Sport arbitration as an emergent process in a complex system: Decision-making variability is a marker of expertise in national-level football referees

RUSSELL, S, RENSHAW, I and DAVIDS, Keith <http://orcid.org/0000-0003-1398-6123>

Sport arbitration as an emergent process in a complex system: Decision-making variability is a marker of expertise in national-level football referees

Abstract

This study examined the experiential knowledge of eight Australian national-level football referees (3.2 yrs mean national-level experience) about the notion of consistent decision-making during competitive matches. Using a grounded theory approach, the analysis revealed that participants view 'consistency' as context-dependent, rather than a rigid process of uniformly responding to isolated foul-play transgressions with putatively correct responses. Our results present two key conceptual abstractions - 'referential and game dependent' and 'purpose and context' – as a framework for understanding referee decision-making consistency. Data indicated that these performance intentions for consistency directed referee attention (attunement) to key contextual information that is necessary to prospectively control player behaviours and thus, the emergent trajectory of each game. Results support the view that consistent decision-making performance is an emergent process defined by the decision-making actions of the referee, and exploitation of specifying contextual game factors. Our discussion outlines how this process can be understood as dynamical transactions within a complex system (i.e. competitive football game), in which varying decision-making responses to superficially similar incidents is a marker of expertise, rather than inconsistency. These findings draw attention to the limitations of isolated foul-play video assessment for training and advocate for more representative game opportunities for referees to practice making decisions with key contextual information sources present. Future research could explore how significantly specific performance goals and contextual factors interact to shape emergent decision-making choices across different sports.

Keywords: ecological dynamics, emergent decision making, grounded theory, complex systems, qualitative research, football, referees.

Lay summary: Australian national-level referees consider consistent decision making as more than just the similar identification of fouls. These higher-level referees draw on contextual factors to institute a foul 'standard' unique to each game, in the interests of maintaining each game's control and integrity. Referees portray consistency as remaining faithful to the reference points they have co-developed with the players and the conditions that defined them.

Implications for practice:

- Assessment of decision-making accuracy and/or consistency should include key contextual factors.
- Higher-level referee training should focus less on 'what the foul is' and more on 'what the foul can offer' in terms of running the game.
- Modified game-based opportunities for referees to practice making decisions, shaped by 'context' and at game speed, would encourage referees to explore a broad range of decision-making responses to superficially similar incidents.
- Performance reflections about how the referee's decision making has contributed to the context of the game (i.e. reference points) would stimulate consideration of what decision-making options remain available to positively influence each game's trajectory.
- These findings draw attention to the limitations of isolated foul-play video assessment for training and advocate for more representative game opportunities for referees to practice making calls with key information sources present.

Sport arbitration as an emergent process in a complex system: Decision-making variability is a marker of expertise in national-level football referees

Introduction

Historically the referee's role has been portrayed as "implementing the rules of the game and... to keep up with play to be in a good position to notice infringements" (Reilly & Gregson, 2006, p. 795). This description of their role implies that the main task of a referee is to make decisions on fouls and player misconduct, as though they are judgements of fact (Mascarenhas et al., 2006). The media spotlight has tended to consolidate this view of foul judgements, emphasising that single decisions by referees can "greatly impact on the outcome of the match, leading to criticism and impacting club revenue" (Kittle et al., 2019, p. 261). Consequently there has been significant support for video technology to aid referee decision-making framed as a more "objective representation and comparative standard… to increase decisional accuracy" (Spitz et al., 2018, p. 1). As such, many research protocols and governing bodies rely on panels of experts to agree on a single putatively correct interpretation of an incident. This is often reported statistically as an accuracy rating, with organisations viewing the attainment of higher accuracy scores as a marker of greater consistency and better performance (Pina et al., 2018).

Despite support for video-led interventions (for a review see Kittle et al., 2019), the level of agreement reported between international-level football officials ranges from 64% - 72.4%, only a moderate to high level of accuracy (Catteeuw et al., 2009; Mascarenhas et al., 2009). As standard, many of these testing protocols are designed to view fouls or incidents in isolation, separated from contextual factors like time of the match, game-score, position on the field of the perceived infringement and the previous decisions made by the arbitrator (e.g.

Plessner & Betsch, 2001; Plessner & Haar, 2006). The dominant rationale is that by minimising the impact of contextual determinants on cognitive serial sub-tasks of participants, the methodology maximises the individual's capacity to accurately detect and interpret the key input information needed to deliberate on the associated response (e.g. Anderson, 1983; Lweiwick et al., 1992; Masters et al., 2008). Thus, research protocols have sought to quantify the extent that particular contextual factors, such as external crowd noise (e.g. Unkelbach & Memmert, 2010), knowledge and infraction priming (e.g. MacMahon et al., 2007) and pre-event expectancies (e.g. Dosserville et al., 2011), may have on a "shift in the standard and reference point used for evaluation" from a pre-determined correct and accurate mean for each decision-making outcome (MacMahon & Starkes, 2008, p. 758).

This characterisation of consistent expert performance has been questioned by an ecological dynamics approach (e.g. Araújo et al., 2005; Passos et al., 2008), which argues that "in open, dynamic systems there is no 'best decision' since the most functional decision at any moment may compromise future decisions" (Araújo et al., 2017, p. 5). An example of this perspective is evidenced in MacMahon and Starkes' (2008) study of baseball umpires, players and coaches. They found that when pitches, conclusively judged as not a strike or ball, were presented to participants following video clips of definite balls and strikes, umpires, more than any other group, tended to call strikes. The authors suggested that calling strikes was an intentional strategy of "hastening the game" because strikes lead to batters being called 'out'. MacMahon and Starkes (2008) portrayed the use of this contextually specifying information as a *positive* function of expertise aimed at "minimising the cost of future borderline calls by forcing batters to swing and make future decisions earlier" (p. 759).

Other research however, has generally framed any shift in the distribution and/or islolated interpretation of fouls during a game as forms of either negative 'game circumstance bias' (i.e. taking into account the score of the game) or 'previous decision bias' (i.e.

sequential biasing). For example, Anderson and Pierce (2009) reported that when the home team in men's colleague basketball games is leading during a game, there is a 6.3% higher chance that the next foul would be awarded to the away team. They attributed this effect mainly to a referee's intention to appear to treat teams fairly, making it likely that fouls would even out during the match. Similarly, Schwarz (2011) analysed 12902 matches spanning 1963-2006 from the German Bundesliga and found that two penalties in a match occurred at a rate that was "larger than expected by chance" on the basis of independent penalty decisions (p. 441). Schwarz argued that these results provided further evidence that referees were using an "adaptive, equality-orientated decision rule that creates contingencies to balance things out" (p. 447). In both studies the investigation methods promote a theoretical viewpoint that incidents or penalties should be viewed independently of each other. ,. Moreover, they implied a fairer assessment of game incidents occurs when Prior penalties and previous events are treated by the referee as separate contextual factors, minimising their influence on the likelihood of future fouls or penalties.

In the field of refereeing research, this position has generally been countered from a game management perspective (e.g. Macarenhas et al., 2002), which contends that "good officials are mindful of the current game situation and make adjustments in their decision-making process to accommodate this" (MacMahon & Mildenhall, 2012, p. 157). For instance, Unkelbach and Memmert (2008) reported that referees are less likely to award a yellow card if an incident is perceived to have happened early in the game, even if it really occurred later in the match. The authors suggested that this represented an intentional overall game management strategy by referees to protect against instituting an overly severe judgement scale early in the game, which, if consistently applied throughout the remainder of the game would lead to an excessive distribution of yellow and potentially red cards. From a game management perspective, these differences in foul judgements related to game-time or prior

decisions from the referee are framed, not as inconsistency or bias, but rather "deliberate attempts to dynamically manage the game" (p. 96).

