12
SCORING/JUDGING APPLICATIONS

Anthony (Tony) N. Kirkbride
CSIR SPORTS TECHNOLOGY, SOUTH AFRICA

Summary
The evolution of sports scoring and judging systems is driven by a need to accurately assess sports excellence. Sports performance analysis is constantly striving to migrate from a subjective assessment of performance to a more objective appraisal. Generally, expert judges and referees are engaged in assessing events, actions and movements that occur in their field of observation. These observations are, by and large, subjective in nature. Consequently, modern scoring and judging systems have developed in an attempt to address this limitation and provide a consistent framework within which reliable assessments of performance can be undertaken. As levels of competition become ever closer, the margins separating performances are decreasing, often necessitating the use of technology to adjudicate some occurrences. This area is often catalysed by the development of innovative and engaging broadcast technologies, which subsequently find redeployment as an integral part of the adjudication process.

Generally, scoring and judging systems are designed specifically to minimise bias, be fair and equitable, reduce errors in observation and improve decision making and communication. Nevertheless, some challenges remain and a review of the current developments in this critical area of competition was undertaken by examining relevant peer-reviewed journals which included a technology scan. The chapter concludes with an assessment of future trends in scoring/judging systems and provides a summary of the salient points of this section.

Introduction
The term sportsmanship can be defined in terms of the aspiration of fair play; competition is relished more when there is an equitable chance to gain victory. Ill-conceived and ill-designed tournaments, in which the top seeds eliminate each other in the early rounds, are generally disparaged. However, in the English FA Cup, the lack of seeding in this particular knockout tournament has injected a degree of romance, with potentially exciting draws that would be impossible in a seeded competition. The objective is for the best athlete or team to win and this should not be unduly influenced by the tournament design or scoring/judging system (Appleton, 1995; Clarke et al., 2009; Pollard and Pollard, 2010). The well-known idiom it’s not cricket is often used to imply the unacceptable. The saying originates from the fundamental
sportsmanship principle of fair play or gentlemanly conduct historically applied to a cricket player’s obligation. It derives from a time when instant television replays were non-existent and, consequently, there was an unwritten onus on the batsman to *walk* (i.e. the act of giving himself or herself out prior to a decision by an umpire). The general trend, at least at the elite international level, is for the batsman to stand his/her ground and rely on the technology available to the umpire to make the correct judgement. Errors can be made in the umpire’s analysis and decision-making process and the batsman is hoping for a decision in his favour. Batsmen can stand their ground aiming to deceive the umpire, even when blatantly aware of their obligation. In close decisions, typically in cases of uncertainty and where there is an absence of evidence, the batsman is always given the benefit of doubt. In many sports, players can deliberately attempt to deceive, the more theatrical footballers by diving, particularly in the penalty area (known as embellishing in ice hockey). Soccer players feign injury to draw free kicks/penalties and rugby players have faked blood injuries to gain substitutions. While the debate continues at pace regarding sportsmanship, sport in general strives for fairness.

Soccer, rugby and cricket are three of the most popular spectator and participation sports worldwide and considered the least subjective when deciding a winner when compared to, for example, boxing, ice skating or martial arts. The aim of these ball sports is to score more goals, points or runs than the opposition. The primary objective is so simple and yet why would judging play such an important role? These sports are played in accordance with, and reference to, the notion of the ‘spirit’ of the game and ‘letter’ of the laws. In cricketing parlance, the ‘spirit’ of the game refers to the way the game should be played and incorporates ideas such as fair play or gentlemanly conduct. In the early days of football (circa 1800s), there was self-regulation of rule infringements of players by the offending captain and, as competition increased, so did professionalism, and this led to external regulation and the appointment of neutral umpires and referees (Colwell, 2000). In the modern game, one referee and two assistant referees are the most common in soccer and rugby; cricket generally has two umpires and can have additional technical support from so-called Third Umpires and TV Umpires, depending on the nature of the tournament. Referees permit the game to flow by interpreting the laws in the ‘spirit’ of the game. If they were to apply the ‘letter’ of the law, then every minor infringement could potentially stop the game. Consequently, referees are constantly making subjective judgements and this level of discretion can have a profound influence on the outcome of games. Controversy usually surrounds referees perceived not to be enforcing laws or, conversely, applying the letter of the law too rigidly. Many laws do not permit room for misinterpretation and are relatively straightforward (e.g. goals, throw-ins). However, where a decision or judgement is made on an opinion, this is much more problematical and controversial. As TV coverage of these sports increases, so does the ability to review, interrogate and scrutinise (in increasingly fine detail) umpire and referee decisions. This is where adjudication mistakes are exposed and where proposals are made and debated to reduce inconsistencies. Many sports codes openly welcome aids to assist scoring and judging, while others are more reticent. In an attempt to minimise human intervention in scoring and judging of sports, automated image processing and sensor processing technologies are also finding some applications and these areas are arousing researchers’ interests. Nevertheless, the majority of sports currently rely on human cognition, and concepts that influence our human perception are outlined in the following section.

