REFEREES’ DECISION MAKING ABOUT TRANSGRESSIONS: THE INFLUENCE OF PLAYER GENDER AT THE HIGHEST NATIONAL LEVEL

Nicolas Souchon
Cardiff University and Université de Paris X

Geneviève Cabagno and Olivier Rascle
Université de Rennes

Alan Traclet and Fabrice Dosseville
Université de Caen

Gregory R. Maio
Cardiff University

Prior research has found that referees are harsher toward sporting offenses in regional-level matches between women than in regional-level matches between men. We tested whether this bias also occurs at a higher, national level of competition, despite the greater pressures for objectivity and fairness at this level. Referees’ decisions were examined in 15 national-level handball matches between women and 15 national-level handball matches between men after transgressions that varied in severity. The results suggest that referees made harsher decisions in female than in male matches. Although more research is needed, this study supported the hypothesis that referees may use the gender of players as a powerful judgmental heuristic for deciding how to respond to aggression.

The study of the psychology of refereeing in team contact sport has been growing in recent years (e.g., Mascarenhas, O’Hare, & Plessner, 2006; Plessner & Haar, 2006). Referees must pay particular attention to the sporting behavior of players of both teams to guarantee fairness and to protect the participants. Referees act as socializing agents who fix authorized limits and apply the rules of the games, occupying a central role in regulating and controlling players’ aggressive behavior (Widmeyer, 2002). Several studies have shown that referees use different judgmental cues to help them to make their decisions (e.g., Jones, Paull, & Erskine, 2002; Nevill, Balmer, & Williams, 2001; Plessner & Betsch, 2002). In particular, studies have found that referees tend to penalize female players’ aggressive behaviors more severely than male players’ aggressive behaviors (Coulomb-Cabagno, Rascle, & Souchon, 2005; Souchon, Coulomb-Cabagno, Traclet, & Rascle, 2004). Nevertheless, these studies examined lower levels of competition, and there are important reasons to test whether the same bias occurs at higher levels of competition. Although it might seem reasonable to predict that a greater level of refereeing expertise at the highest national competition level should be associated with greater decision accuracy (i.e., a reduced gender effect), research has shown that the influence of gender is highly pervasive (Eagly, 2007). Also, any referee bias toward female players may affect female players’ performances, perception of their own abilities, and individuals’ stereotype toward female players. This article describes reasons for examining this bias at a higher level of competition and then describes an experiment that tested whether the gender bias does occur at a higher level of competition.

Understanding the Context: Officiating in Team Sports

In team contact sport (e.g., handball, ice hockey, soccer, rugby), illegal behaviors are usually defined according to rigorous criteria. For example, pushing is a transgression in basketball and soccer. After seeing a push, referees must quickly decide whether it is a foul and make two types of decisions: a sporting decision and a disciplinary decision. An example of a sporting decision is the “advantage rule.” This rule indicates that referees should not intervene when the player in possession of the ball retains the chance to score or pursue his or her action although 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 victim (Souchon et al., 2004). The advantage rule illustrates how a sporting decision is aimed to ensure that play is not disrupted in a way that is advantageous to the aggressors (Widmeyer, 2002).

Disciplinary decisions involve punishing the aggressor more strongly for dangerous offenses. The goal of disciplinary punishment is to discourage players from using dangerous behavior against their opponents (Mascarenhas et al., 2006). Depending on the type of sport, disciplinary punishment can range from the threat of suspension (e.g., yellow card in soccer) to temporary suspension or final suspension (e.g., red card in football). Disciplinary punishment can be given immediately after an infraction if this does not give a sporting disadvantage to the victim. Otherwise, the disciplinary punishment can be delayed until the potential sporting disadvantage no longer exists.

Although a certain degree of objectivity is expected in refereeing (Mascarenhas et al., 2006), a whole range of stress-generating factors make objectivity difficult (e.g., Kaisissidis-Rodafinos, Anshel, & Sideridis, 1998). In addition, time pressure makes it impossible to thoroughly contemplate and revise a decision (Plessner, 2005). Also, a large number of players in motion forces referees to continually change position to have the best possible view of transgressions (Ste-Marie, 2003).