MacMahon and Mildenhall (2012) attempted to reconcile these assertions about the role of contextual factors on decision-making consistency, suggesting that use of context is "evidence of the human processing system filling in gaps in information to facilitate decision making" (p. 157). For example, in an offside situation in football, they may be rely on contextual information such as the player's running speed, to infer whether it was plausible for the player to ultimately be in an onside position. Conversely, they explain that certain contextual factors can work in a non-adaptive fashion, such as crowd noise to judge a foul translating to home team favour (e.g. Unkelbach & Memmert, 2010) or memories biases from pre-competition observation leading to higher routine scores in gymnastics judging (e.g. Ste-Marie, 2003). The implication is that these external or within-event contextual factors either: (i) negatively augment perception of existing foul information by distorting and/or biasing the accuracy of an incident or, (ii) positively work as a 'gap filler' when there is a deficit of perceptual information.

Rather than positioning context as a backdrop that clarifies or colours perception of an incident, ecological dynamicists view perception of events as an integral, exploratory process required for performers to attune to specifying information sources in the performance environment that may be more or less relevant to regulate actions (Araújo & Davids, 2016; Gibson, 1979). In this respect, properties of the performance environment that receive the highest attensity or notice are a "direct function of the psychological attitude assumed by the person" (Shaw, 1982, p. 218). In this way, Shaw (1982) underlined the deeply entwined relations between actions, cognitions and perception in human behaviour. Due to this entanglement, there is a continued need for investigations on how intentions specify attention to support decision-making actions in performance environments (for some existing examples

on sport performance see Connor et al., 2018; McCosker et al., 2019). An issue for researchers is that psychological attributes such as emotions, beliefs and intentions of performers are inherently subjective and qualitative, "making it challenging for science to apply to cognition, the kinds of explanatory techniques that have worked so successfully in studying other natural dynamical systems" (Araújo et al., 2019, p. 538). A useful method for overcoming this perceived challenge, is undertaking qualitative research that seeks to consider the perspective of the participants themselves, whilst they are engaged in performance.

Russell and colleagues (2019) have exemplified this, using a qualitative grounded theory approach, reporting that values and beliefs of national-level football referees directed perceptual attention when considering foul judgements. Adopting an ecological dynamics framework to discuss their findings, they argued that referees were drawn to features of the performance environment that supported key strategic decision-making goals of fairness, safety, accuracy and entertainment (termed the 'four pillars'), nested within over-arching, task-orientated goals (i.e. intentions) of 'maintaining control of the match' and 'preserving the game's integrity'. For instance, in a given context, if a referee recognises that player *safety* is becoming compromised, he/she might award a foul in the interests of maintaining match control. However on another occasion, if the referee considers that the game is under control, other opportunities in the performance environment may become more salient, such as playing an advantage to encourage *entertainment*. Rather than inconsistency, ecological dynamicists argue that "even if the context is exactly the same" variability in decision-making actions is "desirable and an advantage… that allows adaptability and functionality for different performance task goals, framed as intentions" (Chow et al., 2016, p. 18).

Despite this important progress, it is not well understood *how* and *in what ways* the state or context of the game may influence the decision-making priorities of referees with

high levels of expertise. For instance, questions arise about: under what conditions do referee decision-making choices vary and what is the nature of that context? Significantly, what might this information imply for interpretations and evaluation of referee decision-making with respect to expectations concerning performance consistency? To address these questions, the current study builds iteratively on the original 'four pillars' grounded theory proposed by Russell et al., (2019), by accessing experiential knowledge of national-level referees to better understand how their perceptions of 'consistency' may constrain and shape their strategic performance goals. We do not intend to present all facets of consistency or even suggest that how interviewed referees view consistency is correct but instead aim to scrutinise how various conceptions of consistency can have evaluative and behavioural consequences (Glăveanu, 2011). The aims of this study, therefore, were to conceptualise: (i) what decision-making consistency *is* (conceptualised by high level referees), and (ii), how their perspective of consistency manifests in decision-making actions as reported by the referees. The analysis ultimately seeks to generate substantive theoretical conceptions about decision-making consistency, from first hand insights of experienced referees, that can guide and inform methods used when researching refereeing decision-making expertise and to use in referee education and training programmes at higher levels.

Methodology

Grounded theory approaches argue that research, which is "counterintuitive… or challenges well-accepted ideas is often worthwhile" (Tracy, 2010, p. 840). Our study sought to reconsider long-held beliefs concerning what constitutes referee decision-making consistency. In the following sections, we outline how we used "systematic, yet flexible

methodology for collecting and analysing qualitative data" (Charmaz, 2006, p. 2) to produce theory grounded in data 'with and about' referees and their cultural contexts.

Background

The perspective of human behaviour adopted for the purposes of this study is *ecological*, that is, living things and their ecosystems are not seen as logically independent of each other (Araújo et al., 2019). This view of reality considers the nervous system, body, and surrounding environments as deeply integrated, 'open systems' which are continuously and simultaneously engaged in shaping the nature of one another. A behavioural framework that has attempted to explain this interdependent relationship is ecological dynamics. This approach, from ecological realism, also draws upon concepts and tools from dynamical systems theory and the complexity sciences to investigate specific measurable aspects of behaviour in order to predict a system's evolution over different timescales (Araújo et al., 2020). There is considerable evidence, however, that depending on the particular psychological attitude held "various [performers] do not perceive the same objects in the same way, and that even the same [performer] does not perceive the same object in the same way on different occasions" (Shaw, 1982, p. 210). More research is needed on human behaviours in performance contexts like sport to better understand how psychological attitudes constrain emergence of action(s) and decisions that will ultimately uniquely define the system's functionality (Araújo et al., 2017). Ecological dynamics draws upon many rich frameworks and guidelines for collecting and analysing data on human behaviours and experiences, using qualitative and quantitative methodologies (e.g. see Araujo et al., 2019; McCosker et al., 2019). Here, we provide a case that a qualitative grounded theory approach provides a methodological framework to consider "human experience holistically, contextually and as a dynamic process" (Demuth, 2015, p. 210). We also attempt to indicate how "methodological significance [can] emerge from the qualitative study of a concept that

has previous been examined… [primarily] quantitatively or experimentally" (Tracy, 2010, p. 846).

Grounded theory derivatives

When determining *how* we would undertake our grounded theory methodology, we considered: evidence from the original authors of each recognised derivative (e.g. Charmaz, 2006; Glaser & Strauss, 1967; Strauss & Corbin, 1990); those prominent in its critique in our field (e.g. Holt, 2016; Smith & McGannon, 2017; Weed, 2009) and; their subsequent recommendations (e.g. which led to the lead author attending the "The Kathy Charmaz Masterclasses", 2017). A significant focus of this literature is discussion concerning what grounded theory 'is', that is, whether it is a method or methodology, or a combination of techniques, procedures, principles and/or practices. Underlying this critique of its 'essence' is a view that a failure to contemplate ontological and epistemological assumptions of various approaches can lead to a diminishment of the truth and/or knowledge claims that grounded theory can offer (Demuth, 2015; Weed, 2017).

