

The impact of a team's aggressive reputation on the decisions of association football referees

MARC V. JONES,* GEOFFREY C. PAULL and JOHN ERSKINE

Sport, Health and Exercise, School of Health, Staffordshire University, Stoke-on-Trent ST4 2DF, UK

Accepted 24 September 2002

It has been suggested that individuals may use heuristic methods of reasoning and rely on schemata when a quick decision is necessary. Accordingly, it is possible that decisions made by sport officials may be influenced by prior knowledge they have about teams they are officiating. The aim of the present study was to determine whether sport officials are more likely to penalize individuals who participate in a team with an aggressive reputation. In a balanced design, 38 football referees were randomly assigned to either an experimental or control group and were presented with the same 50 video clips of incidents from football games, all involving a team in a blue strip ('blue team'). The incidents were categorized before the study by five experienced referees into fouls committed both by, and against, the blue team, where all participants agreed that a foul had been committed (certain incidents), disagreed it was a foul (uncertain incidents) or agreed that there was no foul (innocuous incidents). Both groups received written instructions before the task; in addition, the experimental group was informed that the blue team present in all of the clips had a reputation for foul and aggressive play. For each incident, the participants were required to indicate what action they would engage in if refereeing the game. Although there was no difference in the number of decisions made, the experimental group awarded significantly more red and yellow cards against the blue team both overall and for the 'certain' incidents. It is suggested that prior knowledge may impact referees' behaviour in a laboratory setting, although future research should explore whether a similar effect is observed in the behaviour of referees during football matches.

Keywords: aggressive reputation, bias, decision-making, football, recognition-primed decision model, referees, soccer, sport officials.

Introduction

Understanding the factors that impact upon the decisions of individuals when a quick and accurate response is required (e.g. fire-fighters, military commanders, police officers) is important. One arena that provides an excellent opportunity to study these factors is sport, as both players and officials are often required to make decisions under intense time pressure that may have a substantial impact on the outcome of a game. Although most research has focused on decision-making among athletes (e.g. Abernethy, 1987), the present study was concerned with decision-making among sport officials; specifically, whether prior knowledge of a team's aggressive reputation would influence the decisions made by those officials.

The notion that, under certain circumstances, sport officials may exhibit a bias in their decision-making and favour a particular team is supported by previous research. For example, it has been suggested that officials may make more decisions in favour of the home team (Courneya and Carron, 1992) as a consequence of pressure from the home fans (Sumner and Mobley, 1981; Nevill *et al.*, 1996; Nevill and Holder, 1999). However, of particular relevance to the present study is research suggesting that prior knowledge and expectations held by officials could influence the decisions they make. For example, Frank and Gilovich (1988) suggested that officials had a tendency to judge black-clad players more harshly, because the colour black is associated with aggressiveness, and officials would perceive more aggression or intent in the actions of players wearing a black uniform. In support of this, Frank and Gilovich reported that professional American football and ice hockey teams wearing black uniforms were

* Author to whom all correspondence should be addressed. e-mail: marc.jones@staffs.ac.uk

penalized significantly more often than teams with uniforms of a different colour. In addition, strong identification with one team over another is also proposed to influence officials' behaviour. Mohr and Larsen (1998) investigated whether Australian rules football teams from outside the state of Victoria were penalized more frequently than teams from Victoria, in matches when all the umpires came from Victoria itself. Mohr and Larsen hypothesized that, in such instances, all the match umpires would share a regional identity with the teams from Victoria and, accordingly, they would perceive the players from the Victorian teams (in-group) more favourably than players from teams outside the state (out-group). In line with this hypothesis, they found that teams from outside Victoria were penalized on 11% more occasions than teams from Victoria. Bias may also be demonstrated towards individual athletes. For example, Plessner (1999) reported that altering the order with which judges were presented with participants competing in a range of gymnastic routines (pommel horse, vault and horizontal bar) had a significant effect on the mark awarded. When the judges believed that the routines were being executed by a gymnast competing in the fifth position for their team (where the strongest member of the team would compete), gymnasts were awarded significantly higher marks than when judges believed that the same routines were being executed by a gymnast who was competing in first position for their team (where the weakest member of the team would compete).

Interestingly, it is not only expectations held before a sporting contest that influence decisions. Plessner and Betsch (2001) reported that decisions may also be influenced by events that occur during the contest itself. Having asked licensed referees and players to watch and adjudicate on videotaped scenes from a football match in the Spanish Primera Division, Plessner and Betsch found a negative correlation between participants' successive penalty decisions involving the same team. That is, participants were less likely to award a penalty to a team if they had previously awarded the same team a penalty. Conversely, they found a positive correlation between successive penalty decisions concerning first one and then the opposing team. In other words, participants were more likely to award a penalty to a team if they had previously awarded a penalty to the opposing team.