**Challenges**

Almost a third of all sports registered with the International Olympic Committee (IOC) rely on a system of human judging to partially or fully assess performance (Wolfram, 2010).
Unfortunately, many researchers and commentators have raised doubts about judging bias, both natural and nationalistic (Balmer et al., 2005; Emerson et al., 2009; Hawson et al., 2010; Zitzewitz, 2006, 2010). Emerson et al. (2009) claimed to have found one occasion where the medal standings could have been changed as a result of unbiased judging.

So, what is the process of judging and just how do we perceive? The ability of not being able to perceive things that are in plain sight has coined the term *inattentional blindness*. One of the best-known examples, and one frequently shown to students of performance analysis, is the ‘Invisible Gorilla Test’. This was conducted by Simon and Chabris (source: en.wikipedia.org/wiki/Inattentional_blindness). Subjects were asked to watch a short video in which two competing teams wearing two different coloured T-shirts were passing a basketball around. In the original test, the subjects were asked either to count the number of passes made by one team or to keep a tally of the number of times the ball was passed either by bounce or by an aerial pass. In the famous video experiment, a person is seen to enter the field of play and walk around the basketball court for approximately 30 seconds, wearing a gorilla suit, before walking off screen. After watching the video, the subjects were asked to recount if they noticed anything unusual take place. A subsequent review of the data indicated that, in most groups tested, 50 per cent of the subjects failed to see the gorilla. One solution to this problem, in scoring and adjudication applications, is to utilise a video review to minimise potential inattentional blindness.

Observation, by its very nature, is subjective. What one person ‘sees’ or ‘perceives’ is never identical to what another person sees. Performance appraisal of aesthetic sports, such as ice skating, dressage or gymnastics, is almost entirely ‘in the eye of the beholder’. We assume and take for granted that judges are impartial, fair and do their job to the best of their abilities. However, the scoring system can be flawed or the appraisal process can have some inherent, built-in biases. No two people think alike. How people construct their own subjective reality is a separate branch of psychology and is influenced by our individual knowledge, experiences and thought processes. In making a judgement, the brain runs through a sequence of processes. Initially, the first step involves perception of the situation or, in the case of sport, this could be the execution of movement. Second, our minds place this event in context; it is given meaning, which is characterised using a mental process that relies heavily on stored prior knowledge and past experiences. This occurs within a framework of what is known about classifying performance. Lastly, the classified executed movement is augmented with additional information, such as relevant past memories, circumstantial information, venue, team or athlete reputation. One particular recognised bias is the so-called ‘halo effect’ (see en.wikipedia.org/wiki/Halo_effect). This is a cognitive bias where perceived traits influence interpretation of events due to expectations. You could say, ‘an athlete’s reputation precedes them!’ In the scoring/judging process, errors can creep in at any of the three stages outlined previously. Perhaps a judge’s view is obscured at the time of an event. For instance, in boxing, the judge’s view could be hindered by the position of the referee. Even if the event was observed equally by two judges, they may disagree on a number of aspects and be inconsistent in awarding or deducting points. This problem can be further compounded if these two judges are aware of both their individual scores. In this instance, a conformity bias effect may kick in and the two judges may adapt and align their scores. This could be considered a form of peer pressure and is a normative human characteristic of psychological behaviour. In dressage and gymnastics, bias has been detected when scoring/judging is done later in the competition (Ansorge and Scheer, 1988), with higher scores being reserved for gymnasts, or riders and horses that enter later in the draws; Wolfram (2010b) termed this order bias. Ice dancing is particularly prone to patriotic or nationalistic bias (Zitzewitz, 2006, 2010), where scores awarded are influenced by the country the athlete is representing. This effect has also been observed in other sports, including dressage and gymnastics.
Scoring/judging applications

(Popović, 2000), as well as the Eurovision Song Contest (Wolframm, 2010a). As a consequence of all these tendencies to bias, the first step in striving to improve the scoring and judging of sports is to be aware of these pitfalls when designing and implementing objective performance analysis systems for competition. The ultimate objective is fairness by excluding prejudice and to be achievable through accurate recall, consistent observation and scoring.

