Visual Scan Patterns and Decision-Making Skills of Expert Assistant Referees in Offside Situations

Peter Catteeuw, Werner Helsen, Bart Gilis, Evelien Van Roie, and Johan Wagemans
Katholieke Universiteit Leuven

The offside decision-making process of international and national assistant referees (ARs) was evaluated using video simulations. A Tobii T120 Eye Tracker was used to record the eye movements. Two hypotheses for explaining incorrect decisions were investigated, namely, the flash-lag effect and the shift of gaze. Performance differences between skill levels were also examined. First, results showed a bias toward flag errors for national ARs as expected by the flash-lag effect. Second, ARs fixated the offside line before, during, and after the precise moment the pass was given, implying there was no shift of gaze from the passer to the receiving attacker. Third, no differences were found in scan patterns between international and national ARs. In conclusion, international ARs seem to have found a strategy to better deal with the perceptual illusion resulting from the flash-lag effect. Based on their experience, they have learned to correct for this illusion, and, consequently, show fewer flag errors.

Keywords: expertise, flash-lag effect, eye movements, decision making, offside

In association football (known as soccer in North America), not only the players, but also the referees strongly depend on their decision-making skills to perform well on the field (Lane, Nevill, Ahmad, & Balmer, 2006). Considering the total amount of 137 observable technical decisions (Helsen & Bultynck, 2004) and changes in activity in relation to the evolving game incidents (referees: 1,268; assistant referees: 1,053) (Krstrup & Bangsbo, 2001; Krstrup, Mohr, & Bangsbo, 2002), a top-class referee and the assistant referees (ARs) take a game-related decision approximately every 2 s, given that there is an effective playing time of 55 min (FIFA, 2006).

During match play, the major responsibility of ARs is the correct and consistent application of the offside rule. Ever since the Laws of the Game were first set

Catteeuw, Helsen, Gilis, and Van Roie are with the Laboratory of Perception and Performance, Research Center for Movement Control and Neuropsychology, Department of Biomedical Kinesiology, Faculty of Kinesiology and Rehabilitation Sciences, Katholieke Universiteit Leuven, Heverlee (Leuven), Belgium. Wagemans is with the Laboratory of Experimental Psychology, Department of Psychology, Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Leuven, Belgium.
down in 1863, the offside law has been one of the most controversial. The modern version of the Fédération Internationale de Football Association (FIFA) Law 11 states that “a player is in an offside position if he is nearer to the opponents’ goal line than both the ball and the second last opponent. A player in an offside position is only penalized if, at the moment the ball is touched or is played by one of his team, he is, in the opinion of the referee, involved in active play by interfering with play, or interfering with an opponent, or gaining an advantage by being in that position” (FIFA, 2007).

Helsen, Gilis, and Weston (2006) analyzed the accuracy of the offside judgments during the 2002 FIFA World Cup in Japan and Korea and revealed that 26.2% of the offside situations were assessed incorrectly. Specifically, they found a bias toward flag errors (the AR raises his flag while the attacker is in an onside position at the moment the ball is played) in comparison with non-flag errors (the AR fails to raise his flag while the attacker is in an offside position). Three major perceptual-cognitive explanations have been proposed that may affect the quality of offside decisions. First, Oudejans et al. (2000, 2005) suggested the optical error hypothesis to explain erroneous judgments in offside decisions. These authors stated that the observation point of the AR relative to the offside line, and the corresponding angle of view, is an important determinant in judging offside. Specifically, an inappropriate positioning toward the offside line (i.e., trailing or leading the line) may cause errors as an inevitable result of the relative optical projections of the players on the retina of the AR.

A second explanation for the underlying mechanisms leading to errors in judging offside was suggested by Baldo, Ranvaud, and Morya (2002). They referred to a perceptual illusion, called the flash-lag effect (Nijhawan, 1994), in which a moving object is perceived as spatially leading its real position at an instant defined by a time marker (usually a briefly flashed stimulus). In the case of offside situations in football, the attacker receiving the ball is a moving stimulus and is perceived ahead of his actual position at the moment the ball is passed (an abrupt event). This leads to the expectation of an overall bias toward flag errors in comparison with non-flag errors.