This philosophical unravelling of grounded theory has led to Glaser and Strauss' (1967) original statement to become positioned as a more objectivist and positivist version; due mostly to its emphasis on single truth to be discovered from data. Straussian and constructivist grounded theory approaches railed against the goal of research to "seek parsimonious explanations and generalisations devoid of context", and instead argued that the voices of both researchers and participants were integral considerations to the development of knowledge (Charmaz, 2008, p. 402; Strauss & Corbin, 1990). Despite ontological differences, ecological approaches share ground with this epistemological view of knowledge, suggesting that it is participant perception of "experiences that provide evidential bridges between knowing and doing… [and thus] a source of necessary a posteriori

knowledge about the world" (Shaw, 1982, p. 192). In this respect, accessing the perspective and opinion of referees "makes no attempt to answer the ontological question of what the environment is in any absolute sense (i.e., metaphysics), but rather attempts to answer the pragmatic question of what an environment means to [referees]" (Shaw, 1982, p. 196).

Worthy topics

One commonality between all the approaches is that the function of grounded theory is to generate theory or an analytic handle on experience(s) - grounded in data of the cultural group of interest (Weed, 2017). This view has at times led to a belief that 'true' grounded theory represents beginning with a completely blank slate or tabula rasa. Contrary to this, worthy topics (and data) often emerge from a range of sources, such as the researchers' "own experiences, general knowledge or reading, and the stories of others" (Glaser & Strauss, 1967, p. 67). In this sense, having a feel for both the researchers' and the participants' experiences can "illuminate the readers understanding of the cultural event, place or practice" and help contextualise the reported findings (Krizek, 2003, p. 149). More importantly, it can equally challenge science to consider the power of *situated knowledge*, such as the significance of participants "memories of emotion, family, and personal experience through sport" when developing 'truth' (Popovic, 2010, p. 237). Rather than viewing these influences as a source of bias, they indicate that the researcher's necessary involvement calls for rendering transparent the collection and presentation of data (Horsburg, 2003).

At the time of this project, the lead author had 15 years' experience as an Australian football referee, which included refereeing in both local and national premier league competitions. This experience involved refereeing at times alongside national and FIFA level referees, as well as enabling first-hand observations of referee training and practice. As part of the research process, the lead author attended the national A-league seminar, as well as the

local member association seminars as an active registered member of the referee organisation. These sessions occur annually and then monthly over the course of a single season. Exposure to these sessions indicated a heavy emphasis on watching video foul play incidents for skill development, which usually involved watching a pre-determined foul and then receiving a FIFA sanctioned correct decision for each clip.

The opportunity to witness and participate directly in training involved in refereeing actual games, highlighted the complicated paradoxical relationship referees had with achieving consistent decision-making. Moreover, these sessions allowed the lead author to identify that most disagreement on specific incidents was often between elite referees. For example, among the best referees decision-making choices for a single incident – could be as diverse as no foul to the highest sanction of issuing a red card. These performers regularly discussed conceptually complex applications of consistency, such as varying foul judgements in the interests of specific games, using different techniques to manage or diffuse situations and varying fouls depending on changing game demands.

Immersion in the field is part of the *theoretical sensitivity* process, where these experiences can be compared with "relevant previous theories to provide the conceptual context for this study" (Holt et al., 2008, p. 665). This iterative research process highlighted how scientific examination and testing of refereeing expertise in the field has been biased towards sources of knowledge gained from empirical evidence. Investigators have rarely sought to access experiential knowledge in the form of insights, experiences and the views of referees themselves (see Pina et al., 2018), tending not to capture the complex nature of their expertise (Russell et al., 2019). In this study, the process of gaining perspective and opinion of what referees believe 'works' to achieve consistent performance – true or otherwise - allows us experientially-based insights into how "choice is made under mitigating

circumstances that direct opinion down paths of decision that are relevant to given goals" (Shaw, 1982, p. 212).

Participants

Ethics approval for the study was granted from a local xxxxx University and participation granted by the refereeing body of Football Federation Australia. Participants provided written consent pertaining to the conditions of involvement, with their anonymity preserved at all times. The final interview group of comprised of 8 referees that were A-League and/or FIFA level referees and had 3.2 yrs mean national-level experience. The experience of the referee cohort ranged from a minimum of 1 year to a maximum of 8 years' experience of arbitration at the level of the national league.

Interview process

Initially, eight individual open-ended interviews were conducted with each participant, with sessions lasting on average 30-45 minutes. These were all conducted by the lead author. Each interview had questions directed towards discussions around our stated aims: (i) what is decision-making consistency, and (ii), how does consistency manifest in decision-making actions. Rather than the same questions being adopted, we instead sought consistency in the values that guided the questioning process, aiming to have participants share and discuss match-situations and contexts where decision-making consistency might be relevant. We asked questions that encouraged participants to consider moments that their decision-making behaviour: (i) affected play (ii) influenced the behaviour of players or (iii) constituted what they believed to be good refereeing. We sought to more broadly contextualise claims made by referees, by asking participants to clarify whether there were any scenarios where their decision-making response would have differed. For example: "It is interesting that you indicated you would change your decisions if it had been earlier in the game – why so?".

Following these initial interviews, we engaged in *member reflections* (Tracy, 2010), to gather a degree of "correspondence between the researcher's findings and the understandings of the participants being studied" (p. 844). This led to a total of twenty interviews including the original eight. During member reflections, we would ask follow-up questions that had emerged from initial data collection, such as: "many colleagues indicated that if the incident was only minor and early in the game, they would refrain from issuing a caution. Is this a view you share?". These types of questions gave participants an opportunity to consider, explore and share deeper values, opinions, and motivations guiding their decision-making process rather than focussing on descriptive accounts of incidents. This reflection process generated diverse and plentiful data, allowing the lead researcher during data analysis to 'map out' and recognise implicit, complementary and/or contradictory recurring themes. In the following section we outline the process we took to compare between key events and their associated reasons/actions, to help guide further data collection and developing theory (Smith & Sparkes, 2016).

Data analysis and theoretical saturation

Upon completion of an interview the lead author conducted and transcribed the interview verbatim to ensure implicit meanings and contextually specific language were not overlooked (Seve et al., 2006). The *initial coding* was line-by-line open coding in conjunction with memo-writing, a process which facilitates identifying tacit assumptions, explicating actions with meanings, comparing data with data, and noticing gaps in the data (Charmaz, 2006). In the case of this work, this led to numerous early descriptions emerging that resembled language used by participants (e.g. "when the temperature rises"), while simultaneously the researcher developed codes that encapsulated the meaning of these actions with respect to achieving 'consistency' (e.g. "reference points"). These emerging codes represent "leads, and hunches that [the researcher] find or identify in the data… then they may gather more data,

ask more questions, and check their developing categories" (Charmaz, 1990, p. 1162). In this respect, further interviews with referees allowed for intricacies and tensions to be contrasted and explored among referees about the study's findings, providing opportunities for "questions, critique, feedback [and] affirmation" (Tracy, 2010, p. 844).

