Referees’ use of heuristics: The moderating impact of standard of competition

NICOLAS SOUCHON1, GENEVIEVE CABAGNO2, ALAN TRACLET3, DAVID TROUILLOUD4, & GREGORY MAIO1

1School of Psychology, Cardiff University, Cardiff, UK, 2STAPS, Université de Rennes, Rennes, France, 3ISSEP, Université de Lausanne, Lausanne, Switzerland, and 4STAPS, Université de Grenoble, Grenoble, France

(Accepted 7 March 2009)

Abstract
The influence of players’ standard of competition on referees’ decision making was investigated. Standard of competition was examined by studying 30 handball matches, 15 at League and 15 at National standard. At both standards, we examined two types of refereeing decision (sporting and disciplinary) after transgressions that varied in ambiguity. The results indicated that referees made more lenient decisions at higher standards of competition, but this tendency depended on the type of decision being made and ambiguity in the offence. The hypothesis that referees use standard of competition as a judgemental heuristic, but in varying ways, was supported.

Keywords: Competition standard, judgemental heuristics, refereeing, sport performance, transgressions

Introduction
Psychologists have exhibited a growing interest in sport refereeing (Plessner, 2005; Ste-Marie, 2003). Although members of the public frequently treat fouls and misconducts as judgements of fact (Mascarenhas, O’Hare, & Plessner, 2006), several studies have shown that referees use different judgemental cues to help them make their decisions (Plessner & Haar, 2006). In the present research, we propose that standard of competition is one important cue used to make refereeing decisions and describe a field study testing this hypothesis.

Understanding the context: Officiating in team sports
In team contact sports (e.g. handball, ice hockey, soccer), referees must quickly make two types of decision after observing an action that might be a foul: a sporting decision and a disciplinary decision. An example of a sporting decision is the “advantage rule” in many team contact sports. This rule indicates that referees need not intervene when the player in possession of the ball retains the chance to score or pursue his or her action, despite being the victim of one or more defensive fouls. If the player misses a pass or shot after being fouled, the referee can give the ball back to the fouled player’s team (Souchon, Coulomb-Cabagno, Traclet, & Rascle, 2004). The advantage rule helps to ensure that play is not disrupted in a way that is advantageous to the aggressors (Widmeyer, 2002). In contrast, disciplinary decisions involve punishing the aggressor more strongly for dangerous offences. Depending on the type of sport, disciplinary punishment can range from the threat of suspension (e.g. yellow card in football) to temporary suspension or final suspension (e.g. red card in football). The goal of disciplinary punishment is to discourage players from using dangerous behaviour against their opponents and from intentionally attempting to injure them (Mascarenhas et al., 2006).

Referees have to make both types of decision under duress from the actions of players, coaches, spectators, and the media (Kaiissidis-Rodafinos, Anshel, & Sideridis, 1998). Decisions can also be made spontaneously, under time pressure (e.g. Plessner & Betsch, 2001) and with varied, suboptimal viewing positions (e.g. Plessner, 2005). Moreover, most players’ fouls are ambiguous. That is, they can be interpreted in different ways (Mascarenhas et al., 2006; Plessner, 2005; Ste-Marie, 2003). Thus, refereeing is a complex decision-making activity that is undertaken in circumstances that preclude...
extensive cognitive information processing (Plessner & Haar, 2006).

Referees can deal with these difficulties by relying on judgemental heuristics (Nevill, Balmer, & Williams, 2002; Plessner & Betsch, 2001). Heuristics are simple ways of reasoning to help guide judgements of uncertain events in complex environments (Tversky & Kahneman, 1974), and they might or might not be valid. For instance, soccer referees make more sporting decisions in favour of the home team because of noise from the home fans (Nevill et al., 2001). In addition, referees are influenced by their stereotypes (Souchon et al., 2004; see also Frank & Gilovich, 1988; Jones, Paull, & Erskine, 2002). Stereotypes, which are a kind of judgemental heuristic, can be defined as the sum of beliefs, knowledge, and expectations individuals develop towards members of social categories (Hamilton & Sherman, 1994). Research has shown that referees use stereotypes linked to the sex of players to help them make decisions (Souchon et al., 2004). Also, referees develop expectations about individual players’ aggressiveness and are influenced by players’ reputations for aggression (Frank & Gilovich, 1988; Jones et al., 2002).