A third explanation for misjudgments in offside situations was suggested by Belda Maruenda (2004) and Sanabria, Cenjor, Márquez, Gutierrez, Martinez, and Prados Garcia (1998). These researchers hypothesized that errors in judging offside were due to the time taken to shift gaze from the player releasing the ball to the player receiving the ball, implying a time delay. When the eyes move to focus on objects, they perform saccadic movements, smooth pursuit movements, vergence movements, vestibular movements, and accommodation, all causing a latency time during which the players move and the positions change (Belda Maruenda, 2004). However, Oudejans et al. (2000) found that these saccadic movements are an unlikely explanation for offside errors because an AR equipped with a head-mounted camera showed no shift of gaze from passer to receiver at the moment the ball was played. In case there is no shift of gaze, the question remains what ARs exactly focus on. If the ARs fixate the second last defender, it is worth investigating whether they are able to perceive the passer in their peripheral field of vision.

In the current study, the first aim was to investigate the scan pattern of ARs in offside decisions in football. We hypothesized that ARs fixate the second last defender at the precise moment of the pass when assessing offside situations in laboratory tasks, thereby ruling out the shift-of-gaze hypothesis (Belda Maruenda,
2004; Sanabria et al., 1998) and supporting the flash-lag effect (Baldo et al., 2002; Gilis, Helsen, Catteeuw, & Wagemans, 2008; Gilis, Helsen, Catteeuw, Van Roie, & Wagemans, 2009; Helsen et al., 2006) as a potential explanation for flag errors.

Furthermore, we were interested to know if the scan pattern could account for differences in response accuracy between international and national ARs. On the one hand, it can be hypothesized that international ARs show a more efficient scan pattern and thus look at different sources of information in comparison with national ARs. According to previous research in time-constraint decision-making tasks, efficient scan patterns can be quite ambiguous. In some studies, experts typically show fewer fixations of longer duration when compared with novices (Goulet, Bard, & Fleury, 1989; Helsen & Starkes, 1999; Ripoll, Kerlirzin, Stein, & Reine, 1995). In other studies, experts show more fixations of shorter duration (Vaeyens, Lenoir, Williams, & Philippaerts, 2007; Williams & Davids, 1998; Williams, Davids, Burwitz, & Williams, 1994). Alternatively, both groups of ARs may look at the same source of information, but may use it in a different way.

A second aim of the current research was to test whether the two groups differed in their response accuracy and to examine which factors might contribute to a possible difference. Based on Gilis et al. (2008), the international ARs were expected to be more accurate in their offside decisions than their national counterparts. In line with previous findings supporting the flash-lag effect (Gillis et al., 2008; Helsen et al., 2006), we assumed an overall bias toward flag errors in comparison with non-flag errors.

**Methods**

**Participants**

Two groups of five Belgian ARs were selected. The first group consisted of international ARs (n = 5; M = 40.0 years, SD = 2.7, range = 35.5–42.4) who were on the FIFA list for 4.6 years on average (SD = 1.2, range = 3.0–6.0) and had experience at the highest level of the Belgian professional football league for 6.1 years on average (SD = 1.2, range = 4.3–7.4). The second group of ARs (n = 5; M = 39.4 years, SD = 5.1, range = 32.6–43.8) had no exposure to international football refereeing and had experience at the highest level of the Belgian professional football league for 5.7 years on average (SD = 3.7, range = 2.3–12.0).

Written consent was obtained from the Belgian referees’ committee. The study was designed and conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and approved by the Committee for Ethical Considerations in Human Experimentation of the Faculty of Kinesiology and Rehabilitation Sciences from the Katholieke Universiteit Leuven.