As concepts started to take shape, *focussed codes* were developed that represented stand-alone conceptual elements of theory (Charmaz, 2006). To enhance rigor, as these conceptual ideas were forming and being finalised, they were discussed as often as necessary with members of the research team to "encourage reflection upon, and exploration of, multiple and alternative explanations" (Smith & McGannon, 2017, p. 113). This process ensured a proportioned view of the evidence, allowing greater self-reflexivity for the lead author to consider the nature of how hisresearch background and refereeing experience "inevitably impact[s] upon the meaning and context of the experience under investigation" (Horsburgh, 2003, p. 308). We concluded that *theoretical saturation* had occurred when the ongoing process of constant comparison was no longer extending our higher-level concepts (Charmaz, 2006). To this end, we present our methodology as a case that the analytic insight and theory put forward in the following sections is authentic, evocative, plausible, and can be trusted enough to "act on and make meaningful decisions in line with" substantive practical applications (Tracy, 2010, p. 843). Throughout our Results section, referee names are substituted with numerical labels (e.g., R2 – referee two) to retain anonymity.

Results

This work set out to establish what views Australian national-level referees held about consistent decision-making and how those beliefs shapes their decision-making practice. Two key interrelated concepts were constructed. These were that consistent decision-making should be understood as 'referential and game dependent' and related to 'purpose and

context' (see Figure 1). The concept 'referential and game dependent' was characterised by an interdependent relationship between: (i) decisions by the referee that instituted a standard for defining what a foul is or is not, ongoingly (*"once I set my threshold"*) and (ii) information from the specific match at hand ("*what the game's giving you")* to mediate that standard. 'Purpose and context' suggested that referees perceived moments in a football match not necessarily in terms of whether they met the criteria of a foul but instead: (i) how their decision-making choices worked together during the match to achieve overarching match goals *("does that [decision] achieve anything"*), as well as foreshadowing (ii) how their decision-making response to an incident in the immediate context ("*what is needed right now"*) might influence the construction of context throughout the game. Perhaps most significantly, referees repeatedly suggested that the "laws of the game are made out of paper… they are not rigid… that is one of the beautiful things about our game is that you have that flexibility" (R3). As such, each conceptual category uses actual decision-making examples to exemplify how referees opt for varied and unique decision-making approaches to achieve performance goals.

Referential and game dependent

Referees viewed consistency as 'referential' (*"once I set my threshold"*) and dependent on information from specific match at hand ("*what the game's giving you")*. Participant R2 explains that "once I set my threshold in terms of a decision… and that tackle is a yellow card and there is another tackle that is similar, then it will affect my decision because I need to then become consistent throughout the game based on what I do initially" (R2). Participant R1 explains how this can then evolve during game play:

> Oh yeah, I think you need to be consistent. You know, if you have players standing on the ball for example and you tell the guy to get out of there and then if you go in to the opposite team that is doing the

> same thing ten minutes later and you go in book the bloke then you are not seen to be consistent. (R1)

This example in football refers to situations where, after a foul has occurred, the opposing team deliberately *delays the restart of play* by "standing on the ball". It shows how once players have received information from the referee it becomes a 'reference point' (i.e. 'threshold') for what is and *will be* considered an infringement. The information that comprises a reference point can however vary depending on what time in the game the incident happens or how significant the impact was.

For instance, R7 says "I booked a guy for his second yellow in the 90th minute for kicking the ball away and sometimes you have to give it, even when it seems small… but he had three opportunities to get away from the ball and still hadn't taken it". In this case, even though the action of delaying was "small", the referee considered the second yellow card to be warranted, which results in a dismissal from the match. This was due mainly to recognition that delaying the restart is more significant when there is less time remaining, but also the nature of the attempt to delay (e.g. "three attempts"). While this was the outcome in this scenario, when circumstances surrounding the incident differ, it may afford the referee the opportunity to treat the 'same' incident 'differently'. For example, R2 explains: "let's say in the 5th minute you give a free kick in the middle of the park… and someone runs across in front of the ball to try and delay it. In the 5th minute you might call him out and say 'that is it', but that is a public warning for everyone and if it happens again there are no excuses… because I had already set my reference point".

Both examples highlight how due to differing contextual considerations (e.g. such as the time of an incident and incidents prior), what is considered in principle an identical incident ("delaying the restart of play"), can result not only in varying decision-making responses, but also contrasting thresholds being instituted *ongoingly* that may influence the

trajectory of both their own decisions as well as the players actions, throughout the match. In this sense what comprises a 'reference point' is not just defined by the actions of the players or the prevailing situational conditions, but also the information provided by the referee's decisions. Rather than seeing this as inconsistency, participant R7 explains that recognising how your decisions have helped define particular foul situations is an important skill that can influence the game's development:

> He dug a hole for himself too early by creating expectations with the first yellow card. He probably gave it to 'set the standard', but that meant that once he had drawn that line in the sand, all the other decision after that meant he had to keep going down that path… And by drawing that line in the sand too early affected all the subsequent decisions to come… 14 cautions shows that your decision early on, can continue to affect the ones you make later on.

In order to avoid the spiralling circumstances described, referees clarify that they need to carefully balance 'setting the standard' against "*what the game's giving you*" (i.e. the nature of the standard the players may wish to play at or within). Participant R7 explains that "the line does change from week to week… adaptability is important. If you just go out and draw the same line every week that may not necessarily suit the game". Participant R8 reinforces this belief stating that "you need to treat the game as a completely different game; each game is completely different" because there are some games "where there is not much at stake" (e.g. "round 24 and that have no chance of making the finals", R1), some games where there is a "fierce rivalry" (R6), and other games where the players might just play "really nice fluid football" (R8). Participant R4 elaborates:

> The standard or tone that you might set is different from game to game, and that may be dependent on the teams rankings or just the nature of the game itself, you might get two teams that might want to be pussycats, so it is easy to set the standard or they are setting the standard for you – as it may be. And other game you have got to set a really firm standard. (R4)

Participant R6 explains that information from the game helps inform the threshold you eventually set because if the players act in a way that indicates "potential for a bit of argy-bargy… you are better off pulling up a few 'little things', especially early on, to avoid anything simmering and getting out of control later on". Participant R7 clarifies that identifying this context can be really important, yet sometimes taking the opposite approach is more effective in developing future context: "if you know a team wants to push you, don't book the first tackle. If you are not sure how they play though, you might go charging into book [award a sanction] the first player and that might not work for you over the course of the game". Participant R7 is reinforcing here that it is important for referees to consider information the players are providing about how they want to play the game – so that when do contribute decision-making information to match incidents, the eventual 'reference point' set serves to align and *facilitate* play. For example, the type of play that the players and referee is looking to create for that given match:

> In the first 10 or 15 minutes, you're roughly trying to attack [the game] in a similar way, in that you are trying to gain control of the game… you would also take information from the type of game it is, because after 5 minutes, if it turns out that players are playing really nice fluid football, well then you can maybe relax a bit… (R7)

In summary, referees suggested that consistency is a developing conversation with the players, whereby the referee aims to institute a 'standard' that ensures the game is played within acceptable boundaries but also accounts for the unique nature of the contest. To this extent, referees argued that "getting a level of consistency between referees is actually very difficult because there are so many variables that take place" and it is instead more relevant to "make sure that you get the same level of fouls throughout the match" (R8). As such, referees generally portrayed consistency as remaining faithful to the references points, they had outlined to the players and the conditions that defined them.