Research to date suggests that although sport officials may strive for fairness and impartiality, favouritism towards one team may occur, based on such diverse factors as game location (Nevill and Holder, 1999), origin of the team (Mohr and Larsen, 1998), colour of the outfit (Frank and Gilovich, 1988), order of competition (Plessner, 1999) and previous decisions (Plessner and Betsch, 2001). However, one factor that

has not been investigated to date is the impact of prior knowledge about the aggressive reputation of the players or teams concerned. This is an important variable to study given that players deemed to have behaved aggressively may be sent from the field of play in many sports, possibly weakening their team considerably and having a substantial impact on the outcome of the game. In addition, for players to accept and respect the decisions of officials, it is important that players believe these decisions have been made in an impartial manner. Interestingly, though, there is anecdotal evidence from association football where both players and managers have indicated that individuals with aggressive reputations are treated differently to other players. For example, the Leeds United player Danny Mills was quoted by Daniel Taylor (2002) in *The Guardian Sport* as saying, 'It might come across as being petty and I don't want to look like I'm being childish. But it's obvious to me that certain players are looked upon differently. Every tackle is under the magnifying glass' (p. 6). To explain why aggressive reputations may have an impact on the behaviour of sport officials, it is necessary to consider in detail the role of prior knowledge and expectations in the way information is processed and decisions are made.

Individuals may rely on heuristics to make decisions about uncertain events in what are often complex environments. These are simple ways of reasoning to help guide judgement (Tversky and Kahneman, 1974; Kahneman and Tversky, 1984), whereby an individual may draw on prior knowledge to interpret a situation, rather than relying on information gleaned solely from the situation itself. Thus the use of prior knowledge, categorized in mental structures called 'schemata', is considered valuable in enabling individuals to process information, particularly when a quick decision is necessary (Fiske and Taylor, 1991; Hogg and Vaughan, 1995). Schemata can be acquired through either direct experience or communication with others (Fiske and Taylor, 1991). Therefore, an official could consider a team aggressive if he or she had personally witnessed aggressive behaviour by members of the team, or had simply been informed of such aggressive behaviour by other officials. If an official relies on heuristics and draws on schemata to process information, this may impact the way an incident is interpreted. For example, a football referee, having witnessed a crude challenge by a player from a team with an aggressive reputation, may consider this crude challenge as a deliberate attempt to injure an opponent, rather than a mistimed tackle.

To further clarify how schemata impact on decision-making, it is necessary to consider the process of decision-making in sport settings; one model that appears to have utility in this regard is Klein's (1989, 1993) 'recognition-primed decision' model. This

model was specifically developed to take account of decision-making in natural settings, where the decision-maker is usually very experienced and decisions have to be made under intense time pressure (e.g. fire-fighters, military commanders). According to Klein (1993), decision-making is not about choosing among a range of possible options (as analytical strategies of decision-making suggest), but about recognizing the situation and responding in an appropriate way on the basis of prior knowledge and experience. Thus, decisions are based on features of the situation and the person's knowledge and experience (schemata), from which they generate a possible option and evaluate its appropriateness (Orasanu and Connolly, 1993). Decision-making involves four aspects: plausible goals, relevant cues, expectancies and actions. For example, consider a football referee who witnesses an ill-timed challenge (relevant cues). He or she believes that they need to act and discipline the player quickly (plausible goals) to prevent other players reacting inappropriately (expectancies) and, therefore, speaks to the player concerned (actions). In many cases, recognition of the offence would lead to an obvious course of action (e.g. stamping repeatedly on an opponent's head during a ruck in rugby will lead to an automatic sending off). Klein also suggested that, in some circumstances, the decision-maker would consciously evaluate the impact of their decision before deciding whether to implement it, as in the case of a football referee deciding to penalize a player for a careless foul. In this instance, the referee may consider that penalizing the player is sufficient and no additional verbal reprimand is necessary. However, if the evaluation indicates a significant likelihood of escalating violence from that player, then the referee may move on to consider the next most favourable option (e.g. speak to the player concerned) and so on in a sequential format.

One relevant sports officiating study based on Klein's recognition-primed decision model was conducted by McLennan and Omodei (1996). They conducted post-match interviews with three Australian football league umpires about their decision-making processes during matches, concluding that the umpires were able to respond rapidly to incidents as they had already executed mental simulations of what was likely to occur on the basis of their prior knowledge of the way the game is played. This finding suggests that prior knowledge and expectation of events play an important role in the decision-making process (Paull and Glencross, 1997). The notion of anticipatory bias and priming through probability setting has been investigated extensively in perceptual-motor skills domains (e.g. Abernethy *et al.*, 1993; Paull and Glencross, 1997). As a strategy in elite performance, this process prepares for faster recognition of more probable scenarios if they do, in fact, arise. This is held to occur through the

establishment of sub-threshold neural priming in cognitive structures to facilitate rapid final activation when 'condition' recognition occurs. Of course, while prior knowledge and expectations may facilitate correct and rapid decision-making in most circumstances, McLennan and Omodei (1996) pointed out that it may also help to explain incorrect decisions by officials who may become trapped into an anticipated set of possible outcomes that do not occur.