**Task**

First of all, simulations of offside situations, played by elite players of a team of the second division in Belgium, were recorded. A digital video camera (Canon, DV Camcorder MV4) was placed at the touch line in such a way that these video simulations were recorded from the perspective of the AR. Three attackers were playing against two defenders and one goalkeeper. Out of a total of 90 offside actions, 40
Expert Perception in Offside Decision Making

high-quality situations were selected. Initially, the selection of these clips was based on the positioning of the second last defender relative to the camera position. Only those situations exactly in front of the camera were selected to eliminate an incorrect angle of view. Next, we categorized the situations in four groups of 10 situations according to the position of the attacker relative to the offside line. First, in situations far behind the offside line, the attacker was more than 1 m behind the second last defender. Second, in situations slightly behind the offside line, the attacker was less than 1 m behind the second last defender. Third, in situations on the offside line, the attacker was at the same level of the second last defender. And fourth, in the situations with the attacker slightly ahead of the offside line, the attacker was less than 1 m ahead of the second last defender.

**Apparatus**

The video simulations were played on a Tobii T120 Eye Tracker. Visual scan patterns are recorded at a constant frame rate of 120 Hz using five infrared lights and a camera mounted below the 17-inch screen. In addition, it allows for gaze angles up to 40° in the semicircle above the camera and up to 10° below the camera. Considering potential head movements over longer periods of time, the Tobii T120 Eye Tracker boasts an effective accuracy of approximately 1° across the entire screen (user manual; Tobii Technology AB, Danderyd, Sweden).

**Procedure**

The ARs were seated in front of the Tobii monitor, with their heads fixated on a chin rest to standardize the distance between the screen and the participant’s eyes at precisely 60 cm. Before the actual test, standardized instructions were read. In addition, five familiarization trials were given. The video simulations were randomly distributed in two sets of 20 clips. A simple five-point eye calibration was performed to verify point of gaze before each block of offside situations. For every situation, the ARs had to decide whether the attacker was positioned offside by pressing the appropriate button on a keyboard, one corresponding with “offside” and another corresponding with “no offside.” The participants were asked to respond as accurately as possible after the final pass. These answers were immediately and automatically registered with the Tobii Studio 1.3.21 software.

**Dependent Variables and Analysis**

*Response Accuracy.* First, the percentage of correct and incorrect decisions was calculated. For the correct decisions, a distinction was made between correct flag (CF) or correct non-flag (CNF) signals when the attacker was in an offside or onside position, respectively. When the AR made an incorrect decision, a distinction was made between a flag error and a non-flag error: The AR pressed offside on the keyboard although the attacker was in an onside position (flag error), or he decided no offside when the attacker was in an offside position (non-flag error).

*Point of Gaze Data.* A fixation was defined as the period of time when the eyes remained stationary within 58 pixels or 1 degree of movement tolerance for a
period equal to, or greater than, 120 ms (Helsen & Starkes, 1999; Savelsbergh, Van der Kamp, Williams, & Ward, 2005; Williams, Davids, & Williams, 1999). To investigate the scan pattern, the number of visual fixations and the mean fixation duration per trial were calculated. To examine the point-of-gaze data in more detail, two areas of interest were defined, namely, the passer and the offside line. Only two areas of interest were used because there was a considerable overlap between the attacker and the second last defender in most of the incidents. In addition, the percentage viewing time was examined (i.e., the amount of time the ARs spent fixating the various areas of interest). These point-of-gaze data were explored at three important moments of each clip: the precise moment of the final pass (“pass”), the period just before the pass (“before”), and the period between the pass and the AR’s decision (“after”). The start of the period “pass” was chosen 100 ms before the pass, whereas the end of the time window was 100 ms after the pass. Likewise, the period “before” had a total duration of 200 ms and ended when the period “pass” started. Finally, the period “after” started when the period “pass” was finished. However, the duration of this last period depended on the decision time of the AR. Therefore, this period was different for every participant, but the main interest was on the spatial position of the fixations.