These pressures may force referees to use different judgmental heuristics to arrive at decisions. Heuristics can be defined as simple ways of reasoning to help guide judgment about uncertain events in complex environments (Tversky & Kahneman, 1974). For example, football referees are less likely to penalize players who have vocal crowd support, even when the players have clearly committed a foul (Nevill et al., 2001). In addition, several studies have shown that referees are influenced by their stereotypes (Frank & Gilovich, 1988; Jones et al., 2002). Stereotypes, which are a kind of judgmental heuristic, can be defined as the sum of beliefs, knowledge, and expectations individuals develop toward the members of social categories (Hamilton & Sherman, 1994).

Gender stereotypes are among the strongest stereotypes. They can be used automatically or completely unconsciously (Greenwald & Banaji, 1995) because gender is a fundamental criterion of social categorization and stereotype development (Fiske, 1998). Indeed, individuals find it difficult to identify the true content of their gender stereotypes and the influence of these stereotypes on their behavior, perceptions, and decisions (Schneider, 2004). Notwithstanding this introspective difficulty, research has shown that people implicitly expect and prefer that women be communal, manifesting traits such as kindness, concern for others, warmth, and gentleness, and that men be agentic, manifesting traits such as confidence, aggressiveness, and self-direction (Eagly, 2007; Deaux & Lafrance, 1998).

Individuals who violate these gender prescriptions experience negative reactions and disapproval from others (Burgess & Borgida, 1999; Heilman, 2001). For example, women who conform to the traditional female role (e.g., not engaging in agency behaviors) may activate benevolent sexism or prosocial behaviors (Glick & Fiske, 1996, 2001; Jackman, 1994), whereas women who violate gender-role prescriptions (e.g., engaging in aggressive behavior) can trigger hostile sexism (e.g., Eagly & Karau, 2002; Swim, Aikin, Hall, & Hunter, 1995). Consequently, it is common for individuals of either gender to be more shocked by the aggressive behaviors of women than by the aggressive behaviors of men (Deaux & Lafrance, 1998; Knight, Guthrie, Pige, & Fabes, 2002) and to perceive female interactions as more aggressive than male interactions, even when both situations are similar (Condry & Ross, 1985). Also, meta-analyses have indicated that this process may be stronger in male-dominated contexts (Eagly, 2007; Eagly & Diekmann, 2005; Eagly & Karau, 2002). For example, female leaders who employ stereotypically masculine styles tend to be devalued when they are in male-dominated contexts like sport and when the evaluators are male (Eagly, Makhijani, & Klonsky, 1992).

Some researchers have tested the notion that male referees are more shocked by women’s aggressive behavior than men’s aggressive behavior, using observations of referees at soccer (Coulomb-Cabagno et al., 2005) and handball matches (Souchon et al., 2004). These investigators found that men exhibited more aggression in both sports (see also Coulomb-Cabagno & Basse, 2006). More relevant to our hypothesis, these researchers also found that male referees more frequently gave the ball back to the victim of aggressive behaviors in female games than in male games. An additional experimental study also revealed that referees gave disciplinary punishments more frequently when fouls were made by women than by men (Souchon et al., 2004, Study 2). Thus, referees applied the rules of the game differently according to the gender of the players in the game.