Purpose and context

Referees explained that consistency is more than considering fouls in isolation but instead evaluating what purpose (*"does that [decision] achieve anything"*) and context (*"what is needed right now"*) a decision offers to the games functioning. In this respect, referees commonly stated that "each incident needs to be treated in insolation as an incident itself but within the context of the whole match" (R6). Participant R1 unpacks this:

> If you have a good rapport with the players and things are going well, you are not going to sanction those piddly small things which may technically should have been cautioned or whatever, you are going to use a bit of man-management and try to deal with the situation in other ways because if you go then go and sanction a player for small type things … often you can lose your respect to the players. Given that the match didn't require that attention to be given. Although you could be sticking to the laws and doing the 'correct' thing, it can then work against you… because then you have to deal with the repercussion of that decision for the next 40-50 mins. (R1)

Participant R1's comments evidence how fouls serve to build relationships and set expectations for the rest of the game. For example, participant R2 explains that "every little thing [the ref] does builds links for the whole game and every bit of communication has a purpose to it". Participant R6 agrees, "absolutely, they [decisions] are linked" and that owing to this, decisions of the referee are not simply reactions to moments but rather become interconnected and intertwined with what events eventually transpire in the game. Participant R6 explains: "you try and referee proactively. If you have to give someone a red card, you have to do it, but you try to make sure that that doesn't take place".

This capacity to referee the game in an interconnected proactive way, was described as having a "smell" for the game; defined by an individual's capacity to sense the context of the match (i.e. state of play) as well as foreshadowing how their treatment of an incident might positively (or negatively) influence the creation of context throughout the game (e.g.

avoid said red cards occurring 'at all'). Participant R7 explains that recognising how and when to act "comes back to the smell" and that this might mean considering "why or when someone does something and *what is needed right now*… it could be a control thing because you know that if you give a yellow right now the whole game will explode". Whereas "if the players are out there to play 'football', they are going to have a bit more leniency, and you might be able to say to a player in the 60th minute 'hey come on, this isn't the way the game is being played' and they might respond to it" (R4). Participant R2 elaborates how this can subsequently manifest in quite different and varied decision-making outcomes:

> It would depend on the feeling of the game at the time, in terms of the decision, even whether it is a free kick or if you are going to play advantage, or if it is a yellow or a send-off. Obviously there are decisions which we have no choice, reds red, yellows yellow, but perhaps on the marginal decisions… the games, if you feel like the game needs a caution for your control, if you feel like the game can flow because nothing is happening, the players are happy, umm then you might play advantage as much as you can. In a game where you might give someone a public warning because nothing is happening, everyone's happy, but in another game, where things aren't going so well, you might have a yellow card out straight away. (R2)

As can be seen, referees view what decision-making options they pursue during the game as *fluctuating* in line with the purpose of the decision and the broader context of the game (e.g. "keeping the lid on it without boiling over the top and spoiling the game", R3). In this way, referees see the game as an interconnected system of events and use a balloon analogy to describe their contribution to how much pressure is in the balloon (i.e. the games current state of play). R5 explains," so as the games expanding the referee lets a little bit of pressure out, then it comes back up and you let a little bit of the pressure out and pulling things up and dealing with them as you go rather than letting them all escalate until it explodes and you have got chaos everywhere". Participant R8 explains that this can make life difficult for referees with how 'consistent' their decision-making 'appears', because "it

depends on what is happening in the match, and they [managers/coaches] aren't fully aware of why you're making certain decisions at certain points… it might be a control thing". Participant R8 further contends that different decision-making strategies are sometimes required at different points in the game to attend to competing priorities *("what is needed right now")* regardless of previous thresholds or context created by decisions:

> You've got to give a foul, give anything, because the temperature has risen, the players are out of control, and you know that the referee has lost control. For example, if you give a contentious free kick, of if a player fouls someone and the 'temp' has risen, the players are unhappy then you think, right for the next five minutes I've got to be on top of everything. I've got to hammer all fouls, no advantages; again you have got to come back to control. (R8)

Participant R7 reiterates that even though "most players, coaches and commentators ask for consistency, as in 'what is good for one should be good for the other'" he doesn't believe that that is what they truly want: "They just say it… looking at consistency is multifaceted… I am not sure they want robots running around, interpretations will always be different". In this sense, it is more important for referees to use decisions to 'consistently' maintain interconnected priorities and goals, such as the level of 'control' (e.g. "you have to spoil the game in a way for those few minutes because you are trying to regain control"), to regain 'authority' (e.g. "when the referee has lost control, you have got to close it, you've got to stamp your authority on it") and to moderate the 'temperature' (e.g. "just take a foul there because they are starting to kick each other", R5). These priorities also shift depending on other factors, such as where it is on the field (e.g. "if it is a minor foul in the defending third where there is not much scope for advantage, you might want to pull that up straight away", R4) yet always remain grounded in the greater interests of the game (*"does that achieve anything")*. Participant R2 explains:

> We have got a responsibility to the game that is going to depend on a whole heap of circumstances… I mean if you caution someone in the

> 2nd minute of the game for a tackle and then in the 70th, they just poke the ball away after you blow the whistle, technically we can show them a second yellow. Does that achieve anything? What were they trying to do? What was the context of the game? What was the score? Are you going to achieve more by calling him out and saying "that is it". I think there is scope for difference. (R2)

The idea of what a foul 'achieves', and thus what the players 'need' from the referee at a given point, is connected to deep beliefs about how the game should be expressed (e.g. that the referee ensures a just and desirable outcome). For example, "if is the 80th minute… if you are giving a free kick in the favour of a team that wants to score, then you want to go very, very quickly…you have to play along with what the particular team that's got the decisions wants" (R8). Referees however clarified that while indeed "there is definitely room for movement" on how decisions or moments are handled, the counter is also true, where you simply do not consider the broader context or purpose but just the action itself: "you just rule them out straight away. If you come in with a rugby tackle, take someone over the fence, see you later" (R4).

In summary, evidence based and informed consistent decision-making as considered by Australian national-level referees, plays a much more expansive role than just the similar identification of fouls. Rather, referees viewed decisions as opportunities to communicate the way the game *can* be played in the interest of managing the whole event. Accordingly, some referees described equal treatment of incidents as erroneous and unachievable - "it is a bit of a 'furphy' to be honest" (R7) - arguing that discretionary and varied decision-making responses were essential to deal with game demands. Moreover, they explained that contextual sensitivity (e.g. such as attention to game time, game score) were necessary to provide a purposeful decision that would support what the players were attempting to achieve at a given moment (e.g. quick goal scoring opportunity with little time remaining). Referees would sometimes completely disavow previous referential markers of what had constituted a

foul (and in turn, contravene contextually developed understandings of 'consistency'), if other decision-making priorities were more critical (e.g. ensuring the game does not "explode", R7).

Discussion

Using a grounded theory approach, we explored perspectives held by national football referees on: (i) what decision-making consistency is and (ii) how consistency manifests in decision-making actions. The analysis revealed that participants view 'consistency' as a context-dependent outcome, rather than a process of uniformly responding to isolated foul-play transgressions with putatively correct responses. How consistency of decision making was achieved could be conceptually explained by a decision-making process that was 'referential and game dependent' and related to 'purpose and context'.

Our discussion represents a core component of a grounded theory approach, which is that the "integrity of a GT [grounded theory] study is maintained by conducting the detailed and substantive review of the literature as part of the iterative process" (Weed, 2017, p. 152). In this respect, we contend that decision-making consistency in referees *can* be understood as dynamical transactions within a competitive football game – an emergent process between the decision-making actions of the referee, and exploitation of specifying contextual factors in a complex system. We propose that varying decision-making responses to similar incidents are a marker of expertise which work to maintain consistent value-orientated performance outcomes. Finally, we encourage alternative explanations of our results and findings to broaden *theoretical significance*, by "extending, building, and critiquing disciplinary knowledge" (Tracy, 2010, p. 846).