The aim of the present study was to determine whether prior knowledge influences decisions made by sport officials. Association football referees were randomly assigned to one of two groups. One group of referees (experimental group) was told that the team involved in all the incidents had a history of violent play and an aggressive reputation. In line with suggestions that individuals may use heuristic methods of reasoning and rely more on schemata when a quick decision is necessary (Fiske and Taylor, 1991; Hogg and Vaughan, 1995), we anticipated that decisions made by sport officials would be influenced by the prior knowledge and expectations they have about teams they are officiating. We hypothesized that the experimental group would award significantly more decisions, and more red and yellow cards, against the team with an aggressive reputation than the control group would. Furthermore, we anticipated they would report an intention to speak more often to the team with an aggressive reputation than the control group. Finally, based on the notion of anticipatory bias and priming (e.g. Abernethy *et al.*, 1993; McLennan and Omodei, 1996; Paull and Glencross, 1997), we hypothesized that participants in the experimental group would respond faster than participants in the control group to aggressive play by the team with an aggressive reputation (i.e. incidents conforming to expectations for the particular team).

Methods

Participants

After ethical approval was obtained from the relevant university panel, contact was made with a regional English football association for permission to get in touch with their members. After receiving permission, 120 referees, qualified to Class 1 and registered with the association, received a standard letter informing them of the study. Class 1 ensured a common standard of professional expertise, with participants typically officiating in semi-professional games, although some had previous experience of officiating professional matches in the English Football League. In total, 43 participants replied, indicating that they were interested in participating. The participants were aged 40.5 ± 9.9 years and had 11.7 ± 7.4 years of experience (mean \pm s).

Five of the participants took part in a pilot study; the remaining 38 participants took part in the main study.

Apparatus and materials

The experimental scenarios comprised a series of incidents on video from Argentinean Football League games in which opposing players challenged for the ball. This league was chosen because we thought that the participants would have limited prior knowledge of the teams involved and, therefore, no pre-existing biases. A common factor in all of the incidents was the participation of one team who played in a predominantly blue strip with a yellow horizontal stripe (hereafter referred to as the 'blue team'). The incidents were taken from three matches involving the blue team against different opponents. Similar to Plessner and Betsch (2001), all of the video files were stopped before any cues indicating the decision made in the game (e.g. a view of the referee or reaction of the players after the incident).

The digital video files were edited in Adobe Premier software and the display was controlled by a Pentium III personal computer via an interface program written in National Instruments LabView software and projected in real-time onto a 1.42 × 1.88 m display. The interface program: (a) determined a random playing order for the clips; (b) ran the digital files; (c) responded to signals placed in the digital file at the onset of an incident by starting a timer; (d) recorded participants' responses from a hand-held push button to determine decision time; and (e) occluded the display on the participant's response to preclude additional cues becoming available after a decision was made.

Pilot study

In total, 108 suitable incidents were identified and digitized. A pilot study with five participants (age 38.0 ± 13.0 years, experience 13.4 ± 5.4 years) was conducted for two reasons. First, to determine if the task had face validity for the referees and enough information was provided in the video file to make a decision; and, second, to classify the incidents according to their severity. Altogether, five categories were chosen to reflect a range of possible incidents that may occur during a game:

1. Incidents for which all participants agreed that a foul had been committed by the blue team ('certain blue team').
2. Incidents for which there was disagreement between the participants as to whether a foul had been committed by the blue team ('uncertain blue team').
3. Incidents for which all participants agreed that a foul

had been committed against the blue team ('certain other team').

4. Incidents for which there was disagreement between the participants as to whether a foul had been committed against the blue team ('uncertain other team').
5. Incidents for which all participants agreed that no foul had been committed ('innocuous incidents').

A range of incidents of varying severity committed both by and against the blue team were used to maintain the ecological validity of the task and prevent the referees engaging in a response set. In addition, if prior knowledge did indeed have an impact on the decisions made, we wished to determine if this was primarily observed in incidents in which there was some doubt as to whether a foul had been committed or in incidents in which there was little doubt that a foul had been committed. This extends the work of Frank and Gilovich (1988), who only required participants to make a decision for two identical incidents, one in which an American football team in black was the defensive team and one in which an American football team wearing white was the defensive team.

Procedure of pilot study

The participants were welcomed to the laboratory and informed that they were about to take part in a study investigating the decision-making processes of football referees. They were given a written outline of the methodology and asked to sign an informed consent form. The participants were then randomly presented with 108 incidents involving the blue team. Each incident was preceded by an on-screen countdown to the action (4 s long), followed by the play leading up to and showing the incident (5–10 s in length). There was a 7 s pause between the end of one incident and the beginning of the visual countdown for the next incident.

