development of coordination. In M. G. Wade, & H. T. A. Whiting (Eds.), *Motor development in children: Aspects of*

coordination and control (pp. 341-360). Martinus Nijhoff, Dordrecht. doi: 10.1007/978-94-009-4460-2_19

Niebuhr, J., Plessner, H., & Wittbrodt, J. (1998). *Lehrvideo für das Kunstturnen der Männer zum Code de Pointage '97 [Video]* [*Educational video for men's gymnastics to the Code de Pointage '97*]. Hannover: NTB.

Nijhawan, R. (2001). The flash-lag phenomenon: Object motion and eye movements. *Perception, 30*(3), 263-282. doi: 10.1068/p3172

O'Connor, D., Wardak, D., Goodyear, P., Larkin, P., & Williams, A. M. (2018). Conceptualising decision-making and its development: a phenomenographic analysis. *Science and Medicine in Football, 2*(4), 261-271. doi: 10.1080/24733938.2018.1472388

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

Oudejans, R. R. D., 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 of their offside judgments during actual match play. *International Journal of Sport Psychology, 36*(1), 3-21.

Oudejans, R. R. D., 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

Page, K., & Page, L. (2010). Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, *31*(2), 192-199. doi: 10.1016/j. joep.2009.08.007

Paradis, K., Larkin, P., & O'Connor, D. (2016). The effects of physical exertion on decision-making performance of Australian football umpires. *Journal of Sports Sciences, 34*(16), 1535-1541. doi: 10.1080/02640414.2015.1122205

Paull, G., & Glencross, D. (1997). Expert perception and decision-Making in baseball. *International Journal of Sport Psychology, 28*(1), 35-56.

Philippe, F. L. Vallerand, R. J., Andrianarisoa, J., & Brunel, P. (2009). Passion in referees: Examining their affective and cognitive experiences in sport situations. *Journal of Sport and Exercise Psychology, 31*(1), 77-96. doi: 10.1123/jsep.31.1.77

Pizzera, A., Möller, C., & Plessner, H. (2018). Gaze behavior of gymnastics judges: Where do experienced judges and gymnasts look while judging? *Research Quarterly for Exercise and Sport, 89*(1), 112-119. doi: 10.1080/02701367.2017.1412392

Pizzera, A., & Raab, M. (2012a). Perceptual judgments of sports officials are influenced by their motor and visual experience. *Journal of Applied Sport Psychology, 24*(1), 59-72. doi: 10.1080/10413200.2011.608412

Pizzera, A., & Raab, M. (2012b). Does motor or visual experience enhance the detection of deceptive movements in football? *International Journal of Sports Science and Coaching, 7*(2), 269-283. doi: 10.1260/1747-9541.7.2.269

Plessner, H. (1997). *Urteilsverzerrungen bei Kampfrichtern im Kunstturnen – Der Einfluß von Erwartungen* [*Judgement distortions of judges in artistic gymnastics – The influence of expectations*]. Aachen: Shaker.

Plessner, H. (1999). Expectation biases in gymnastics judging. *Journal of Sport and Exercise Psychology, 21*(2), 131-144. doi: 10.1123/jsep.21.2.131

Plessner, H., & Betsch, T. (2001). Sequential effects in important referee decisions: The case of penalties in soccer. *Journal of Sport and Exercise Psychology, 23*(3), 254-259. doi: 10.1123/jsep.23.3.254

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognition perspective. *Psychology of Sport and Exercise, 7*(6), 555-575. doi: 10.1016/j.psychsport.2006.03.007

Plessner, H., & Raab, M. (1999). Kampf- und Schiedsrichterurteile als Produkte sozialer Informationsverarbeitung [Umpire and referee judgments as products of social information processing]. *Psychologie und Sport* [*Psychology and Sport*]*, Schorndorf 6, Heft 4*, 130-145.