Nevertheless, these studies were carried out on matches occurring at a regional level of competition (Coulomb-Cabagno et al., 2005; Souchon et al., 2004), and there are important reasons to test whether a gender bias also occurs at higher levels of competition. First, it is conceivable that professionalization and experience negate the sexist bias at a higher level of competition. Indeed, higher levels of competition may induce more pressure for accuracy and objectivity because of larger audiences and consequences for the competitors and officials. Also, at the highest levels of competition in many sports, including handball and soccer, referees can be appointed to officiate both the men’s and women’s game so players benefit from an equal level of expertise. In contrast, women tend to receive different referees from men at lower levels of competition in some sports, including the sports studied in the past research demonstrating a gender bias. Thus, the higher level of pressure for accuracy and the presence of the same referees for both genders might mitigate against sexist bias at a higher level of competition.
Moreover, gender differences in aggression decrease significantly as the level of competition mounts (Coulomb-Cabagnol & Rasetle, 2006). At higher levels of play, female players are more likely to consider aggressive behavior to be a legitimate means of gaining a tactical advantage over an opponent (Conroy, Silva, Newcomer, Walker, & Johnson, 2001; Silva, 1983). In other words, at the highest level of play, the perceptions and behavior of both male and female players may overlap. This similarity would seem to work against a heuristic that penalizes female players more strongly.

Nevertheless, some anecdotal evidence suggests that these anti-bias factors at a higher level of competition may not be enough to remove gender bias. Specifically, Kolnes (1995) explained that some women playing for the top international Norwegian handball team thought that they are refereed very severely and unfairly when they behave too aggressively or are perceived as acting in a masculine manner. According to these women, severe punishments could be avoided by maintaining a seductive and feminine (e.g., nonargumentative) attitude toward the referees, who are usually men. Other research indicates that men can react negatively toward women who show behavior that the men regard as being overly masculine (Dovidio, Glick, & Rudman, 2005; Eagly, 2007; Glick & Fiske, 1996; Swim et al., 1995). It is therefore possible that referees’ decision making demonstrates the same bias against female transgressions at a high level of competition as at a low level of competition.

To examine these possibilities, we conducted the first investigation of the role of player gender in officiating at the highest national level of competition. We supposed that revealing a gender bias at the highest national level could contribute importantly to illuminating the powerfulness and pervasiveness effect of player gender in refereeing. We chose to study bias at this level by examining decisions in handball because robust procedures for examining gender bias in this context have been established (Souchon et al., 2004). Overall, we expected that referees would punish aggressive behavior from female players more frequently than aggressive behavior from male players. Nevertheless, referees in team contact sport can make different kinds of decisions that differ in severity. For example, they can decide simply to give back the ball to the player victim of a foul or award a penalty (i.e., direct shot at the goal) according to their interpretations. Also, they can warn an aggressor with a yellow card or decide to eject him or her from the game. Because the relatively severe decisions are less common, it is more difficult to detect bias in these types of decisions (Plessner & Betsch, 2002). It was hypothesized then that the bias should be more evident for the relatively common decisions than for the very rare ones.

**METHOD**

**Participants**

The study examined 30 matches taken from the First Division of the French Handball Championships, which is the highest national level for this sport in France. Fifteen women’s matches and 15 men’s matches were filmed in their entirety, with the agreement of the French Handball Federation. The matches were selected by the National Technical Director responsible for appointing officials to obtain a large number of referees and teams. Thus, 27 different pairs of referees were filmed and observed (14 from the Women’s Championship and 13 from the Men’s Championship). All of these referees were men. In addition, 32 teams were filmed and then observed (16 from the Women’s Championship and 16 from the Men’s Championship). This means that most of the referees who had the ability to officiate at the highest national level and each team within the male championship and the female championship were videotaped at least one time.

**Procedure**

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. In handball, like in other team contact sport, the members of the team in possession of the ball are precisely defined as the “attacking players,” while the members of the team not in possession of the ball are identified as the “defending players.” According to rules of handball (International Handball Federation, 2005), referees can interrupt the game and give the ball back to attacking players (players on the offensive) when one or more defenders commits a transgression. Nevertheless, under the advantage rule, referees must not intervene before they are sure that the team on the offensive has lost possession of the ball or cannot pursue their actions because of the foul (International Handball Federation, 2005, Rule 13.2). Because this judgment depends principally on the actions concerning players in possession of the ball (Souchon et al., 2004), we decided to concentrate our observations on the 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 finds himself behind a line of at least four opposing defenders. These attacks make up the greater part of the game (Souchon et al., 2004).