For each incident, the participants were required to indicate whether a foul had been committed. If they did not believe a foul had occurred, then they were to say 'play on'. If they believed a foul had been committed, they were instructed to press a hand-held key as soon as possible. This provided an indication of decision time and automatically stopped the video playback of the incident, after which the participants were asked to verbally inform the researcher which of the following actions they would engage in if refereeing the game: (a) issue a card (red or yellow), (b) speak to the player concerned or (c) do nothing. We were aware that during a game decisions made by referees are based on several factors, such as whether there is a possibility of playing advantage or whether they have spoken to the player involved about a previous incident. Accordingly, as the

referees did not have that information available to them during this task, we emphasized that they should judge each incident *on its own merit*. Following data collection, which took approximately 60 min, the participants were thanked and received £20 payment for taking part in the study. At the end of the study, all participants received a follow-up debrief letter in which the aim of the study was clearly explained and our preliminary findings were reported.

The referees in the pilot study participated individually. From the data collected, 50 incidents were extracted for use in the main study. These comprised 10 incidents in each of the five categories selected (see above).

Procedure of main study

In a balanced design, 38 participants (age 40.8 ± 9.6 years, experience 11.5 ± 7.7 years) were randomly assigned to one of two groups: experimental or control. The information given to the participants in the main study was identical to that provided in the pilot study, apart from the experimental group being primed (via both written and oral instructions) that the team involved in all the incidents (blue team) had a history of aggressive behaviour. We based our decision to provide both written and oral information in this explicit manner to replicate the way this information is often presented in a real-world setting. For example, a referee may read of a team's aggressive reputation in a newspaper article and also be informed of it in conversation with other officials, players or managers. The written instructions presented to the experimental group contained information that the team wearing the blue strip in all of the clips 'had a reputation for foul and aggressive play' and that this team had been chosen specifically for inclusion in this study on decision-making because 'they had a history of playing in a confrontational manner and have a deserved reputation for foul and aggressive play'. This was further reinforced by the researcher verbally stating immediately before the task: 'To recap, you are going to watch a series of video clips involving a team wearing blue who have a reputation for aggressive play'. The referees participated individually and the incidents were presented in the same manner as in the pilot study. To familiarize themselves with the task and apparatus, all participants were presented with five example incidents. After viewing these, each participant watched the same 50 incidents, presented at random, and responded by indicating whether a foul had been committed in a procedure identical to that outlined in the pilot study.

After completing the task, the participants were asked to indicate on a 5-point Likert scale how aggressive they believed the team in blue was compared with their

opponents, where 1 = 'substantially less aggressive', 3 = 'similarly aggressive' and 5 = 'substantially more aggressive'. In addition, they were asked whether they recognized any of the teams included in the incidents.

After data collection, which took approximately 45 min, the participants were thanked and received £15 payment for taking part in the study. At the end of the study, all participants received a follow-up debrief letter in which the aim of the study was clearly explained and our preliminary findings were reported.

Results

The 10 'innocuous' incidents were included to increase the ecological validity of the task and prevent the referees from entering a response set whereby they would be indicating a foul for most incidents. Therefore, only data collected from the 20 'certain' and 20 'uncertain' incidents were analysed. Analysis of the data supported the classification of the incidents by the five pilot study participants. During the main study, the participants indicated that a foul had occurred during 90% of the 'certain' incidents and during 45% of the 'uncertain' incidents. In addition, the participants indicated that a foul had been committed during 11% of the 'innocuous' incidents. No cards were awarded for any of the 'innocuous' incidents. Interestingly, three participants in the control group each penalized a member of the other teams for diving on one occasion in the 50 incidents they viewed, while two participants in the experimental group each penalized a member of the blue team on one occasion for diving in the 50 incidents they viewed. These decisions were not included in the analyses.

The data for the number (frequency) of decisions awarded against both teams, the number of red and yellow cards given, and the number of times the referee reported an intention to speak to the player, were not normally distributed, demonstrating both skewness and kurtosis. Accordingly, it was not possible to analyse the data using parametric techniques. Addressing this through deletion of univariate and multivariate outliers would have resulted in a substantial reduction in data points and, although transforming the data was considered, the varying distributions of the data sets meant there was no single transformation that could be applied to the data set as a whole. Thus, for clarity, and because we were dealing with frequency data, we considered chi-square analysis to be most appropriate.

Initially, three separate chi-square analyses were performed to determine whether the experimental and control groups differed in their behaviour towards the blue team, for both the certain and uncertain incidents combined, in terms of the number of decisions, the

number of red and yellow cards and the number of times the referee intended to speak to the players involved. Because multiple chi-square tests were conducted, a conservative significance criterion of $P = 0.01$ was adopted per hypothesis, so that the risk of Type I error was not too great (Cohen, 1992). Even with this, the total number of cases involved in the analysis meant that the tests for the number of decisions and number of red and yellow cards still had sufficient power (> 0.8) to detect a small to medium effect size (Cohen, 1988). The test for the number of occasions the referees reported an intention to speak to the player concerned had sufficient power (> 0.8) to detect a medium to large effect size (Cohen, 1988).