Plessner, H., & Schallies, E. (2005). Judging the cross on rings: A matter of achieving shape constancy. *Applied Cognitive Psychology, 19*(9), 1145-1156. doi: 10.1002/acp.1136

Plessner, H., Schweizer, G., Brand, R., & O'Hare, D. (2009). A multiple-cue learning approach as the basis for understanding and improving soccer referees' decision-Making. In M. Raab, J. Johnson, & H. Heekeren (Eds.), *Progress in brain research, Mind and motion: The bidirectional link between thought and action* (pp. 151-158). Elsevier Press.

Raab, M., Avugos, S., Bar-Eli, M., & MacMahon, C. (2020). The referee's challenge: A threshold process model for decision making in sport games. *International Review of Sport and Exercise Psychology.* Published online 09 July 2020. doi: 10.1080/1750984X.2020.1783696

Renden, P. G., Kerstens, S., Oudejans, R. R. D., & Cañal-Bruland, R. (2014). Foul or dive? Motor contributions to judging ambiguous foul situations in football.

European Journal of Sport Science, 14(Suppl. 1), S221-227. doi: 10.1080/17461391.2012.683813

Reilly, T., & Gregson, W. (2006). Special populations: The referee and assistant referee. *Journal of Sports Sciences, 24*(7), 795-801. doi: 10.1080/02640410500483089

Royce, R. (2012). Refereeing and technology – Reflections on Collins' proposals. *Journal of the Philosophy of Sport, 39*(1), 53-64. doi: 10.1080/00948705.2012.675066

Samuel, R. D., Tenenbaum, G., & Galily, Y. (2020). An integrated conceptual framework of decision-making in soccer refereeing, *International Journal of Sport and Exercise Psychology*. Ahead of print. Published online 22 May 2020. doi: 10.1080/1612197X.2020.1766539

Scheer, J. K., Ansorge, C. J., & Howard, J. (1983). Judging bias induced by viewing contrived videotapes: a function of selected psychological variables. *Journal of Sport Psychology, 5*(4), 427-437. doi: 10.1123/jsp.5.4.427

Schweizer, G., & Plessner, H. (2016). The accuracy-adequacy model: A theoretical perspective for understanding referees' decisions. *Research Quarterly for Exercise and Sport, 87*(S1), S82.

Schweizer, G., Plessner, H., Kahlert, D., & Brand, R. (2011). A video-based training method for improving soccer referees' intuitive decision making skills. *Journal of Applied Sport Psychology, 23*(4), 429-442. doi: 10.1080/10413200.2011.555346

Schwarz, W. (2011). Compensating tendencies in penalty kick decisions of referees in professional football: Evidence from the German Bundesliga 1963-2006. *Journal of Sports Sciences, 29*(5), 441-447. doi: 10.1080/02640414.2010.538711

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P. et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results [Online version published by the Department of Economics, University of Maryland, under the title: Crowdsourcing data analysis: Do soccer referees give more red cards to dark skin toned players?]. *Advances in Methods and Practices in Psychological Science, 1*(3), 337-356. doi: 10.1177/2515245917747646

Simmons, P. (2010). Communicative displays as fairness heuristics: Strategic football referee communication. *Australian Journal of Communication, 37*(1), 75-94.

Slack, L. A., Maynard, I. W., Butt, J., & Olusoga, P. (2013). Factors underpinning football officiating excellence: Perceptions of English Premier League referees. *Journal of Applied Sport Psychology, 25*(3), 298-315. doi: 10.1080/10413200.2012.726935

Souchon, N., Cabagno, G., Rascle, O., Traclet, A., Dosseville, F., & Maio, G. R. (2009). Referees' decision making about transgressions: The influence of player gender at the highest national level. *Psychology of Women Quarterly, 33*(4), 445-452. doi: 10.1111/j.1471-6402.2009.01522.x

Souchon, N., Livingstone, A. G., & Maio, G. R. (2013). The influence of referees' expertise, gender, motivation, and time constraints on decisional bias against women. *Journal of Sport and Exercise Psychology, 35*(6), 585-599. doi: 10.1123/jsep.35.6.585

Spitz, J., Put, K., Wagemans, J., Williams, A. M., & Helsen, W. F. (2016). Visual search behaviors of association football referees during assessment of foul play situations. *Cognitive Research: Principles and Implication, 1*(1), Article 12. doi: 10.1186/s41235-016-0013-8