We propose that standard of competition is another source for stereotyping players. Players of a lower standard are less skilled than their higher-standard counterparts and, as a result, referees might consider that the latter can continue play despite being the victim of a defensive foul (Souchon et al., 2004). As a result, referees interpret players’ fouls differently according to their preconceptions from the standard of play. For example, a referee officiating at a low standard of competition might decide that a specific occurrence of defensive contact against a player in possession of the ball is a foul because the referee believes that the contact caused the attacking player to lose his or her balance. Conversely, the same contact at a high standard of competition could be perceived as fair, because referees believe that the player at a higher standard could have recovered his or her balance. Referees would intervene more frequently during competition of low standard, but would apply the advantage rule more frequently during competition at higher standard. Accordingly, referees would penalize fewer sporting infringements that occurred in high-standard competition.

In addition, high-standard players are more likely to believe that aggression is legitimate and hence use aggressive behaviours as a tactical tool (Conroy, Silva, Newcomer, Walker, & Johnson, 2001). Consequently, they commit more fouls than less-skilled players (Coulomb-Cabagno & Rascle, 2006). Referees could therefore develop stereotypes of high-standard players being more motivated and aggressive than their lower-standard peers. This difference is important because it has been demonstrated that players with a reputation for aggressive behaviour are more severely punished for breaches of discipline than other players (Jones et al., 2002). There is a need for greater deterrence against more aggressive players, whose actions are more likely to cause severe injury. Thus, referees may be more severe with high-standard players so as to discourage their aggression and reduce the chance of serious danger to their opponents.

These effects of standard of competition might also depend on the nature of the transgressions, because different transgressions can have different effects on the victim. For example, players in possession of the ball could more easily continue their action after a “pushing” offence than after a “holding back” offence in handball (Souchon et al., 2004) and in soccer (Coulomb-Cabagno, Rascle, & Souchon, 2005). Accordingly, “pushing” offences can be considered to be ambiguous. Because prior stereotyping influences referees’ responses to these situations where several interpretations are possible (Plessner & Haar, 2006; Schneider, 2004), referees could interpret ambiguous fouls, such as “pushing”, differently from clearer ones such as “holding back”. Standard of competition could cue different decisions more for ambiguous offences than for the less ambiguous ones.

To test these predictions, we examined the role of competitive standard in handball players at local and national tournaments. Handball is a team contact sport where two teams of seven players have to throw or strike a ball into the opposite goal and defend their own goal against opponent attacks. To score a goal, attacking players have to throw the ball past the goalkeeper into the opponent's goal, without entering the goalkeeper’s area (6 m away from the goal). The simple scoring, contact during play, high skill and fitness, and team coordination make handball a prototypical team contact sport in many ways. These elements also lend it to an investigation of sporting and disciplinary decisions.

By examining actual play and decisions at local and national tournaments, we achieved construct validity in our design. Simultaneously, we maximized measurement accuracy and internal validity by unobtrusively recording play and decisions, thereby avoiding any intrusion into matches. Using our recordings, we were able to assess transgressions objectively, without the distractions and temporal pressures imposed on officials, and then compare officials' decisions with observed transgressions. Thus, a strong feature of this research was that we were able to compare actual transgressions with refereeing decisions in an ecologically sound context without intrusion by the researchers.
Methods

Participants

The study examined 30 matches in France’s National Handball Championships for men. Fifteen matches at the highest league standard (i.e. pre-national) and 15 matches at the highest national standard (i.e. first division) were videotaped. Referees at national and league standard comprise two distinct groups. Matches were selected with as many referees (12 at national vs. 14 at league standard) and teams (14 at national vs. 18 at league standard) as possible. The research was conducted in accordance with the ethical guidelines of the American Psychological Association and the British Psychological Society, and was approved by the Ethics Committee of Rennes University.