Current trends

In the preceding section, some common pitfalls in scoring and judging aesthetic sports were outlined. The current trend to reduce controversy and improve fairness in performance appraisal is to reduce the variability whilst increasing the accuracy of modern scoring/judging systems.

The driver for judging reform is often a perceived or real injustice that robs one athlete of a potential victory on the basis of bias detected in the scoring process. In the recent past, this has galvanised a number of sports, including ice skating and boxing, to review and amend their scoring/judging systems. The next section highlights the different approaches adopted by each of these codes in order to improve the reliability of adjudication so that the resultant process is robust and improves the probability of a fair outcome. It is beyond the scope of this chapter to catalogue the development of scoring systems in these two codes. The examples provide insight into modern approaches to solving a long-standing problem and could be equally applicable to other sports codes.

Ice skating

A highly controversial ice skating judging scandal at the 2002 Winter Olympics in Salt Lake City was the catalyst for speeding up adjudication reform in the sport. In the 2002 Games, the single panel of nine international judges utilised a 6.0 grading scale and deducted points for errors, mistakes or problems in execution. In the pairs figure skating competition, a firestorm of controversy broke out when the Russian pair of Yelena Berezhnaya and Anton Sikharulidze was awarded the gold medal over the Canadian pair of Jamie Salé and David Pelletier. The subsequent investigation into the judges’ scoring revealed that the French judge alleged that she had been pressured, by the head of the French skating organisation, to vote for the Russian pair, regardless of how the others performed (Swift, 2002). A subsequent investigation into the controversy by the world’s governing body and the IOC led to the award of a second gold medal to the Canadians.

The replacement international judging system (IJS) started at the fundamental level, by reviewing the decision process used to appoint judges. To reduce nationalistic bias, the International Skating Union (ISU) decided to appoint the judges directly rather than permit the national federations to do so. Furthermore, in an attempt to make collusion more difficult, the judging pool was increased from 9 to 14 and now currently consists of two panels, a judging panel and a new technical panel. The technical panel comprises of a technical specialist (also known as the caller), assistant technical specialist, technical controller, data operator and video replay operator. In the new system, all 14 scores would be publically reported, but only 9 would be used in computing the results. This means which scores were counted and which judge gave which score was not to be revealed, thus making any future determination of nationalistic bias impossible. Two more reforms were introduced: first, the technical merit and artistic impression scores were to be replaced with a system similar to that used in diving and gymnastics in which elements are assigned a predetermined score or degree of difficulty, known as Base Value.
A.N. Kirkbride

(BV), which is linked to a Scale of Values (SOV) calculated by a group of skating experts. It is the responsibility of the judges to grade the quality of execution, taking cognisance of the BV and SOV. Second, scores were to be aggregated and the trimmed mean or average of the middle five scores out of nine determined. For example, a salchow has a BV of 0.4, while a quadruple salchow has a BV of 8.5 in the SOV. The panel of judges must evaluate each technical element identified by the caller in the routine and assign a Grade of Execution (GOE, from +3 to –3). Consequently, judges have seven grades of execution from which to choose and these can either add or subtract from the score. These seven grades are also listed in the SOV. As an example:

- Salchow – BV = 0.4, and the possible GOEs are: –0.3, –0.2, –0.1, 0, +0.3, +0.6, +1.
- Quadruple salchow – BV = 8.5, and the possible GOEs are: –3, –2, –1, 0, +1, +2, +3.