Variability as a marker of expertise

Traditional descriptions of the referee's decision-making role have tended to characterise them as operating 'above and outside' of the game, that is, they primarily "perceive sporting actions and react to whether an infringement has occurred" (Kittel et al., 2019, p. 261). To this end, there has been a heavy emphasis on assessing refereeing expertise by judging incidents in isolation - irrespective of previous decisions – to pair correct interpretations with appropriate decision-making responses (Schweizer et al., 2013). While these models do acknowledge the influence of contextual factors, they tend to be positioned as either negatively augmenting existing foul information (e.g. sequential or calibration bias) or as a substitute to overcome "errors [that] are due to missing information and uncertainty pertaining to the judgement" (MacMahon & Mildenhall, 2012, p. 157). Our findings suggest instead that incidents do not contain all the information needed for interpretation and that a referee's decision-making actions *are* the 'missing' specifying information. In this way, each decision contributes to ongoing decision-making opportunities and outcomes (Renshaw & Gorman, 2015) by contributing contextual information which defines the nature of each incident. For example, referees in this study emphasised how varying and adapting 'the line' (what they determined was a foul) from game to game reflected a critical component of skilled behaviour, rather than reflecting inconsistency. Moreover, they suggested that an inability to institute a 'standard' that matched the nature of the contest, often had immense implications for the emergent trajectory that each competitive game eventually took (e.g. "14 cautions shows that your decision early on, can continue to affect the ones you make later on, R7). To this extent, our work supports the contention of Unkelbach and Memmert (2008) that referees need to provide a judgement scale anew each time they referee a match by "develop[ing] a feeling for the game" (p. 97). Our work also gives dimension to this 'feel', suggesting that new interpretations are not only possible but a *necessary*, emergent and ever-changing outcome of 'dynamical transactions' between their own decision-making actions and exploitation of

specifying contextual information within the complex system of the match. In this respect, expertise is not reflected in the acquisition of similar responses to incidents but rather by "very unique, individually-adapted, efficient solutions" which work to define and influence the nature of the systems (the football games') competitive functioning (Komar et al., 2019, p. 135).

The referee as part of a complex system

How referees contribute to each game's competitive functioning can be understood from a complex systems perspective, whereby the referee's decisions act as information that helps regulate the movements of individuals (Button et al., 2020; Renshaw & Chow, 2019). This is a symbiotic process. Information perceived by the players, such as referee's decisions, constrains their actions. Yet equally, actions from players related to those refereeing decisions, generates new information impacting on the decision-making behaviours of the referee. This reciprocating cycle of perception and action is continuous in all self-regulating human behaviour. These collective actions from both teams and the referee, are constantly evolving and emerging states of (re)organisation that a game passes through, and therein characterise the games trajectory. The results of this study indicate that expert officials not only recognise these 'states' but seek strategies to use their decision-making actions to create bifurcation points. Bifurcations points can be considered as decisions by the referee that serve to redirect the games trajectory, to influence each game's functionality (e.g. "just take a foul there because they are starting to kick each", R8). At times, when the game as a complex system appeared unstable (e.g. "because the temperature has risen, the players are out of control", R8) the referee would 'find a foul' and minimise certain decision-making strategies like advantage. Yet, when the system is stable and functioning (e.g. "if you feel like the game can flow because nothing is happening, the players are happy", R2), referees look to 'avoid a

foul' and use alternative management techniques to maintain order (e.g. "you are going to use a bit of man-management", R1).

These findings provide perspective on the sentiment that "every umpiring decision may be critical and have a direct impact on the result of the game" (Larkin et al., 2011, p. 427). While it is indeed true that single decisions can have a significant outcome on a game, our work suggests *every* decision is aiming to influence the trajectory of the match towards goal-orientated outcomes in line with the referee's intentions (e.g. such as maintaining control or reducing red cards). In this respect, what constitutes a foul *at all* often depends on an intersection between often competing, emergent priorities, such as where it is on the field (e.g. "minor foul in the defending third where there is not much scope for advantage", R4), whether the referee needs to regain control (e.g. "you feel like the game needs a caution for your control", R2) or whether a foul will be helpful at that given movement (e.g. "if you are giving a free kick in the favour of a team that wants to score, then you want to go very very quickly"). This study therefore extents on the 'four pillars' grounded theory (Russell et al., 2019) by conceptually defining how specific game factors, purpose, context and prior decision-making interventions, lead to *varying* decision-making responses to consistently 'maintain control' and 'preserve the game's integrity'.

Theoretical and methodological implications

Previous studies have reported that "the message to get out to all referees is that we should treat every game the same" (ex-elite referee cited in Webb et al., 2018, p. 1033). Our work provides clarification to this sentiment, indicating it should be interpreted more as a philosophical vision to treat all games with integrity, rather than to literally treat actions as identical and their contribution to the competitive environment equal. Our findings also indicate that although there are naturally some pre-determined expectations of what

constitutes a foul during a football match, what ultimately is classed as a foul by the referee will "depend on the feeling of the game at the time, in terms of the decision, even whether it is a free kick or if you are going to play advantage, or if it is a yellow or a send-off" (R2). In turn, researchers could consider the implications these findings have for statistical measures of accuracy and/or consistency when contextual factors and/or the decisions of referee's are excluded from research methods and design.

A recent integrative review on football referees suggested a "need to extend the scope of empirical research in refereeing", noting that less than twenty percent of research on referees had adopted qualitative methods (Pina et al., 2018, p. 10). By adopting a qualitative approach, the current study has highlighted how methods that seek to preference objectivity, generality and prediction have diminished the complexity of referee expertise (Russell et al., 2019). Moreover, "conducting grounded theory inquiry means learning about the empirical world" (Charmaz, 2017, p. 4). The approach taken in this study has raised questions about the over-reliance of research in general on quantitative tools to describe human experiences. While relevant at times, this tendency to over-rely on quantitative approaches, potentially rests in a belief that mathematical language and physics, provides a 'factual bridge' to represent 'objective accounts' of observed actions and behaviours (Shaw, 1982). The notion that quantitative approaches are more objective, impartial or value-neutral, has been questioned extensively (see Guba & Lincoln, 1995; Horsburg, 2003).

Without having all the answers, grounded theory approaches offer behavioural frameworks such as ecological dynamics, onto-epistemological contemplation and reflection around how knowledge of performance and research action is "no more than correct doxa in the sense of motivating pragmatically true (useful) action that leads to success as relatively defined in a given semantic context" (Shaw, 1982, p. 213). Reciprocally for grounded theory, key questions surround how to "move from a substantive theory grounded in the particular

area researched, to more generically applicable formal theory" (Weed, 2017, p. 152). This process can be achieved by comparing grounded participant data with existing behavioural frameworks. However, questions arise as to whether ongoing study remains under the banner of grounded theory or shifts to an approach resembling the framework that the grounded data is beginning to support – or both.