**Type 1 and Type 2 transgressions.** The observation criteria were strictly in accordance with the rules of handball. Any Type 1 (i.e., pushing, pushing away, and barging) and Type 2 (i.e., holding, catching, and seizing around the waist) defensive actions are considered to be transgressions (International Handball Federation, 2005, Rule 8.2). We observed these two types of transgressions during the organized attacks. For instance, a defensive player can “push from the side” an attacking player during a shot or “catch” him or her just before he or she makes a pass. Before carrying out our final observations, four matches (two from the women’s championship and two from the men’s championship) were observed by the main author and two
handball experts. The agreement rate among the observers was between 0.86 and 0.92 (Kappa coefficient). Because the level of the coefficients was satisfactory, the complete matches were watched by only two people: the main author and a handball expert, with an agreement rate of between 0.89 and 0.93. The observations by the expert were made without his being aware of the aim of the study. For each variable, the mean of both observations was determined to derive one measurement.

Observations of refereeing decisions. Each transgression can be penalized from a sporting and/or disciplinary point of view. Anywhere a foul was committed by a player on the court, referees could make one of two possible sporting decisions: (a) award a 9-meter throw against the defending team (simply give back the ball to the victim where there foul occurred) or (b) award a 7-meter throw against the defending team (i.e., a direct shot at the goal like a penalty in soccer). Three possible disciplinary sanctions could be applied: (a) a yellow warning card, (b) a 2-minute suspension from the game, or (c) a red card (i.e., permanent exclusion). The sanctions during a game can depend on the amount of aggression. So, to compare referees' decisions between men's and women's games, a ratio of sanctions to transgressions was calculated for each game (Coulomb-Cabagno et al., 2005).

For sporting sanctions in each match, we calculated the ratio of sporting sanctions awarded by referees after Type 1 fouls against the total number of Type 1 fouls committed for (a) 9-meter throws (SS9/T1T) and (b) 7-meter throws (SS7/T1T). In the same way, the ratio between the number of sporting sanctions awarded by referees after Type 2 transgressions against the total number of Type 2 fouls was calculated for (a) 9-meter throws (SS9/T2T) and (b) 7-meter throws (SS7/T2T).

For disciplinary sanctions in each match, we calculated the ratio of disciplinary sanctions awarded by referees after Type 1 fouls against the total number of Type 1 fouls committed for (a) yellow cards (DSY/T1T), (b) 2-minute suspensions (DS2/T1T), and (c) red cards (DSR/T1T). Finally, the ratio between the number of disciplinary sanctions awarded by referees after Type 2 transgressions against the total number of Type 2 offenses was calculated for (a) yellow cards (DSY/T2T), (b) 2-minute suspensions (SD2/T2T), and (c) red cards (DSR/T2T).

RESULTS

Transgressions

Average transgressions per match were analyzed with a 2 × 2 (player gender × type of transgressions) mixed analysis of variance (ANOVA), with player gender and type of transgression as between-subjects factors. The adjusted $R^2$ was 0.90, and the model was significant, $F(1,56) = 183.54$, $p < .001$. Results indicated main effects of the type of transgression, $F(1,56) = 534.22$, $p < .0001$, and player gender, $F(1,56) = 10.37$, $p < .002$, $n^2 = .16$, such that players exhibited more Type 1 transgressions than Type 2 transgressions, and male players committed more transgressions than female players.

These main effects were qualified by an interaction between the variables, $F(1,56) = 4.23$, $p < .05$, $n^2 = .07$. Male players ($M = 131.2$, $SD = 13.9$) displayed more Type 1 transgressions (pushing, pushing back, and bumping into) than female players ($M = 113.7$, $SD = 12.7$, $p < .0005$). Nevertheless, male players ($M = 47.6$, $SD = 13.59$) displayed the same number of Type 2 transgressions (pulling, holding, and catching) as female players ($M = 43.73$, $SD = 11.01$, $p = .41$). These differences in transgression frequency were consistent with our predictions and supported our decision to analyze refereeing decisions as a proportion of transgressions committed (i.e., controlling transgression frequency) in the analyses to follow.