Coding

Referees can interrupt a game and restart it with a free-throw for the attacking players when a defensive player commits a violation of the rules. Nevertheless, according to the “advantage rule” in handball, referees should not intervene until and unless it is clear that the attacking team has lost possession of the ball or cannot continue its attack (rule 13.2 of the International Handball Federation, 2005). Consequently, handball referees consider that the advantage rule applies mainly to actions involving the player in possession of the ball (Souchon et al., 2004). This led us to concentrate our observations on transgressions committed on players in possession of the ball and, in particular, fouls by defenders against players with possession during an organized attack. An organized attack begins at the moment a player in possession of the ball is located behind a line of at least one set of four opposing defenders. Referees consider these situations to be the biggest part of the handball game (Souchon et al., 2004). Moreover, during an organized attack, several defensive players simultaneously or consecutively defend against the player in possession of the ball. Referees then have numerous opportunities to apply the advantage rule if the defenders foul the attacking player. In contrast, counter-attacks are mainly one-on-one situations; hence referees have fewer opportunities to apply the advantage rule.

Transgressions. The observation criteria were strictly in accordance with the rules of handball. Any Type 1 (“pushing”, “pushing away”, “bumping into”) and Type 2 (“holding back”, “catching and holding”, “seizing the player with possession around the waist”) defensive actions are considered to be transgressions (rule 8.2 of the International Handball Federation, 2005). We recorded both types of transgression because, as noted in the introduction, Type 1 pushing offences can be less clear and unambiguous than Type 2 pulling offences. In other words, Type 1 offences can be equivocal and more likely to yield different decisions (Souchon et al., 2004; some evidence presented later supports this assumption).

Before undertaking our final observations, two games (one at League standard and one at National standard) were observed by the first author and two handball experts. The rates of agreement between the observers (kappa coefficients) were high, between 0.85 and 0.94. Two further games were observed two weeks later, with agreement coefficients between 0.89 and 0.95. These games were then observed by two people only: the lead author and one expert (agreement coefficients between 0.92 and 0.96). Importantly, observations by the expert were made without him being aware of the research hypothesis. The mean of both of the observations was calculated to provide one measurement.

Refereeing decisions. For each defensive transgression that we coded, we noted the relevant sporting decision. In handball, referees can award either a 9-m or 7-m throw against the defending team. We also noted disciplinary decisions. Referees can choose between giving a yellow card (a warning), a 2-min suspension or a red card (expulsion from the game).

Statistical analysis

Because the response variable was categorical (9-m throw vs. 7-m throw vs. no intervention and/or yellow card vs. 2-min suspension vs. red card vs. no intervention), the assumptions necessary to conduct hypothesis tests using analysis of variance (ANOVA) are likely to be violated. A more appropriate technique to analyse categorical data is logistic regression. This analysis estimated the probabilities (or more correctly the odds) associated with the three (sporting) or four (disciplinary) categorical options and how these probabilities will vary due to differences in the predictor/independent variables. Note that to describe the strength of an effect (equivalent to an effect size in traditional ANOVA), logistic regression calculates the odds ratio – that is, the ratio of the odds of, for example, giving the ball back to the fouled player in a regional game, compared with (divided by) the odds of giving the ball back to the fouled player in a national game. Binary logistic regression was used to assess the effect of the independent variables, competition standard and type of transgression, on each outcome variable/option separately. For example, when analysing the 9-m throw option, binary logistic regression estimates the probability of awarding a 9-m
throw \((p)\) versus not awarding a 9-m throw \((1 − p)\) and how this probability will vary because of the differences in, or the effects of, the independent variables. Differences in decision according to competition standard and type of transgression were investigated separately for each option (9-m throw, 7-m throw, no sporting intervention and yellow card, 2 min suspension, and no disciplinary intervention) using binary logistic regression. The statistical software Statistica 6.00 was used to analyse the binomial responses variables. The method of model simplification adopted was backward elimination, in which the least important variable was dropped from the current model at each step.