The trimmed mean is determined when an equal number of the highest and lowest judges’ scores are excluded from the calculation and the average of the remaining grades is determined. For example, assuming a panel of nine judges and the assigned GOEs for a quadruple salchow are as follows: –2, 1, 1, 1, 2, 2, –1, 2, the two highest (2, 2) and the two lowest (–2, –1) grades are removed from the grouping. The average of the remaining numbers is calculated (1+1+1+1+2 = 6; 6÷5 = 1.2). The panel’s score for each element is then calculated by adding the trimmed mean (1.2) to the technical element’s base value (8.5) for a total (9.7). The panel’s scores for all the elements are added together to create a Total Element Score (TES). The next part of the adjudication process is to calculate the Programme Component Scores (PCS). The overall presentation and technical mastery is assessed on a scale of zero to ten and represents their overall skating ability and performance level and on five different components. These are:

- skating skills
- transitions/linking footwork
- performance/execution
- choreography/composition, and
- interpretation.

(Ice dancing is also graded on timing.)

The TES is then added to the five Programme Component Scores and any adjustments are applied to determine the Total Segment Score (TSS). Some examples of possible deductions include penalties incurred due to falls, time violations, music violations, illegal element violations and costume or prop violations. The final score or competition score is the total of the combined TSS, to produce the final competition ranking. More detailed information on the current system is available from the IJS Handbook (2010). The results of this modification are that the skaters earn points on what they successfully achieve and not what they attempt. Generally, modern ice skaters strive to undertake the most difficult elements for which the maximum grades of execution are available.

Boxing

Boxing is awash with controversial decisions in both the amateur (Olympic) and the professional codes of this contact sport. Knocking an opponent to the canvass in an amateur bout has no influence on the scoring/judging system, whereas in a professional bout, it counts for significantly more. Why should this inconsistency prevail? What is the rationale for the two
different scoring/judging systems in essentially the same sport? Once again, the impetus for review and change in the scoring/judging systems can be traced back to controversial decisions, in both codes.

Firstly, let’s examine the scoring system in the Olympics prior to 1988. In essence, the scoring was simple. Five judges scored the bout, a point was scored each time the white part of the glove landed on the front of the head/torso and each judge tallied the score on his card as the bout progressed. Controversy erupted when Roy Jones Jnr was robbed of the gold medal by a South Korean fighter, Park Si-Hun, despite pummelling Park for three rounds and landing 86 punches to Park’s 32 (www.royjonesjr.com/biography). A subsequent official IOC investigation revealed that three of the judges were wined and dined by South Korean officials and all three who voted against Jones were suspended. The incident led to the IOC insisting on the development of a new scoring system.

1992 new scoring system

1. Five judges score the bout.
2. One point is scored if the white part of the glove lands on the front part of the head/torso, if point (3) is satisfied.
3. Each judge has two electronic buttons, one for each of the contestants. Three out of five of the judges must press the button within one second for the boxer to receive a point.

Bouts are scheduled for four, two-minute rounds. No points are awarded or deducted for a knockdown in Olympic boxing. Consequently, the introduction of this system has led to a radical change in fighting style (IOC, 2005). Competitors no longer strive to knock out an opponent. The aim has shifted to scoring as many points as possible through correct blows to the head and torso. The introduction of electronic scoring was an attempt to make officiating more objective. Interestingly, a series of studies by Coalter et al. (1998) and Mullan and O’Donoghue (2001) failed to improve on the accuracy of the scoring system and found that a minority (less than one third) of actual scoring punches are scored. Judges focussing on counting blows may lead to secondary errors, such as inattentional blindness, and may outweigh the aim of the new system. Furthermore, when there is a flurry of blows being traded, the time taken for the judge to review and process the information may be outside the one-second window, even for highly experienced judges (Brown, 2008).

Scoring a professional bout is an inherently subjective process. A prize fight is judged on four basic criteria, namely:

1. Clean punching
2. Effective aggressiveness
3. Ring generalship
4. Defence.

Unless the bout ends in a knockout or other stoppage, the winner is decided on the basis of the score cards submitted by three judges. The boxing rounds, of varying duration and number depending upon the competition, are scored on a ten-point must system in which the boxer winning the round is awarded the maximum permissible, namely ten. Consequently, his or her opponent will receive nine (or lower). Knockdowns count and usually result in the deduction of an extra point from the opponent’s score. The abovementioned basic criteria are relatively self-explanatory, with perhaps the exception of ring generalship, which is the ability of a fighter
to control the pace and style of a fight. This basket of measures is subjective and consequently creates the necessary conditions under which boxing results, based on judges’ decisions, are too readily challenged. Furthermore, each criterion should have equal weighting (25 per cent); however, clean punches tend to count more than the other three areas. Interestingly, boxing broadcasters often keep a tally of the punch statistics, but this data is not currently used by the judges to objectively assess performance. Judges cannot see each other’s scores and cannot modify scores assigned from previous rounds. One of the most controversial results occurred on 13 March 1999 when heavyweight champions Lennox Lewis and Evander Holyfield fought to a draw. Many observers felt that Lewis had outperformed Holyfield, and one judge’s results, in particular, drew much of the attention (Lee et al., 2002). Eugenia Williams became the focus of speculation that the result of the contest was ‘fixed’ and an official investigation was prompted.