Practical implications

Qualitative research has significance when it encourages curiosity in the reader, can invite questioning in other settings and/or causes a potential shift in the way a craft is practised (Tracy, 2010). Our work has interesting practical implications for video referee training, suggesting a move away from focussing on equating specific movements with fouls, and instead encouraging exposure to situations and circumstances that invite referees to strategically use decisions to foster competitive circumstances (affordances) favourable to them (Passos et al., 2016). For example, when decision-making scenarios are presented to referees, rather than emphasising predominately 'what the foul is' they could consider 'what the foul can offer'? To elaborate, the findings of this study suggest that decision-making to be conceived as aligning with Bernstein's (1967, p. 134) notion of 'repetition without repetition', where multiple functional decision-making options are explored with respect to their capacity to achieve similar performance outcomes and meet varying task goals. To facilitate this conceptualisation, vignettes could accompany video-led practice sessions, where a referee has to ascertain how various decision-making strategies may influence the game trajectory in certain performance scenarios. For example, referees could be encouraged to identify changes in system stability (e.g. the game is deteriorating) and then to outline what decision-making responses could potentially serve goals connected to those situations (maintaining control more important than facilitating entertaining play).

These approaches, however, are somewhat limited, as the passive, reductionist and compartmentalised nature of video training rarely replicates the dynamic, emergent nature of the flow and interconnectivity of the competitive performance environment. In this sense, video training does not provide referees with the opportunity to continuously adapt to changing task constraints, whilst engaging with a variety of contextualised, emotional and situationally-specific competitive pressures (Headrick et al. 2015). Ideally, modified game-based opportunities for referees to practice making decisions in 'context' and at game speed, would allow referees to attune to key performance information to support decision-making goals and potentially, encourage faster development of expertise. Moreover, intentionally designing the performance environment to contain relevant informational constraints (e.g., the score as nil-nil, or suggesting a player is on a second yellow card), would encourage referees to develop effective and decision-making solutions to adapt to context-specific problems (Renshaw & Chow, 2019).

Future work and limitations

Our work only interrogated the notion of outcome 'consistency' within a small sample size of Australian referees at national-level competition, thus limiting the generalisability of our findings to other sports beyond football. To what extent decision-making actions are intentionally varied across competitions, countries and cultures would also be valuable for training and development, particularly since referees are a geographically dispersed group (e.g. Webb, 2016) that work in teams during the most significant competition events (e.g. World Cup). Future work could explore how the role of contextual factors varies between different sports, particularly those that have similar structural characteristics of play, such as invasion games like ice hockey, Australian Rules football or American football. Moreover, exploration of the role of context in fielding sports like baseball and cricket, would provide

valuable insight as to whether sources of information used to support action vary in line with domain-specific tasks.

Conclusions

We conclude that decision-making consistency in referees can be understood as dynamical transactions emerging from their own decision-making actions and exploitation of specifying contextual information in a complex system. Results suggest that attunement to key contextual information is necessary to prospectively control player behaviour and, thus, the trajectory of the game. We propose that *varying* decision-making responses to similar incidents are a marker of expertise, representing intentional attempts to maintain consistent value-orientated performance outcomes, rather than inconsistency. This conceptualisation of the referee's role would be useful in training programs at an advanced level, where the consistency of decision-making choices needs to be evaluated with respect to how effectively they constrain and shape player behaviours during competitive performance.

References

Anderson, J. R. (1983). The architecture of complexity. Cambridge, MA: Harvard University Press.

Anderson, K., & Pierce, D. A. (2009). Officiating bias: The effect of foul differential on foul calls in NCAA basketball. *Journal of Sports Sciences, 27(7)*, 687-694. doi:10.1080/02640410902729733

Araújo, D., & Davids, K. (2016). "Team synergies in sport: Theory and measures." *Frontiers in Psychology*, 7:1449. doi:10.3389/fpsyg.2016.01449

Araújo, D., Davids, K. & McGivern, P. (2019). The irreducible embeddedness of action choice in sport. In Handbook of Embodied Cognition and Sport Psychology (Edited by Massimiliano L. Cappuccio). MIT Press: Cambridge, MA.

Araújo, D., Davids, K. & Renshaw, I. (2020). Cognition, emotion and action in sport: An ecological dynamics perspective. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of Sport Psychology (4th edition., Vol. 1, pp. 535-555).* Wiley & Sons, Inc.

Araújo, D., Davids, K., & Serpa, S. (2005). An ecological approach to expertise effects in decision-making in a simulated sailing regatta. *Psychology of Sport and Exercise, 6(6),* 671-692. doi:10.1016/j.psychsport.2004.12.003

Araújo, D., Hristovski, R., Seifert, L., Carvalho, J., & Davids, K. (2017). Ecological cognition: expert decision-making behaviour in sport. *International Review of Sport and Exercise Psychology*. DOI: 10.1080/1750984X.2017.1349826

Bernstein, N. A. (1967). The Control and Regulation of Movements. London: Pergamon Press.

Catteeuw, P., Helsen, W., Gilis, B., & Wagemans, J. (2009). Decision-making skills, role specificity, and deliberate practice in association football refereeing, *Journal of Sports Sciences, 27(11),* 1125-1136. doi:10.1080/02640410903079179

Charmaz, K. (1990). 'Discovering' chronic illness: Using grounded theory, *Social Science & Medicine, 30(11),* 1161-1172. https://doi.org/10.1016/0277-9536(90)90256-R.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* London: Sage.

Charmaz, K. (2008). *Handbook of emergent methods.* New York, US: Guilford Press.

Charmaz, K. (2017). Special invited paper: Continuities, contradictions, and critical inquiry in Grounded Theory. *International Journal of Qualitative Methods, 16(1).* https://doi.org/10.1177/1609406917719350

Connor J. D., Farrow, D. & Renshaw, I. (2018). Emergence of skilled behaviors in professional, amateur and junior cricket batsmen during a representative training scenario. *Frontiers in Psychology,* 9*:*2012. doi:10.3389/fpsyg.2018.02012

Cunningham, I., Simmons, P. Mascarenhas, D. (2018). Sport officials' strategies for managing interactions with players: Face-work on the front-stage. *Psychology of Sport & Exercise*, 39, 154-162. https://doi.org/10.1016/j.psychsport.2018.08.009

Davids, K., Button, C., & Bennett, S. J. (2008). *Dynamics of skill acquisition: a constraints-led approach*, US: Human Kinetics Publishers.

Demuth, C. (2015). "Slow food" Post-Qualitative research in psychology: Old craft skills in new disguise. *Integrative Psychological and Behavioral Science*, *49, 2*07-215. doi.10.1007/s12124-015-9304-8

Dosseville, F., & Laborde, S., & Raab, M. (2011). Contextual and Personal Motor Experience Effects in Judo Referees' Decisions. *Sport Psychologist. 25*. 67-81. Doi:10.1123/tsp.25.1.67.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. New York: Aldine.

Glăveanu, V. P. (2011). Is the lightbulb still on? Social representations of creativity in a Western context. *The International Journal of Creativity & Problem Solving, 21(1),* 53-72.

Guba, E. G., & Lincoln, Y. S. (1995). Fourth Generation Evaluation. Sage, Newbury Park, CA, USA.

Headrick, J., Renshaw, I., Davids, D., Pinder, R. & Araújo, D. (2015). "The Dynamics of expertise acquisition in sport: The role of affective learning design." *Psychology of Sport and Exercise, 16,* 83–90. doi:10.1016/j.psychsport.2014.08.006.