**Results**

**Sporting decisions**

For 9-m throw, binary logistic regression identified a main effect for type of transgression \((\chi^2(1) = 406.45; \text{odds ratio, } \exp(\beta) = 0.68, P < 0.0001)\), competition standard \((\chi^2(1) = 17.1; \text{odds ratio, } \exp(\beta) = 0.13, P < 0.0001)\), and an interaction between competition standard and type of transgression \((\chi^2(1) = 3.93; \text{odds ratio, } \exp(\beta) = 0.07, P < 0.05)\). The main effect for type of transgression indicated that referees, collapsed across standard of competition, intervened more frequently with a 9-m throw after a Type 2 transgression (55%) than after a Type 1 transgression (31%). The main effect for competition standard indicated that referees, collapsed across type of transgression, intervened more frequently with a 9-m throw at a regional standard than at a national standard (see Table I). To interpret the interaction between competition standard and type of transgression, the odds ratio that related competition standard and 9-m throw was calculated first for Type 1 transgression and then for Type 2 transgression, using two binary logistic regressions. For Type 1 transgression, binary logistic regression identified a main effect for competition standard \((\chi^2(1) = 25.63; \text{odds ratio, } \exp(\beta) = 0.20, P < 0.0001)\). This effect indicated that referees intervened more frequently with a 9-m throw at regional than at national standard after Type 1 transgressions (see Table I). Nevertheless, binary logistic regression did not identify a main effect for competition standard with Type 2 transgressions \((\chi^2(1) = 1.8; \text{odds ratio, } \exp(\beta) = 0.07, P = 0.17)\). This indicated that referees tended to punish as severely players at regional standard as those at national standard with a 9-m throw after Type 2 transgressions (see Table I).

For 7-m throw, binary logistic regression identified a main effect for type of transgression \((\chi^2(1) = 82.55; \text{odds ratio, } \exp(\beta) = 0.70, P < 0.0001)\). This effect indicated that referees, collapsed across standard of competition, were more likely to intervene with a 7-m throw after a Type 2 transgression (8%) than after a Type 1 transgression (2%).

For no intervention, binary logistic regression identified a main effect for type of transgression \((\chi^2(1) = 541.04, \text{odds ratio, } \exp(\beta) = 0.79, P < 0.0001)\) and a main effect for competition standard \((\chi^2(1) = 17.32; \text{odds ratio, } \exp(\beta) = 0.14, P < 0.001)\). The main effect for type of transgression indicated that referees, collapsed across standard of competition, tended to let the game continue without intervention more frequently after Type 1 (74%) than after Type 2 transgressions (37%). The main effect for competition standard revealed that referees, collapsed across type of transgression, were more likely to let the game continue without intervention at national than at regional standard (see Table I).

**Disciplinary decisions**

For yellow cards, binary logistic regression identified a main effect for type of transgression \((\chi^2(1) = 20.39; \text{odds ratio, } \exp(\beta) = 0.43, P < 0.0001)\). This effect indicated that referees, collapsed across standard of competition, punished aggressive players more severely with a yellow card after Type 2 (5%) than after Type 1 transgressions (2%).

For 2-min suspensions, binary logistic regression identified a main effect for type of transgression \((\chi^2(1) = 48.5; \text{odds ratio, } \exp(\beta) = 0.65, P < 0.0001)\), competition standard \((\chi^2(1) = 4.6; \text{odds ratio, } \exp(\beta) = 1.66, P < 0.03)\), and an interaction between competition standard and type of transgression \((\chi^2(1) = 3.9; \text{odds ratio, } \exp(\beta) = 0.47)\)
competition standard. The main effect for type of transgression revealed that referees, collapsed across standard of competition, punished aggressive players more severely with a 2-min suspension after Type 2 (6%) than after Type 1 transgressions (2%). The main effect for competition standard indicated that referees, collapsed across type of transgression, punished aggressive players less severely at national than at regional standard (see Table II). To interpret the interaction between competition standard and type of transgression, the odds ratio that related competition standard and 2-min suspension was calculated first for Type 1 transgression and then for Type 2 transgression, using two binary logistic regressions. For Type 1 transgression, binary logistic regression did not identify a main effect for competition standard ($\chi^2(1) = 0.62$; odds ratio, $\text{exp}(\beta) = 1.25$, $P = 0.42$). This indicated that referees tended to punish players at regional standard as severely as those at national standard with a 2-min suspension after Type 1 transgressions (see Table II). However, binary logistic regression identified a main effect for competition standard ($\chi^2(1) = 4.62$; odds ratio, $\text{exp}(\beta) = 0.6, P < 0.03$) for Type 2 transgression. This revealed that referees were less likely to punish aggressive players with a 2-min suspension at national than at regional standard after a Type 2 transgression (see Table II).

For no intervention, binary logistic regression identified a main effect for type of transgression ($\chi^2(1) = 74.79$; odds ratio, $\text{exp}(\beta) = 0.58$, $P < 0.0001$). This effect indicated that referees, collapsed across standard of competition, tended to let the game continue without punishing players more frequently after Type 1 transgressions (97%) than after Type 2 transgressions (90%).