Sporting Refereeing

Average sporting decisions per match were analyzed with a 2 × 2 × 3 (player gender × type of transgressions × type of sanctions) mixed ANOVA, with player gender (female vs. male), type of transgression (Type 1 vs. Type 2), and type of sanction (no punishment vs. 9-meter throw vs. 7-meter throw) as between-subjects factors. The adjusted $R^2$ was 0.93, and the model was significant, $F(2,168) = 203.75$, $p < .001$. The main effect of severity of sanctions, the interaction between type of transgression and severity of sanctions and the interaction between player gender and severity of sanctions were significant.

The main effect of severity of sanction, $F(2,168) = 1,115.82$, $p < .001$, $n^2 = .89$, indicated that referees were more likely to let the game continue without intervention ($M = 0.64$, $SD = 0.05$) than interrupt the game by granting a 9-meter throw to the victim ($M = 0.32$, $SD = 0.05$), which was awarded more frequently to the victim than the harsher 7-meter throw ($M = 0.04$, $SD = 0.01$). Consequently, referees were less likely to punish players with a direct shot at the goal (i.e., penalty in soccer) than to simply give back the ball to the victims where the defensive fouls were committed.

Table 1 depicts the two-way interaction between type of transgression and severity of sanctions. The table shows that referees were more likely, after Type 1 transgressions, to give no punishment than to award the milder sanction of a 9-meter throw, which was chosen more frequently than the harsher 7-meter throw, $p < .0001$. After Type 2 transgressions, referees were more likely to intervene with the milder sanction of a 9-meter throw than to give no punishment, which was selected more frequently than the harsher sanction of a 7-meter throw, $p < .0001$. Referees intervened more frequently with a nine-meter throw after Type 2 transgressions than after Type 1 transgressions. They also intervened more frequently with 7-meter throws.
after Type 2 transgressions than after Type 1 transgressions. Referees were more likely to give no punishment after Type 1 transgressions than after Type 2 transgressions. Overall, these results indicate that referees preferred more lenient sporting decisions, but more so after Type 1 offenses.

Table 2 depicts the two-way interaction between player gender and severity of sanctions. The table shows that referees were less likely to give the milder 9-meter throws to male players than to female players, \( p = .0006 \), and more likely to give no punishment to male players than to female players, \( p < .001 \). In contrast, referees awarded 7-meter throws to female players as frequently as to male players. Overall, referees were more likely to mete out mild punishment (as opposed to no punishment) to female players, while giving severe punishments to the same extent for both sexes. This result supports our hypothesis that referees would punish aggressive behavior from female players more frequently than aggressive behavior from male players, but only for the more frequent, milder levels of punishment.

Disciplinary Refereeing

Average disciplinary decisions per match were analyzed with a 2 \( \times \) 2 \( \times \) 3 (player gender \( \times \) type of transgressions \( \times \) type of sanctions) mixed ANOVA, with player gender (female vs. male), type of transgression (Type 1 vs. Type 2), and type of sanction (no intervention vs. yellow card vs. 2-minute suspension) as between-subjects factors. Red cards were not entered into the analyses because no red cards were given. The adjusted \( R^2 \) was .99, and the model was significant, \( F(3,224) = 4,851.35, p < .001 \). The main effect of severity of sanctions and the interactions between type of transgression and severity of sanctions were significant. The main effect of sanction, \( F(3,224) = 241,971.09, p < .001, n^2 = .99 \), indicated that referees were more likely to let the game continue without intervention (\( M = 0.95, SD = 0.016 \)) than to punish with a yellow card (\( M = 0.02, SD = 0.008 \)) or a 2-minute suspension (\( M = 0.03, SD = 0.01, p < .0001 \)). In contrast, referees sanctioned players with a yellow card (\( M = 0.02, SD = 0.008 \)) as frequently as with a 2-minute suspension (\( M = 0.03, SD = 0.01, p = .75 \)).