The match was watched by the three judges, using the same basket of measures described previously and following the common and well-developed ten-point scoring system. Table 12.1 shows the variability in scoring for each round and reproduces the score cards for the official judges and well-respected media commentators and pundits.

The first point of note is that there is variability in all the judges’ scores over the bout. The most consistent scoring, however, was between the first three unofficial scorers, who returned precisely the same scores over each round. Nevertheless, focussing on the official scorers, it is important to review the variability in their scores and compare these to the unofficial scorers, since this is the prime source of controversy expressed in the media. On the basis of what we know, how can the discrepancies in the score cards be explained?

<table>
<thead>
<tr>
<th>Scorer</th>
<th>Individual scores for the following rounds</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Officials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams (USA)</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>O’Connell (UK)</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Christodoulou (RSA)</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Media and others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HBO-Lederman</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Sportsticker</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Boxing Times</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Sportsline</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Associated Press</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>ESPN</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>Boxing Monthly</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

Note
H and L denote rounds in favour of Holyfield and Lewis, respectively, and E denotes even rounds (tie). All non-even rounds were scored 10–9 in favour of the fighter indicated.
over multiple rounds. Therefore, such impressions can be unduly influenced by fight statistics and slow-motion replays produced by the broadcasters and not available to the official judges. Furthermore, aggression and ring generalship are very subjective measures of performance, unlike clean punches and defence, which have slightly higher degrees of objectivity. Lewis won the rematch, by a decision.

In challenging bias in boxing or even corruption or incompetence in judging, it is first necessary to determine the variability of scoring that can lead to judges arriving at different conclusions after watching the same bout. The simple fact that they differ in opinion is not necessarily an indication of foul play. People always see things in slightly different ways and it is important to recognise the inherent variability in boxing scoring so that the sport can minimise its scoring controversies.

The role of technology in assisting the referee and umpire

The role of a referee or umpire is to make impartial and accurate judgements on both the spirit and the law of the sport. In some codes, such as soccer, they are the sole authority and their decision is final; in others, such as cricket, they can consult off-field umpires, who have access to a vast array of technology tools and can amend their decision according to the supplementary information available. Generally, referees and umpires are the on-field authority responsible for making discretionary decisions and enforcing the rules. Soccer referees make many hundreds of decisions per game and in the 2010 FIFA World Cup, 8 per cent of the decision events were inaccurate (Solomon et al., 2011). Cricket umpires make fewer decisions, but, in terms of percentage correct, are much more accurate. There are many options available to sports governing bodies to improve the accuracy of decision making in sport. Some opt for additional human intervention and others rely more on technology intervention.

In soccer, the accuracy of decision making has been found to be a function of the distance the referee is away from the infringement or action (De Oliveira et al., 2011). In their study of the 2002 under-20 Brazilian Championship, they indicated the optimum distance to achieve the maximum (80.6 per cent) correct decisions was 20.1 to 25 m. (Coincidently, in cricket, this is the approximate distance the facing umpire stands away from the facing batsman.) In order to minimise errors due to incorrect positioning of referees on the field, FIFA (1982) established a diagonal system of match control so that the referee and his or her assistants work in harmony to provide optimal adjudication cover on the pitch and can use communication devices to maintain contact (Mignerat and Audebrand, 2010). Furthermore, the experience level of the referee and crowd noise both influence the decision-making process. Home-team advantage is recognised and acknowledged in a wide variety of sports. Nevill et al. (2002) observed in their analysis a distinct bias when awarding fewer fouls to the home team. Gender bias has also been reported in handball (Souchon et al., 2009). However, bias in soccer is inherently difficult to quantify and much additional work is required in this area to determine the potential causal mechanism and its magnitude (Lucey and Power, 2009).