Next, the number of decisions awarded, the number of red and yellow cards given and the number of times the referee indicated an intention to speak to the players involved were analysed separately for the 'certain' and 'uncertain' incidents using six separate 2×2 (group \times incident) chi-square analyses. In line with the recommendations of Thomas and Nelson (1996), Yates' correction for continuity was included for each chi-square analysis. Again, because multiple tests were conducted, a conservative significance criterion of $P = 0.01$ was adopted for each hypothesis (Cohen, 1992). Even so, the total number of cases involved in the analysis meant that most of the tests had sufficient power (> 0.8) to detect at least a medium effect size (Cohen, 1988). The tests for the number of cards awarded and the number of times the referees reported an intention to speak to the players for the 'uncertain incidents' had sufficient power (> 0.8) to detect a large effect size (Cohen, 1988). The number of decisions awarded, the number of red and yellow cards given and the number of times the referee indicated an intention to speak to the players for the 'certain' and 'uncertain' incidents are shown in Table 1.

The decision time data (in milliseconds) for the time lag between the onset of the incident and the intervention of the participant to indicate that an offence had occurred were analysed using a 2 (group) \times 4 (incident) general linear model analysis of variance (ANOVA) with repeated measures on the second factor (incident). The group factor refers to experimental versus control group. The incident factor refers to the classification of the incidents into the four categories ('certain blue team', 'certain other team', 'uncertain blue team', 'uncertain other team').

Number of decisions

There was no significant difference [χ^2 (1, $n = 533$) = 3.47, $P > 0.01$] in the number of decisions awarded against the blue team by the experimental ($n = 288$) and control ($n = 245$) groups.

Table 1. Number of decisions awarded, red and yellow cards given and intention to speak to the players involved in the 'certain' and 'uncertain' incidents

	Experimental group	Control group
'Certain blue team'		
Decisions	185	165
Cards	111	76
Intention to speak	30	33
'Certain other team'		
Decisions	160	172
Cards	56	74
Intention to speak	42	29
'Uncertain blue team'		
Decisions	103	80
Cards	20	14
Intention to speak	9	12
'Uncertain other team'		
Decisions	77	85
Cards	8	6
Intention to speak	19	10

Two separate 2×2 (group \times incident) chi-square analyses were conducted to determine the impact of the manipulation on the frequency with which participants penalized the teams involved in both the 'certain' and 'uncertain' incidents. There was no significant difference [χ^2 (1, $n = 682$) = 1.30, $P > 0.01$] in the number of decisions made by the experimental and control groups for the 'certain' incidents; similarly, there was no significant difference [χ^2 (1, $n = 345$) = 2.30, $P > 0.01$] in the number of decisions made by the experimental and control groups for the 'uncertain' incidents.

Number of red and yellow cards

Few red cards were awarded and only in the 'certain' category (experimental group: 'certain blue', $n = 7$; 'certain other team', $n = 8$; control group: 'certain blue', $n = 1$; 'certain other team', $n = 14$). Consequently, the number of red cards was combined with the number of yellow cards awarded. Analysis of the number of cards awarded showed that there was a significant difference [χ^2 (1, $n = 221$) = 7.61, $P < 0.01$] in the number of cards awarded against the blue team by the experimental ($n = 131$) and control ($n = 90$) groups.

Two separate 2×2 (group \times incident) chi-square analyses were conducted to determine the impact of the manipulation on the frequency with which participants awarded red or yellow cards against the teams involved in both the 'certain' and 'uncertain' incidents. There was a significant difference in the number of cards awarded by the experimental group and the control

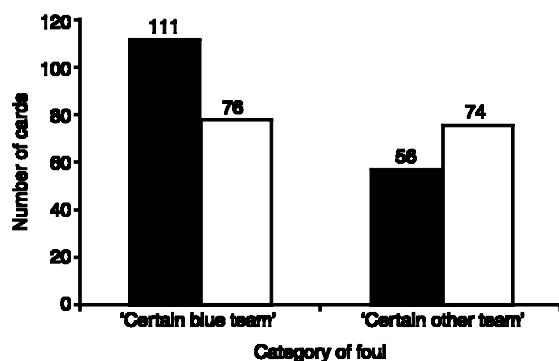


Fig. 1. The number of red and yellow cards awarded by the experimental (■) and control (□) groups after the 'certain' incidents.

group for the 'certain' incidents [$\chi^2(1, n = 317) = 7.52, P < 0.01$]. The numbers of red and yellow cards awarded by the experimental and control groups for the 'certain' incidents are displayed in Fig. 1. However, there was no significant difference [$\chi^2(1, n = 48) = 0.001, P > 0.01$] in the total number of cards awarded by the experimental and control groups for the 'uncertain' incidents.