Spitz, J., Wagemans, J., Memmert, D., Williams, A. M., & Helsen, W. F. (2020). Video assistant referees (VAR): The impact of technology on decision making in association football referees. *Journal of Sport Sciences*, *39*(2), 147-153. doi: 10.1080/02640414.2020.1809163

Ste-Marie, D. M. (1999). Expert-novice differences in gymnastic judging: An information-processing perspective. *Applied Cognitive Psychology, 13*(3), 269-281. doi: 10.1002/(SICI)1099-0720(199906)

Ste-Marie, D. M. (2000). Expertise in women's gymnastics judging: An observation approach. *Perceptual and Motor Skills, 90*(2), 543-546. doi: 10.2466/PMS.90.2.543-546

Ste-Marie, D. M. (2003a). Expertise in sport judges and referees: Circumventing information-processing limitations. In J. L. Starkes, & K. A. Ericsson (Eds.), *Expert performance in sport: Advances in research on sport expertise* (pp. 169-190). Champaign, IL: Human Kinetics.

Ste-Marie, D. M. (2003b). Memory biases in gymnastic judging: Differential effects of surface feature changes. *Applied Cognitive Psychology, 17*(6), 733-751. doi: 10.1002/acp.897

Ste-Marie, D. M., & Valiquette, S. M. (1996). Enduring memory-influenced biases in gymnastic judging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1498-1502. doi: 10.1037/0278-7393.22.6.1498

Stone, J., Perry, Z. W., & Darley, J. M. (1997). "White men can't jump": Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology, 19*(3), 291-306. doi: 10.1207/s15324834basp1903_2

Stride, C., Patterson, M., & Thomas, F. (September, 2011). Cheating in soccer: Team culture, player behavior, or just a question of circumstance? New England Symposium on Statistics in Sports (NESSIS). Cambridge, Mass.

Sutter, M., & Kocher, M. G. (2004). Favoritism of agents – the case of referees' home bias. *Journal of Economic Psychology, 25*(4), 461-469. doi: 10.1016/S0167-4870(03)00013-8

Tiryaki, M. S. (2005). Assessing whether black uniforms affect the decisions of Turkish soccer referees: Is finding of Frank and Gilovich's study valid for Turkish culture? *Perceptual and Motor Skills, 100*(1), 51-57. doi: 10.2466/pms.100.1.51-57

Unkelbach, C., & Memmert, D. (2008). Game-management, context-effects and calibration: The case of yellow cards in soccer. *Journal of Sport and Exercise Psychology, 30*(1), 95-109. doi: 10.1123/jsep.30.1.95

Unkelbach, C., & Memmert, D. (2010). Crowed noise as a cue in referee decisions contributes to the home advantage. *Journal of Sport and Exercise Psychology, 32*(4), 483-498. doi: 10.1123/jsep.32.4.483

Unkelbach, C., & Memmert, D. (2014). Serial-position effects in evaluative judgments. *Current Directions in Psychological Science, 23*(3), 195-200. doi: 10.1177/0963721414533701

Unkelbach, C., Ostheimer, V., Fasold, F., & Memmert, D. (2012). A calibration explanation of serial position effects in evaluative judgments. *Organizational Behavior and Human Decision Processes, 119*(1), 103-113. doi: /10.1016/j.obhdp.2012.06.004

van Quaquebeke, N., & Giessner, S. R. (2010). How embodied cognitions affect judgments: Height-related attribution bias in football foul calls. *Journal of Sport and Exercise Psychology, 32*(1), 3-22. doi: 10.1123/jsep.32.1.3

Voight, M. (2009). Sources of stress and coping strategies of US soccer officials. *Stress and Health, 25*(1), 91-101. doi: 10.1002/smi.1231

Weston, M., Castagna, C., Impellizzeri, F. M., Rampinimi, E., & Breivik, S. (2010). Aging and physical match performance in English Premier League soccer referees. *Journal of Science and Medicine in Sport, 13*(1), 96-100. doi: 10.1016/j.jsams.2008.07.009