**Data Analysis.** The overall data analysis was in line with previous work (Gilis et al., 2008). First, a chi-squared test was used to examine whether there was an overall bias toward flag errors in comparison with non-flag errors and whether the ratio of correct versus incorrect decisions was uniformly distributed across the different conditions.

Second, signal detection theory (Macmillan & Creelman, 2005) was used. The sensitivity index $d'$, found by comparing the hit rate (i.e., CFs) to the false alarm rate (i.e., flag errors), gives valuable information about the sensitivity of the ARs to detect offside. The response bias $c$ examines the preference of ARs to flag or to keep the flag down in case of doubt. To determine whether $d'$ and $c$ were significantly different from zero, 95% confidence intervals (CI) were calculated.

Finally, the Mann–Whitney $U$ test was used to study the differences in response accuracy and eye movement data between the two groups. We preferred this non-parametric test over the parametric $t$ test for independent samples, because, owing to the small sample size, we could not accept the assumption that our samples are normally distributed (Portney & Watkins, 2000). Cohen’s $d$ was calculated as a measure of effect size. Significance level was set at .05 for all analyses.

**Results**

**Response Accuracy**

The Mann–Whitney $U$ test showed that international ARs ($M = 83.5\%, SD = 7.0$) were more accurate in their decisions when compared with their national counter-
parts \((M = 74.6\%, SD = 4.8; U = 2.5, p < .05, d = 1.2)\). A trend toward a difference was found when the response accuracy of both groups was compared for only the situations with the attacker in an onside position (international ARs: \(M = 82.0\%, SD = 9.6\); national ARs: \(M = 70.7\%, SD = 7.2\); \(U = 3.5, p < .10, d = 1.1\)), whereas no difference was apparent for the situations with the attacker in an offside position (international ARs: \(M = 88.0\%, SD = 4.5\); national ARs: \(M = 86.0\%, SD = 8.9\); \(U = 9.5, d = 0.3\)).

Furthermore, both international \((d' = 2.20, 95\% CI 1.56–2.84, p < .05)\) as well as national ARs \((d' = 2.06, 95\% CI 0.54–3.58, p < .05)\) were able to discriminate between offside and no offside above chance level. In doubtful situations, international \((c = –0.09)\) and national ARs \((c = –0.48)\) showed a tendency to raise their flag, although these differences were not significantly different from zero and not from each other.

Regarding the type of error, no bias toward flag errors was found for the group of international ARs (flag errors 27/150, 18.0%; non-flag errors 6/50, 12.0%; \(\chi^2 = 0.82, \chi^2_{CV} = 3.841\)), whereas national ARs tended to make more flag errors \((44/150; 29.3\%)\) compared with non-flag errors \((7/50; 14\%; \chi^2 = 3.46, \chi^2_{CV} = 3.841)\).

When we examined the error scores of the international ARs in more detail, the results showed the highest error percentage for those situations in which the attacker was on the offside line \((21/50; 42\%)\), followed by the situations in which the attacker was slightly behind \((4/50; 8\%)\) and ahead \((6/50; 12\%)\) of the line, and very few errors for the simulations in which the attacker was far behind the offside line \((2/50; 4\%)\) \((\chi^2 = 27.24, p < .001)\). Similar results were found for the national ARs. The situations with the attacker on the offside line were the most difficult to judge, resulting in an error percentage of 72.0% \((36/50)\). Almost every assessment for the situations with the attacker far behind the offside line was correctly handled, with only one flag error in 50 offside simulations \((2\%)\). For the simulations with the attacker slightly behind and ahead of the line, the same error percentage \((7/50; 14\%)\) —significantly lower than the error score for the situations with the attacker exactly in line with the defender \((\chi^2 = 58.41, p < .001)\) —was found.

**Point-of-Gaze Data**

The Mann–Whitney \(U\) test showed no significant differences between the two groups of participants for the number of fixations and the mean fixation duration per clip. In addition, when the data were specified for the offside line, differences were of no significance (Table 1).