Table 3 depicts the two-way interaction between type of transgression and severity of sanctions. The table shows that referees were more likely, after Type 1 transgressions, to give no punishment than to award a yellow card or a 2-minute suspension, \( p < .0001 \). Referees intervened as frequently with a yellow card as with a 2-minute suspension, \( p = .52 \). Referees were also more likely to give no punishment after Type 2 transgressions than to award a
yellow card or a 2-minute suspension, \( p < .0001 \), but intervened as frequently with a 2-minute suspension as with a yellow card, \( p = .27 \). Referees sanctioned with a yellow card more frequently for Type 2 transgressions than Type 1 transgressions, \( p < .001 \). They also penalized more with a 2-minute suspension Type 2 transgressions than Type 1 transgressions, \( p < .001 \). Referees were more likely to give no punishment toward Type 1 transgressions than toward Type 2 transgressions, \( p < .001 \). Overall, the main effect of sanction and the two-way interaction between type of transgression and severity of sanction indicated that referees preferred giving no punishment, but more so after Type 1 transgressions. There were no effects of gender on disciplinary decisions.

**DISCUSSION**

The aim of this study was to observe the influence of player gender on referees’ decision making at a higher level of competition than has been studied in past research. As expected, a sexist bias in officiating was revealed in the analysis of the milder sporting decisions. This bias was not important in size but has a real practical significance due to the very high referees’ level of expertise and the high players’ professionalization at this standard of play. Of interest, this bias did not occur in the sporting decisions that were relatively harsh (e.g., whether to give a 7-meter throw as a penalty) or in the disciplinary decisions. The lack of effects on the harsh sporting decisions and the disciplinary decisions is likely due to the fact that these decisions are much rarer during games (e.g., Plessner & Betsch, 2002). For example, referees awarded only seven or eight yellow cards or 2-minutes suspension in means by match and did not award a red card within the 30 matches being observed, while players averaged more than 150 transgressions in each match. Thus, only the more commonplace, milder decisions provide a powerful test for bias.

The results for these sporting decisions clearly replicate the prior evidence that referees are likely to use judgmental heuristics to help them in making decisions (Mascarenhas et al., 2006; Plessner & Haar, 2006). One relatively benign interpretation of the gender bias observed in this study is that it occurred because male players could have better ability to continue their action after defensive fouls (Coulomb-Cabagno et al., 2005; Souchon et al., 2004). Referees could then perceive that male attacking players are more physically and technically able to succeed in their pass or shot at a higher level after the defensive fouls than female players. Thus, they can more frequently let the game continue without intervention with male players than female players.

However, this explanation may be more plausible for explaining the prior evidence of gender bias at lower levels of competition than for explaining the bias at this higher, national level. At the higher level, male and female players both have sufficient physical and technical abilities to continue actions after a foul from their opposition (who are similar in strength and skills to the victims). To us, this benign interpretation of the difference in referee judgments underestimates the skill of female players at the national level.

Our evidence for national-level bias provides further support for the hypothesis that stereotypes related to player gender are influencing decision making (Chaiken, Liberman, & Eagly, 1989). In theory, these stereotypes might act through both assimilation and contrast mechanisms. Assimilation and contrast effects occur when the interpretation of ambiguous behaviors is determined by constructs that are consciously or subconsciously activated in memory (i.e., the behaviors can be assimilated or contrasted with a perceptually ready interpretation).