**Discussion**

The aim of this research was to improve understanding of the diverse determinants of refereeing decisions. Past research has shown that several contextual variables influence these decisions, but no studies have examined effects of standard of competition (Mascarenhas et al., 2006; Plessner & Haar, 2006). We anticipated that the effect of competitive standards would be greater for more difficult, ambiguous transgressions, because heuristic cues in general are more useful for ambiguous decisions (Chaiken, Liberman, & Eagly, 1989; Tversky & Kahneman, 1974). For these decisions, we anticipated that referees’ decisions would be more lenient in high-standard competition so as to maintain continuity of play, but more severe with disciplinary infringements. Our results revealed that higher-standard competition was associated with different sporting and disciplinary decisions.

With regard to sporting decisions, the effect of competitive standard on referees’ intervention depends on the severity of sanction. Referees tended to award a 9-m throw less frequently at the higher standard of competition. Also, they decided not to intervene more frequently at the higher standard of competition, but did not differ in decisions to award the more severe 7-m throw. In essence, referees avoided giving the relatively minor (9-m) sporting intervention at the higher standard of competition. This finding is consistent with the notion that a high standard of competition involves greater flow, thus effective refereeing is characterized by fewer interventions (Mascarenhas et al., 2006). This finding is consistent with reports that referees’ decision making differs according to factors such as the sex of player (Souchon et al., 2004) and location of the game (Nevill et al., 2002).

This result might have occurred because the need to maintain flow in matches is important to referees’ sporting decisions (Mascarenhas et al., 2006). Referees might wait longer before intervening at a higher standard of competition because they surmise stereotypically that players can continue their actions, despite the seriousness of the fouls (Coulomb-Cabagno et al., 2005). In addition, speed of play at a high standard of competition could make a higher proportion of fouls more ambiguous, and greater player aggressiveness could make it more difficult to note each transgression and cause referees to focus

<table>
<thead>
<tr>
<th></th>
<th>Regional standard</th>
<th>National standard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type 1</td>
<td>Type 2</td>
</tr>
<tr>
<td>Yellow card</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-min suspension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red card</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1419 (100%)</td>
<td>676 (100%)</td>
</tr>
</tbody>
</table>
on the more serious ones (Mascarenhas et al., 2006; Widmeyer, 2002). Thus, there is a variety of factors that might cause fewer sporting penalties to occur at a higher standard of competition, over and above whether the action involved more ambiguous “pushing” or less ambiguous “pulling”. Nevertheless, the results also indicated that for ambiguous transgressions, referees awarded a 9-m throw less frequently but not for the obvious ones. This result is consistent with data showing that judgemental heuristics influence only judgements based on ambiguous information (Chaiken et al., 1989).

For disciplinary decisions, our results indicate that the effect of competitive standard depends both on the type of transgression and the severity of sanction. That is, the standard of competition interacted with the ambiguity of the offence to predict the disciplinary decision. Surprisingly, the more ambiguous fouls were punished as severely at a high standard of competition as at the low standard, while the more obvious fouls were penalized differently according to the standard of competition. After the obvious fouls, referees opted for fewer suspensions at a high standard of competition than at a lower standard. It is not yet clear why officials were more lenient following these offences at a higher standard of competition. One potential explanation is that the aggressive behaviour elicited less punishment because players are perceived to be able to avoid contact. Our evidence provides a provocative indication that future research needs to address these issues by also examining whether referees’ perceptions of fairness change in contexts where the base rate for the behaviour is higher.

In summary, our results indicate that referees tend to make different decisions in handball according to the standard of competition. One possible explanation is that they could develop and use different stereotypes related to the players’ standard of competition. Future research should use experimental and longitudinal designs to assess if the standard of competition influences referees’ decisions across a wider range of transgressions and sports. For example, results obtained in this study could also be affected by differences in referees’ expertise across different standards of competition (Ste-Marie, 2003). Another challenge is to determine whether this evidence can be replicated in other sports, or whether this technique and sport produce different outcomes than in other more commercially compelling sports such as soccer, ice hockey, and basketball. We hope that this investigation helps to stimulate further research on these issues, assist training of officials, and inform the public about the cognitive challenges that referees face.

References


Schneider, D. J. (2004). The psychology of stereotyping. New York: Guilford Press.