The ARs spent the greatest amount of time fixating upon the offside line. However, the international ARs showed a different pattern compared with the national ARs after the pass. They spent less time fixating upon the offside line \((U = 2.0, p < .05, d = 1.5)\) and relative more time fixating upon another area, specifically, the receiving attacker \((U = 1.0, p < .05, d = 1.2)\) (Figure 1).
<table>
<thead>
<tr>
<th>Variable</th>
<th>M (1)</th>
<th>Rank Sum (1)</th>
<th>M (2)</th>
<th>Rank Sum (2)</th>
<th>U</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fixations</td>
<td>3.87</td>
<td>24.0</td>
<td>4.68</td>
<td>31.0</td>
<td>9.0</td>
<td>-0.731</td>
<td>.465</td>
</tr>
<tr>
<td>Mean fixation duration (s)</td>
<td>1.26</td>
<td>32.0</td>
<td>1.08</td>
<td>23.0</td>
<td>8.0</td>
<td>0.940</td>
<td>.347</td>
</tr>
<tr>
<td>Number of fixations on offside line</td>
<td>4.75</td>
<td>28.5</td>
<td>4.68</td>
<td>26.5</td>
<td>11.5</td>
<td>0.209</td>
<td>.835</td>
</tr>
<tr>
<td>Mean fixation duration on offside line (s)</td>
<td>0.44</td>
<td>27.0</td>
<td>0.47</td>
<td>28.0</td>
<td>12.0</td>
<td>-0.104</td>
<td>.917</td>
</tr>
<tr>
<td>Number of fixations not on offside line</td>
<td>2.03</td>
<td>23.5</td>
<td>2.59</td>
<td>31.5</td>
<td>8.5</td>
<td>-0.836</td>
<td>.403</td>
</tr>
<tr>
<td>Mean fixation duration not on offside line (s)</td>
<td>0.08</td>
<td>27.0</td>
<td>0.09</td>
<td>28.0</td>
<td>12.0</td>
<td>-0.104</td>
<td>.917</td>
</tr>
</tbody>
</table>
Figure 1 — Percentage viewing time of the international \((n = 5)\) and national \((n = 5)\) ARs for the different areas of interest (AOI) before, after, and at the exact moment of the final pass. Significant differences between groups are marked with an asterisk (*) . Error bars indicate the standard error.
Discussion

In the current study, our purpose was to gain a better understanding of the underlying factors contributing to errors in judging offside in association football. Therefore, video simulations of offside situations were created in an attempt to construct a laboratory test that is more closely related to the real game than 3-a-side computer animations used by Gilis et al. (2008). Unlike previous research, the Tobii T120 Eye Tracker was now used to investigate the scan pattern of ARs while assessing offside.

One of the possible explanations for errors in judging offside, the optical error hypothesis (Oudejans et al., 2000, 2005), was ruled out by the fixed position of the camera on the offside line.

A more viable explanation in our study may therefore consist of a perceptual illusion, referred to as the flash-lag effect (Baldo et al., 2002). Gilis et al. (2008) have found evidence for the flash-lag effect through the use of computer animations. In the current study, we expected to confirm the conclusions of these authors in video simulations. First, as hypothesized, an overall bias toward flag errors in comparison with non-flag errors was found for the national ARs.

The second finding to support the flash-lag hypothesis is a high increase in flag errors when the distance between the attacker in an onside position and the defender decreases. Both groups of ARs had a similar pattern of errors, namely, the highest error score for the situations on the line. Furthermore, an increase was found in error percentage for those situations in which the attacker was on the line compared with the situations in which the attacker was far and slightly behind the offside line.

There are several reasons to believe that the perceptual illusion induced by the flash-lag effect may indeed contribute to the explanation of errors in judging offside in laboratory tests, especially for national ARs. However, no exclusive confirmation was found for the international ARs, who did not make more flag errors compared with non-flag errors. Therefore, it seems as if these ARs have learned to deal with the flash-lag effect. At this point, we can only offer this as a tentative suggestion because in a previous study (Gulis et al., 2008) also international ARs showed more flag errors than non-flag errors. One reason might be that the current study used Belgian ARs only, who have become quite familiar with our work, by having participated in our studies quite regularly and who have also attended seminars about our research.