Assimilation could occur in referees’ inferences about the victim of the foul. According to gender stereotypes, men have more physical competence and strength than women (Deaux & Lafrance, 1998; Dovidio et al., 2005). Because forceful and aggressive behaviors are related to performance in team contact sport (Sheldon & Aimar, 2001; Stephens, 1998), individuals often think that male players are better than female players in team contact sport (Fredericks & Eccles, 2005; Krane, 2001). Male referees, culturally socialized in the sporting domain, may develop the stereotype that women have less sporting ability than men (Coulomb-Cabagno et al., 2005; Souchon et al., 2004). If referees consciously or unconsciously expect that female players have less ability to withstand fouls than male players, they could interpret failed performance of a female victim of a defensive transgression within their expectation (assimilation effect). As a result, they may be more likely to believe that a woman who falls during a game is not playing (which is a common question in upper-level male competition). The male referees may then be motivated to award a penalty through their benevolent sexism (Glick & Fiske, 2001; Jackman, 1994). Thus, referees may be more likely to award penalties to compensate female victims than to compensate male victims.

Contrast might occur in referees’ perceptions of the transgressor. The role incongruity theory of prejudice states that prejudice can occur when social perceivers hold a stereotype about a social group that is inconsistent with the attributes that are believed to be required for success in certain social roles (Eagly & Diekmann, 2005). In this context, an aggressive act from a woman contradicts stereotypes about women as being less aggressive. Aggressive behaviors are more shocking when they are displayed by a woman than by a man because female aggression violates traditional gender stereotypes (Burgess & Borgida, 1999; Eagly & Karau, 2002; Heilman, 2001; Knight et al., 2002). If referees have in mind the expectation that women should be unaggressive, a female player’s aggressive behavior may be seen as more aggressive through the contrast effect. From this perspective, referees, influenced by implicit associations, may be motivated to punish the counter-stereotypic behavior of female transgressors. This type of
decision would reflect the hostile component of ambivalent sexism (e.g., Glick & Fiske, 1996; Swim et al., 1995).

Overall, referees could have both a level of hostile sexism and a level of benevolent sexism, and both attitudes may explain why they give the ball back to female players more frequently than to male players. This hypothesis concurs with the perceptions of female handball players from the Norwegian national team who thought that male referees penalize them more when they behave in an overmasculine or overaggressive manner (Kolnes, 1995). Nonetheless, it would be useful to test this explanation by examining referees’ benevolent and hostile sexism in future research using the Ambivalent Sexism Inventory (Glick & Fiske, 2001). This design would reveal whether one or both attitudes contribute to the bias observed here.

The null effect on disciplinary sanctions reveals a vital reason for extending experimental research on bias to real contexts. In the present research, there was no effect of player gender on disciplinary decisions: Male and female players were penalized equally in terms of severity. This result, obtained in an ecological context, is different from evidence obtained in a laboratory experiment by Souchon et al. (2004). The laboratory experiment found that video replays of similar transgressions were judged more severely by male referees when the fouls were committed by female players than when they were committed by male players. Similarly, a study of basketball decisions showed that game situations judged out of context by referees (i.e., video situations showing just physical contact between two opponents) were penalized more severely than in reality or when the events leading up to the fouls were presented to the referees (Brand, Schmidt, & Schneeloch, 2006). Laboratory experiments lack the flow of real game situations, audience pressures, and many other factors (e.g., lack of perspective, time pressure) that may cause referees to give fewer harsh decisions and display bias less strongly.

In summary, the results obtained in this study are consistent with the hypothesis that referees use gender stereotypes as judgmental heuristics at the highest national level of competition, despite the greater pressures for objectivity and fairness at this level. The results indicate that referees may use the gender of players as a judgmental heuristic for deciding how to respond to aggression. Future research should explore the mechanisms underlying these effects through additional experimental or longitudinal research. Equally, it would be useful to examine referees’ decision making in professional sports where there are very highly paid male and female athletes. Despite their level of expertise and willingness to make fair decisions for both male and female teams, referees may be influenced unconsciously by implicit associations learned via exposure to cultural gender stereotypes. These subtle effects are important partly because referees’ decisions in sport may also affect player and fan attributions in reactions to athlete and team success and failure. Thus, any gender bias, however unwittingly it occurs, may have widespread repercussions.

REFERENCES


Krane, V. (2001). We can be athletic and feminine, but do we want to? Challenging hegemonic feminity in women’s sport. *Quest, 53*, 115–133.