Number of occasions the referees would have spoken to the player concerned

There was no significant difference [$\chi^2(1, n = 84) = 0.429, P > 0.01$] in the number of times the experimental ($n = 39$) and control ($n = 45$) groups reported an intention to speak to a player on the blue team after an incident.

Two separate 2×2 (group \times incident) chi-square analyses were conducted to determine the impact of the manipulation on the frequency with which participants spoke to the players involved in both the 'certain' and 'uncertain' incidents. There was no significant difference [$\chi^2(1, n = 134) = 1.35, P > 0.01$] in the number of times participants in the experimental and control groups reported an intention to speak to the players concerned for the 'certain' incidents. Similarly, there was no significant difference [$\chi^2(1, n = 50) = 1.70, P > 0.01$] in the number of times participants in the experimental and control groups reported an intention to speak to the players concerned for the 'uncertain' incidents.

Decision time

The data for the time lag between the onset of the incident and the intervention of the participant to indicate that an offence had occurred were analysed using a 2 (group) \times 4 (condition) general linear model ANOVA with repeated measures on the second factor (con-

Table 2. Decision times across the four categories of playing incident (mean \pm s)

	Experimental group	Control group
'Certain blue team'	1205 \pm 143	1036 \pm 179
'Certain other team'	1272 \pm 181	1027 \pm 231
'Uncertain blue team'	1508 \pm 50	1263 \pm 320
'Uncertain other team'	1563 \pm 338	1413 \pm 326

dition). Univariate outliers were removed using box-plots, resulting in a reduction in the number of cases to 11 participants in the experimental group and 18 in the control group. Also, because the Mauchly's test of sphericity was significant, the conservative Greenhouse-Geisser correction was applied to the reported test results. The mean decision time scores for these participants are shown in Table 2. A significant main effect was observed for group ($F_{1,27} = 7.64, P < 0.05, \eta^2 = 0.217$), with the experimental group (mean decision time 1376 \pm 270 ms) reporting higher average decision times than the control group (mean decision time 1203 \pm 321 ms) over the four types of incident. There was also a significant main effect for incident ($F_{3,81} = 26.4, P < 0.001, \eta^2 = 0.494$), indicating that decision-making time changed across the category of incident. Table 2 illustrates how these scores changed across categories, with an increase in decision time as ambiguity increased. That is, decision times were slower for the 'uncertain' than for the 'certain' incidents. The interaction between group and incident was not significant ($F_{3,81} = 0.52$), although power was obviously low in this test, with 18 participants in the control group and 11 participants in the experimental group.

Perception of aggressiveness

As a manipulation check, an independent t -test was applied to determine whether the participants in both groups perceived the blue team to be similarly aggressive to the other teams. There was no significant difference ($t_{36} = 0.580, P > 0.05$) in the perception of aggressiveness between the experimental (mean perception of aggressiveness 3.32 \pm 0.58) and control (mean perception of aggressiveness 3.21 \pm 0.54) group.

Awareness of teams involved

Some participants (experimental group, $n = 2$; control group, $n = 5$) indicated that they recognized at least one of the teams involved; however, they reported that they had limited knowledge and no strong feelings for any of the teams. Therefore, their data were included in the analyses.

Discussion

The results of this study suggest that football referees who are informed that a team has an aggressive reputation do respond differently to identical incidents than football referees who have not received this information. Specifically, football referees informed that a team has an aggressive reputation were more likely to award red and yellow cards against this team than the football referees who did not receive such information. However, they were unlikely to award more decisions against the team with an aggressive reputation or report an intention to speak to them more frequently after incidents. Neither did they respond more quickly to incidents conforming to that expected for the team (i.e. aggressive play).

The awarding by the experimental group of more red and yellow cards against the blue team overall, and following the 'certain' incidents, is interesting. The findings would support the notion that prior knowledge and expectation of aggressive play may result in officials judging players more harshly, in addition to other factors, such as the colour of players' uniforms (Frank and Gilovich, 1988). However, that prior knowledge about the aggressive nature of a team does not impact the number of decisions awarded against them contrasts with our hypothesis. This finding is consistent for incidents in which participants in the pilot study agreed that a foul had been committed (certain incidents), and also for incidents for which there was disagreement that a foul had been committed (uncertain incidents). It is perhaps not surprising that there was no difference in the number of decisions awarded for the 'certain incidents', given that for most referees these were clear-cut decisions. It is more surprising that there was no difference in the number of decisions awarded for the 'uncertain incidents', as it could be suggested that prior knowledge of a team's aggressive behaviour may be more likely to influence a decision when there is ambiguity as to whether a foul had been committed, because referees may use heuristic methods of reasoning and rely more heavily on prior knowledge (schemata). However, this was not the case and, in contrast to other factors, such as game location (Sumner and Mobley, 1981; Nevill *et al.*, 1996) and in-group favouritism (Mohr and Larsen, 1998), prior knowledge of a team's aggressive reputation does not appear to have an impact on the number of decisions awarded against a particular team.