When we examine the difference in type of errors between international and national ARs into more detail, the main conclusion is that international ARs are more accurate in their decisions when the attacker is in an onside position. Mascarenhas, O’Hare, and Plessner (2006) suggested signal detection methodology to test whether ARs are more motivated toward detecting offside and thus make more flag errors than non-flag errors. In this study, both international as well as national ARs were able to discriminate between offside and onside. However, international and national ARs had the tendency to flag in case of doubt. The limited number of participants may have had an impact in this respect.

For a better understanding of the differences between the two groups and the underlying mechanisms leading to errors, we also analyzed point of gaze. In this way, we tried to find out whether international ARs look at different things and
employ a more effective scan pattern when assessing offside in comparison with national ARs, or whether they somehow learned to better deal with the perceptual illusion resulting from the flash-lag effect and representational momentum. To find an answer to this question, the point-of-gaze data of international and national ARs were investigated using a Tobii T120 Eye Tracker. First, the hypothesis of Belda Maruenda (2004) and Sanabria et al. (1998) that misjudgments in offside situations are due to the time delay caused by a shift of gaze from passer to receiving attacker, was investigated. However, Oudejans et al. (2000) found the shift-of-gaze hypothesis an unlikely explanation for errors in judging offside. In the current study, the results clearly showed that the ARs were fixating on the offside line at the exact moment of the pass. More than 80% of viewing time was used to focus on the second last defender. In addition, the moments before and after the pass were mainly characterized by a focus on the offside line instead of other areas of interest. At least in laboratory offside tests, it appears that ARs are most of the time looking at the offside line during offside situations and make no saccadic eye movement within the 300 ms before the final pass. Second, very few differences were found between the two groups of ARs. Although international ARs were more accurate in their judgments, they did not show different scan patterns when compared with national ARs. One exception was found for the period after the pass. The national ARs fixated mostly on the second last defender, whereas international ARs made a shift of gaze and smoothly followed the attacker. When we make the transfer to the real game, ARs, in case of no flag signal, need to follow the attacker in possession of the ball after the final pass.

In conclusion, this study presented further evidence for the perceptual illusion induced by the flash-lag effect (Baldo et al., 2002; Gilis et al., 2008; Helsen et al., 2006). International ARs appear to deal in a more efficient way with this perceptual illusion. These ARs have learned, based on their international experience and related instructions, to deal with flag errors, sometimes even causing an overcorrection and relatively more non-flag errors. However, overall, international ARs score significantly better in laboratory offside tests than their national counterparts. Furthermore, and in line with Oudejans et al. (2000), the hypothesis of a shift of gaze at the precise moment of the pass was rejected for these laboratory tests (Belda Maruenda, 2004; Sanabria et al., 1998). The ARs were already looking at the second last defender at the moment of the pass, and even before the pass, they mainly focused on the offside line. It has to be mentioned that the behavior of ARs can be different on the field, where the passer is probably not always perceivable by peripheral vision. Based on this study, with its limitations in mind, we can state that the difference in response accuracy between two groups of experienced ARs is not the consequence of a different scan pattern. In line with previous research (Bard, Fleury, & Goulet, 1994), subjects can look at an object but extract information through peripheral vision without using saccades associated with temporary blindness. Although the same information might be captured on the fovea of the ARs, the active information pickup and the meaning given to the visual information that is available to both groups can be quite different (Abernethy, Neal, & Koning, 1994; Williams & Davids, 1998; Williams et al., 1999). Therefore, using their experience, international ARs perceive and use the informative sources more accurately and, by consequence, take more correct decisions.
References


*Manuscript received: December 24, 2008*

*Revision accepted: July 30, 2009*