Prior to the introduction of referral technology in cricket, umpires operated at ±93 per cent correct decisions on average. With the introduction of the Umpire Decision Review System by the International Cricket Council (ICC), accuracy went up to 98 per cent (PTI, 2011). Rugby union and league use ‘video refs’ to make judgements on a number of rule infringements. In rugby league, the Television Match Official (TMO) is frequently called upon to review the grounding of the ball in try situations. In the view of many spectators, both in the stadium watching replays on big screen TVs and those viewing the broadcast at home, this anticipation of the result adds a new dimension to the game. Depending on the sport, the video ref can only adjudi-
cate on certain aspects. For instance, in rugby, he/she is not permitted to adjudicate on forward passes, since the reliability of these calls varies on the camera angle and is not dependable.

Today, television replays are the de facto norm during live sports broadcasts. The instant replay of an action almost immediately post-event has become an integral part of the broadcasters’ content package. This provides the armchair viewer the ability to review the decisions of on-field umpires and referees and quickly form their own opinions and conclusions. Since researchers have observed, with only ±80 per cent and ±93 per cent of correct decisions being made in soccer and cricket, respectively, the door was open for spectator criticism of an official’s performance during telecasts. Some sports codes permit the use of video review in official in-game adjudication and others specifically exclude it. The role of the video referee varies between sports and very little consistency exists. However, in many sports, video evidence is almost universally admissible post-match – for example, in cases of misconduct, disciplinary sanctions or other match infringements by players or spectators.

Basketball was one of the pioneering sports to opt for video review and its first recorded use was in the 2002–3 season to determine whether the Lakers player Samaki Walker scored a field goal in Game 4 of the NBA 2002 Western Conference Finals. The video replay showed the ball was still in his hand as the match clock expired.

The ability to show videos in slow motion or from multiple angles provide the TV viewer with much more information for decision support than was available to the referee or umpire. In cricket, normal video cameras placed perpendicular to the wicket are capable of more accurately determining close dismissals via run-out than the unaided eye. Hence, cricket fans watching at home are capable of increasing the probability of a correct umpiring call than on-field officials. Historically, when the broadcast feed was relayed into the stadium’s big TV screens, the images frequently caused frustration in the crowd when poor decisions were exposed. Consequently, in cricket, controversial decisions referred to the Third Umpire are not permitted to be shown on stadium screens until the final decision has been made.

From these simple beginnings, the video replay has evolved immensely, with high-speed (>1000 fps) cameras, thermal imaging, frame-by-frame analysis, graphical overlays, ball-trajectory tracking and such like finding application in a diverse range of sports.

In their quest to obtain the highest percentage accuracy of decision making, the ICC has driven the introduction of the Umpire Decision Review System (UDRS) over a period of years. It is now a cornerstone of all international matches, although not mandatory. It incorporates high-speed video cameras, audio analysis from stump microphones, approved ball-tracking technologies and thermal imaging. The UDRS (2011) is a complex process and a full explanation is available on-line from the ICC. In essence, each team is currently permitted two unsuccessful review requests per innings. The fielding team can challenge and dispute an umpire’s not out call and the batting team an out call. The decision is then referred to an off-field, TV or Third Umpire for review. Subsequently, a number of options are available in the post-event analysis, by either verifying the first decision or correcting it. In cases where there is no clear reason on review to overturn an umpire’s on-field call, then the original decision stands. Cricket is a game that can easily accommodate interruptions to play to permit the verification of decisions. Soccer, on the other hand, is a much more dynamic and fast-paced sport and, consequently, the governing body is reluctant to interfere during play and is currently resisting the introduction of the video replay in adjudication. However, one particular area of controversy and worthy of FIFA’s high-level investigation is that of goal-line technology. The objective is to assist the referee in determining when the ball has crossed the goal-line. It appears FIFA are currently hesitant to introduce technology that interferes with the very fabric of the game, since their opinion is that the human element is an integral part of the sport. There are currently a number of
Scoring/judging applications

proposals being evaluated by FIFA (up to March 2012), which include image processing (Hawk-Eye, Hawk-Eye Innovations, Basingstoke, UK) and sensors (Cairos, Cairos Technologies AG, Munich, Germany). Supporters of the proposed solutions contend that striving to obtain the correct decisions, in critical games during competition, outweighs the arguments regarding non-uniformity of rules and competitions. Due to the costs involved, the use of video tools in adjudication is generally reserved for use at the elite and professional levels.