Given that the experimental and control groups did not differ in the number of decisions awarded against the blue team, it is not surprising that the experimental group did not respond quicker to aggressive incidents by the blue team. Indeed, the participants in the experimental group were slower to make decisions in general,

and this was consistent across all incidents, regardless of whether they conformed to the expectation for the experimental group (i.e. aggressive play by the blue team). In short, the referees did not appear to be anticipating aggressive behaviour from the blue team in the incidents they viewed. One possible explanation of this finding is that the written and oral priming received may have interfered with the normal automatic (implicit) decision-making process, resulting in a more conscious (explicit) form of decision-making (Masters *et al.*, 2002). If a more explicit form of decision-making was used, it did not appear to affect the decision of whether or not a foul had been committed but rather what to do afterwards (i.e. award a red or yellow card).

It is interesting that prior knowledge appeared to affect the number of red and yellow cards, but not the overall number of decisions awarded. One possible explanation for this may be that deciding whether or not a foul has occurred requires recognition of the characteristics of the incident itself, and this decision is made not on information about the teams involved, but on long established schemata acquired over many years of on-field experience and knowledge of the rules. Thus, while recognition of a foul may be based on long established schemata, the decision of how to respond *after* an incident may be affected by a recently acquired, or updated, schema concerning a team's aggressive reputation. Indeed, Klein (1993) suggested that the process of decision-making involves recognizing the incident and responding on the basis of prior knowledge and experience. In terms of Klein's recognition-primed decision model, a referee, in deciding to award a foul, is only required to recognize the incident as a foul (relevant cues). However, in deciding on any further action, a referee may consider all available information to arrive at the most appropriate decision and, accordingly, may be influenced by prior knowledge. For example, once a referee has recognized a crude challenge (relevant cues), the decision to award a red or yellow card may be influenced by expectancies the referee may have regarding the outcome of his or her course of action. In this case, a referee may believe that penalizing players from teams with an aggressive reputation sends a clear signal of intent that aggressive behaviour will not be tolerated, thus preventing an escalation of aggressive behaviour. However, while this strategy may be used in a match, it may not necessarily explain the results of the present study. Although we asked the referees to judge each incident on its own merit and, therefore, respond as they would in a game, there was no real match and the referees could not influence the behaviour of the players in the video clips.

Before discussing any possible implications of our results, it is important to consider both how the nature of the task and any methodological limitations may have

contributed to the findings. First, it is possible that demand characteristics could have led the referees in the experimental group to award more red and yellow cards against the blue team. That is, from the referees' point of view, the explicit provision of information regarding the aggressive reputation of the blue team may have appeared as an instruction to use this information in their decisions. To counter this, we made it clear to the participants that they should judge each incident on its merit. In addition, it is worth noting that we only used qualified and experienced referees who, it is presumed, were used to judging incidents in a fair and impartial manner. Furthermore, the manipulation check indicated that the experimental and control groups were similar in their perception of the blue team being aggressive. That is, the additional oral and written information provided to the experimental group did not influence post-experimental perceptions of the level of aggressiveness of the blue team. Finally, if referees in the experimental group were using this information to impact their decisions, because that is what they believed the task required them to do, then it may reasonably be expected that they would also have awarded more decisions against the blue team. However, this was not the case.

A second possible limitation is that the laboratory setting reduces the external validity of the study, limiting the ability to generalize these results to real-world settings. That is, there is no guarantee that in a match referees will award more red and yellow cards against a team with an aggressive reputation. For example, when officiating teams with an aggressive reputation, referees may adopt certain strategies, such as talking to the players concerned before and during the game, before any aggressive incidents occur to encourage players to maintain their discipline. In addition, teams with an aggressive reputation when playing at home may not receive more red or yellow cards because of pressure on the official from a supportive home crowd (e.g. Nevill *et al.*, 1996; Nevill and Holder, 1999). In the present study, the ability to exert control over the nature of the incidents viewed, to provide balance both in terms of the certainty of an incident and whether it was perpetrated by or against the blue team, was a central feature of the design, which would have been impossible to achieve in a 'live' game.

A final limitation of this study is that the power was only sufficient to detect a large or medium to large effect size for some of the analyses conducted. Thus, small effects of the manipulation may not have been detected for the number of occasions the referees reported an intention to speak to the player concerned over all incidents, for the 'uncertain' incidents, for the number of cards awarded for the 'uncertain' incidents, and for an

interaction between group and incident for decision time.