Holt., N. L. (2016). Doing grounded theory in sport and exercise from *Routledge Handbook of Qualitative Research in Sport and Exercise*. Routledge. doi/10.4324/9781315762012.ch3

Holt, N., & Tamminen, K. (2010). Moving forward with grounded theory in sport and exercise psychology. *Psychology of Sport and Exercise, 11,* 419-422. 10.1016/j.psychsport.2010.07.009

Horsburgh, D. (2003). "Evaluation of Qualitative Research." *Journal of Clinical Nursing, 12 (2),* 307–312. doi:10.1046/j.1365-2702.2003.00683.x.

Kittle, A., Larkin, P., Elsworthy, N., & Spittle, M. (2019), Video-based testing in sporting officials: A Systematic review. *Psychology of Sport and Exercise, 43,* 261-270. https://doi-org.ezp01.library.qut.edu.au/10.1016/j.psychsport.2019.03.013

Komar, J., Potdevin F., Chollet, D. & Seifert, L. (2019). Between exploitation and exploration of motor behaviours: unpacking the constraints-led approach to foster nonlinear learning in physical education, *Physical Education and Sport Pedagogy, 24:2*,133-145, DOI: 10.1080/17408989.2018.1557133

Krizek, R. L. (2003). Ethnography as the excavation of personal narrative. In R. P. Clair (Ed.), Expressions of ethnography: Novel approaches to qualitative methods (pp. 141-151). New York: State University of New York.

Larkin, P., Berry, J., Dawson, B. & Lay, B. (2011). Perceptual and decision-making skills of Australian football umpires, *International Journal of Performance Analysis in Sport, 11(3),* 427-437. DOI: 10.1080/24748668.2011.11868562

Lweiwick, P., Hill, T., & Czyewska, M. (1992). Non-conscious acquisition of information. *American Psychologist,* 47(6), 796-801.

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. https://doi.org/10.1260/1747-9541.7.1.153

MacMahon C., & Starkes, J. (2008). Contextual influences on baseball ball-strike decisions in umpires, players, and controls, *Journal of Sports Sciences,* 26:7, 751-760. DOI: 10.1080/02640410701813050

MacMahon, C., Starkes, J., & Deakin, J. (2007). Referee Decision Making in a Video-Based Infraction Detection Task: Application and Training Considerations. *International Journal of Sports Science & Coaching, 2(3),* 257–265. doi.org/10.1260/174795407782233164

Mascarenhas, 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,* 1-9. doi:10.2174/1875399X00902010001

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

Mascarenhas, D., O'Hare, D., & Plessner, H. (2006). The psychological and performance demands of Association Football refereeing. International Journal of Sport Psychology, 37, 99–120.

Masters, R. S. W., Poolton, J. M., Maxwell, J. P., & Raab, M. (2008). Implicit motor learning and complex decision making in time constrained environments. *Journal of Motor Behavior*, 40, 71-79. doi: 10.3200/JMBR.40.1.71-80

McCosker, C., Renshaw, I., Russell, S., Polman, R. & Davids, K. (2019). The role of elite coaches' expertise in identifying key constraints on long jump performance: how practice task designs can enhance athlete self-regulation in competition, *Qualitative Research in Sport, Exercise and Health,* DOI: 10.1080/2159676X.2019.1687582

Passos, P., Araújo, D. & Davids, K. (2016). Competitiveness and the process of co-adaptation in team sport performance. *Frontiers in Psychology,* 7:1562. doi:10.3389/fpsyg.2016.01562

Passos, P., Araújo, D., Davids, K. & Shuttleworth, R. (2008). Manipulating constraints to train decision making in Rugby Union. *International Journal of Sports Sciences and Coaching, 3(1),* 125-140. doi:10.1260/174795408784089432

Pina, J. A. E., Passos, A., Araújo, D., & Maynard, M. T. (2018). Football refereeing: An integrative review. *Psychology of Sport and Exercise*, 35, 10-26. doi:10.1016/j.psychsport.2017.10.006

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

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognitive perspective. *Psychology of Sport and Exercise,* 7, 555-575.

Popovic, M. (2010). A voIce in the rink: Playing with our histories and evoking autoethnography. Journal of Sport History, 37(2), 235-255.

Renshaw, I., Newcombe, D., Roberts, W., & Davids, K. (2019). *The constraints-led approach: Principles for sports coaching and practice design.* London: Routledge.

Renshaw, I., & Chow, J., (2019) A constraint-led approach to sport and physical education pedagogy, *Physical Education and Sport Pedagogy*, 24(2), 103-116. DOI: 10.1080/17408989.2018.1552676

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

Russell, S., Renshaw, I., Davids, K. (2019). How interacting constraints shape emergent decision-making of national-level football referees, *Qualitative Research in Sport, Exercise and Health*, 11:4, 573-588. DOI: 10.1080/2159676X.2018.1493525

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

Schweizer, G., Plessner, H., & Brand, R. (2013). "Establishing Standards for Basketball Elite Referees' Decisions." *Journal of Applied Sport Psychology,* 25(3): 1533–1571. doi:10.1080/10413200.2012.741090.

Seve, C., Poizat, G., Saury, J., & Durand, M. (2006). A grounded theory of elite male table tennis players' activity during matches. *The Sport Psychologist, 20*, 58-73. http://dx.doi.org.ezp01.library.qut.edu.au/10.1123/tsp.20.1.58

Shaw, R. E., Turvey, M. T. & Mace, W. M. (1982). Ecological psychology. The consequence of a commitment to realism. In W. Weimer & D. Palermo (Eds.). Cognition and the symbolic processes. Vol. 2 Pages 159 – 226. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Smith, B., & McGannon, K. R. (2017). Developing rigor in qualitative research: problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*. doi: 10.1080/1750984X.2017.1317357

Smith, B., & Sparkes, A. C. (2016). Interviews: qualitative interviewing in the sport and exercise sciences. In B. Smith & A. C. Sparkes (Eds.), Routledge Handbook of Qualitative Research in Sport and Exercise (pp. 103–123). Abingdon, UK: Routledge.

Spitz, J., Moors, P., Wagemans, & Helsen, W. (2018). The impact of video speed on decision-making process of sports officials. *Cognitive Research: Principles and Implications, 3*(16). doi: 10.1186/s41235-018-0105-8

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

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury, CA: Sage.

Tracy, K. (1995). Action-implicative discourse analysis. *Journal of Language and Social Psychology, 14*, 195-215. doi.org/10.1177/0261927X95141011

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837–851. doi.org/10.1177/1077800410383121

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

Unkelbach, C. & Memmert, D. (2010). Crowd Noise as a Cue in Referee Decisions Contributes to the Home Advantage. *Journal of sport & exercise psychology.* doi: *32*. 483-98. 10.1123/jsep.32.4.483.

Webb, T. (2016). Referees and the media: a difficult relationship but an unavoidable necessity. *Soccer & Society*. DOI:10.1080/14660970.2015.1133414.

Webb, T., Dicks, M., Thelwell, R., & Nevill, A. (2018). The impact of referee training: reflections on the reduction of home advantage in association football. *Soccer & Society, 19:7*, 1024-1037. DOI: 10.1080/14660970.2016.1267626

Weed, M. (2009). Research quality considerations for grounded theory research in sport & exercise psychology. *Psychology of Sport and Exercise, 10,* 502–510.

Weed, M. (2017) Capturing the essence of grounded theory: the importance of understanding commonalities and variants, *Qualitative Research in Sport, Exercise and* Health, 9(1), 149-156, DOI: 10.1080/2159676X.2016.1251701

Figure 1: Referee decision-making consistency as an emergent process in a complex system