In tennis, errors in line calls by chair umpires in the 2004 US Open stimulated the introduction of Hawk-Eye technology (en.wikipedia.org/wiki/Hawk-Eye#Tennis). By 2007, the system was used as an officiating aid on centre court with each player permitted access to the technology to challenge calls until they have made three incorrect challenges within the current set. The accuracy of the system is subject to much debate; however, it remains an engaging technology and one whose application is increasing worldwide.

On the horizon

The simplest of performance review processes in scoring and judging applications is the video replay. However, this tool is still to find universal acceptance in all sports. Other, more sophisticated approaches have been proposed that tend to focus on solving specific challenges in certain sports. In the majority of video review cases, human intervention interprets and then makes an informed decision based on the observed and reviewed content. Major research focus is shifting towards automated image processing in order to classify specific actions. Researchers are looking into real-time analysis of in-game events, such as goal-mouth incidents (Assfag et al., 2003; Wan et al., 2003) to automatically identify and annotate match events and actions. Although object tracking shows some promising results, the proposed systems are a long way from the accuracy and reliability necessary for adjudication purposes (Chen et al., 2009). In soccer, for the determination of off-side play, Maruenda (2009) has proposed that neither human nor machine is capable of error-free judging when determining an off-side position and, in his opinion, the rule should be scrapped.

The patent area in sports scoring and adjudication is exceptionally strong and a wide range of technologies have been proposed. These range from real-time sensors for determining when a basket is scored in basketball (Klein, 2009) to sensors on cricket wickets to determine whether the ball has been dislodged from the stumps (Eichstädt, 2003) and the ubiquitous ball tracking (Holthouse and van de Greindt, 2009). In taekwondo, the subjective nature of the scoring led to the development of a wearable vest (SensorHogu) that incorporates piezoelectric sensors that monitor the amount of force delivered to a body protector (Chi et al., 2004). In the area of computer-augmented systems, Reilly et al. (2009) have proposed a general-purpose taxonomy system aimed at assisting researchers and designers to clarify these systems with respect to both form and function.

Due to the problems highlighted previously in the scoring of boxing, this sport has not avoided the researchers’ attention (Hahn et al., 2010). In an attempt to develop a system that outperforms the current system of judging, boxing has also witnessed the application of sensors that detect impact in an attempt to automate the scoring. Unfortunately, one of the system’s current limitations is its inability to differentiate between valid scoring and non-scoring hits (Krajewski et al., 2011).

Cricket has seen the further refinement of the Snikometer, a technology utilised by broadcasters that uses the audio component from the stump microphone to indicate when the ball had hit either the bat/pad or glove. Further refinement of the audio signal was proposed as the basis to develop an intelligent cricket decision-making system (Ting and Chilukuri, 2009). It
A.N. Kirkbride

was proposed that the umpiring decision could be automated through audio analysis alone; however, poor accuracy limits its application. Nevertheless, it can be used very effectively in conjunction with video-based evidence.

One of the most entertaining and novel proposals is that of audience participation in scoring and judging. Van Beusekom et al. (2004) proposed that, by monitoring an audience’s response to performance (i.e. clapping, cheering and waving), a fun and entertaining system could be developed that ranks athletes using a system that captures the audience’s expressions and calculates an associated score. As an interesting concept, it has merit, but a number of serious drawbacks would limit its applicability in elite level sport (e.g. audience response along partisan lines or national pride, spectators not familiar with the rules or making inputs based on ‘gut feel’).

Concluding remarks

This chapter has provided an overview of the common problems encountered in scoring and judging of sport. The often subjective nature of the process of adjudication has been described and the most common approaches to improve reliability and reduce bias outlined. Two sports in particular, boxing and ice skating, were chosen as exemplars to highlight the challenges faced in both scoring and judging competitive sport. The role of adjudication in umpiring and refereeing was also described, as well as a technology scan of the aids currently on the horizon aimed at improving the decision-making process.

Note

1 Historically, the ICC published the percentage correct decisions by each umpire on the Elite and International Panel of Umpires on their web site. The publication of this sensitive data in the open press has now ceased. However, performance analysts working with elite level teams often gather this type of information to support the development of their team’s tactical match plans.

References

Scoring/judging applications


decision making about transgressions: The influence of player gender at the highest national level’,