Conclusion

Although it is necessary to be cognizant of possible limitations and, therefore, cautious in drawing strong conclusions from the study, the implications of the findings are interesting. It would appear that in a laboratory setting referees may treat a team with an aggressive reputation differently to teams without such a reputation. Furthermore, although our results are sport-specific, the findings have applicability to a range of contexts in which individuals are often required to make a quick and accurate response. In addition, there is the potential for further research to address a number of pertinent issues arising from the present study. For example, future research may wish to examine whether information of a team's aggressive reputation communicated in a more subtle way or, indeed, acquired by the referees' own observations (i.e. scenarios with less demand characteristics), have a similar effect on behaviour. Similarly, research may wish to examine whether, in real games, referees do indeed award more red and yellow cards against a team with an aggressive reputation. A qualitative methodology may be particularly useful here in helping to explore not only whether referees do treat teams or players with aggressive reputations differently, but why this may be so (e.g. trying to prevent on-field aggressive behaviour escalating). Other potential areas of research include determining whether knowledge of a team's aggressive reputation influences officials' behaviour in other sports (e.g. ice hockey, field hockey, rugby union, American football). Finally, as the data collected here were from skilled but not elite referees, future research may also wish to investigate whether the standard of sport officials moderates the impact of prior knowledge on decisions made.

Acknowledgements

This study was supported by a grant from the British Academy (SG-31074). The authors wish to acknowledge the work of Sam Dunkey and Kim Bird in organizing the data collection, the technical assistance provided by Sharon Gwinneth and, finally, the referees and officers of the Staffordshire Football Association for their cooperation.

References

- Abernethy, B. (1987). Anticipation in sport: a review. *Physical Education Review*, 10, 5–16.

- Abernethy, B., Thomas, K.T. and Thomas, J.R. (1993). Strategies for improving understanding of motor expertise (or mistakes we have made and things we have learned!!). In *Cognitive Issues in Motor Expertise* (edited by J.L. Starkes and F. Allard), pp. 317–356. Amsterdam: North-Holland.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155–159.
- Courneya, K.S. and Carron, A.V. (1992). The home advantage in sport competitions: a literature review. *Journal of Sport and Exercise Psychology*, **14**, 13–27.
- Fiske, S.T. and Taylor, S.E. (1991). *Social Cognition*. London: McGraw-Hill.
- Frank, M.G. and Gilovich, T. (1988). The dark side of self- and social perception: black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology*, **54**, 74–85.
- Hogg, M.A. and Vaughan, G.M. (1995). *Social Psychology: An Introduction*. Hemel Hempstead: Prentice-Hall/Harvester Wheatsheaf.
- Kahneman, D. and Tversky, A. (1984). Choices, values and frames. *American Psychologist*, **39**, 341–350.
- Klein, G.A. (1989). Do decision biases explain too much? *Human Factors Society Bulletin*, **32**(5), 1–3.
- Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In *Decision-making in Action: Models and Methods* (edited by G.A. Klein, J. Orasanu, R. Calderwood and C.E. Zsombok), pp. 138–147. Norwood, NJ: Ablex.
- Masters, R., Law, J. and Maxwell, J. (2002). Implicit and explicit learning in interceptive actions. In *Interceptive Actions in Sport: Information and Movement* (edited by K. Davids, G. Savelsbergh, S.J. Bennett and J. Van der Kamp), pp. 126–143. London: Routledge.
- McLennan, J. and Omodei, M.M. (1996). The role of pre-priming in recognition-primed decision making. *Perceptual and Motor Skills*, **82**, 1059–1069.
- Mohr, P.B. and Larsen, K. (1998). Ingroup favoritism in umpiring decisions in Australian Football. *Journal of Social Psychology*, **138**, 495–504.
- Nevill, A.M. and Holder, R.L. (1999). Home advantage in sport: an overview of studies on the advantage of playing at home. *Sports Medicine*, **28**, 221–236.
- Nevill, A.M., Newell, S.M. and Gale, S. (1996). Factors associated with the home advantage in English and Scottish soccer matches. *Journal of Sports Sciences*, **14**, 181–186.
- Orasanu, J. and Connolly, T. (1993). The reinvention of decision making. In *Decision-making in Action: Models and Methods* (edited by G.A. Klein, J. Orasanu, R. Calderwood and C.E. Zsombok), pp. 3–20. Norwood, NJ: Ablex.
- Paull, G.C. and Glencross, D.J. (1997). Expert perception and decision-making in baseball. *International Journal of Sport Psychology*, **28**, 35–56.
- Plessner, H. (1999). Expectation biases in gymnastic judging. *Journal of Sport and Exercise Psychology*, **21**, 131–144.
- Plessner, H. and Betsch, T. (2001). Sequential effects in important referee decisions: the case of penalties in soccer. *Journal of Sport and Exercise Psychology*, **23**, 254–259.
- Sumner, J. and Mobley, M. (1981). Are cricket umpires biased? *New Scientist*, July, pp. 67–68.
- Taylor, D. (2002). Angel with dirty face evokes the Revie era. *The Guardian Sport*, 12 January, p. 6.
- Thomas, J.R. and Nelson, J.K. (1996). *Research Methods in Physical Activity*. Champaign, IL: Human Kinetics.
- Tversky, A. and Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science*, **185**, 1124–1131.

Copyright of Journal of Sports Sciences is the property of E & FN Spon Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.