*Journal of Sport & Exercise Psychology,* 2013, 35, 585-599


JOURNAL OF
SPORT & EXERCISE
PSYCHOLOGY
Official Journal of NASPSPA
www.JSEP-Journal.com
ORIGINAL RESEARCH

# The Influence of Referees' Expertise, Gender, Motivation, and Time Constraints on Decisional Bias Against Women

**Nicolas Souchon,[1] Andrew G. Livingstone,[2] and Gregory R. Maio[3]**
[1]Université de Paris Ouest; [2]University of Exeter; [3]Cardiff University

The influence of player gender on referees' decision making was experimentally investigated. In Experiment 1, including 145 male handball referees, we investigated (a) the influence of referees' level of expertise on their decisional biases against women and (b) the referees' gender stereotypes. Results revealed that biases against women were powerful regardless of the referees' level of expertise and that male referees' stereotype toward female players tends to be negative. In Experiment 2, including 115 sport science students, we examined the influence of the participants' gender, motivation to control bias, and time constraints on gender bias. Results indicated that participants' gender had no impact on gender bias and that participants were able to reduce this bias in conditions in which they were motivated to control the bias.

***Keywords****:* gender stereotype, judgmental heuristics, player gender, refereeing, sport psychology

Referees in team contact sports are not meant to intervene automatically when players commit a transgression in the field; rather, they must judge both the victims' performance after the defensive transgression and the danger of the foul. Specifically, they must consider whether the victim player can still profit from his or her action after the foul and whether the foul presented a hazard to the safety of the victimized player (Souchon et al., 2010). These judgments are full of ambiguity, and referees tend to use multiple decision cues, or judgmental heuristics, to help them make their decisions (Plessner, Schweizer, Brand, & O'Hare, 2009). Decision cues or heuristics are simple ways of reasoning that help guide judgments of uncertain events in complex environments (Tversky & Kahneman, 1974), and they might or might not be valid. For example, referees tend to be influenced by the color of players' shirts (Frank & Gillovich, 1988), the passage of the game (e.g., Unkelbach & Memmert, 2008), a player's aggressive reputation (Jones, Paull, & Erskine, 2002), the noise of the crowd (e.g., Nevill, Balmer, & Williams, 2002), and the height of the aggressor (Van Quaquebeke & Giessner, 2010). In addition, referees may use their stereotypes to help them make their decisions (Plessner & Haar, 2006).

Stereotypes, which are a kind of judgmental heuristic, can be defined as the sum of beliefs, knowledge, and expectations that individuals develop toward the members of social categories (Hamilton & Sherman, 1994). Research has shown that referees' decisions may be influenced by stereotypes relating to competition level (Souchon, Cabagno, Traclet et al., 2009) and by gender stereotypes (Souchon, Coulomb-Cabagno, Traclet, & Rascle, 2004; Souchon, Cabagno, Rascle et al., 2009; Souchon et al., 2010). For example, if referees hold a stereotype of women as less competent than men in masculine domains such as team contact sports (Deaux & Lafrance, 1998) and that women should not be aggressive (Burgess & Borgida, 1999; Rudman & Kilianski, 2000), then this stereotype may create expectations that female players are more likely to be neutralized or perturbed by the foul and lead referees to appraise the transgressing female player as being more aggressive. Consequently, female players tend to be more severely sanctioned than male players in team contact sports (Souchon et al., 2004; Souchon, Cabagno, Rascle, et al., 2009; Souchon et al., 2010) and pervasive gender bias in refereeing decisions potentially present a significant barrier to women's feeling free to pursue their sporting interests, including the freedom to play like men if they want to or to not conform to arbitrary stereotypes of femininity.

The previously discussed effect is manifest in two ways. First, observations in ecological settings have consistently revealed that male referees apply sporting sanctions (e.g., awarding a free kick or free throw) more frequently to female players than to male players regardless of the competition level (Coulomb-Cabagno, Rascle, & Souchon, 2005; Souchon et al., 2004; Souchon, Cabagno, Rascle, et al., 2009). More precisely, at intermediate competition levels, Souchon et al. (2010) showed that referees tend to apply sporting sanctions more frequently

Nicolas Souchon is with Université de Paris Ouest, Nanterre, France. Andrew G. Livingstone is with the University of Exeter, Exeter, England, UK. Gregory R. Maio is with Cardiff University, Cardiff, Wales, UK.

to female players than to male players in unsuccessful advantage situations (i.e., when the victim misses the pass or shot after a defensive transgression), but not in successful advantage situations (i.e., when the victim completes the pass or shot after a defensive transgression).

These sporting sanctions are different from disciplinary sanctions that referees can apply. 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 full dismissal (e.g., red card in football). While recent ecological observations revealed that referees tend to punish male players more severely with disciplinary sanctions than female players at intermediate competition levels (Souchon et al., 2010), an experiment showed that male referees tend to punish female players more severely than male players for equivalent transgressions (Souchon et al., 2004). This effect appears to arise because referees could be more "shocked" by female players' aggressive behaviors than by male players' aggressive behaviors, because aggressiveness is a characteristic stereotypically associated with men. In contrast, there is a strong stereotype-based prescription that women should not be aggressive (Burgess & Borgida, 1999; Rudman & Kilianski, 2000).

Despite the robustness of this gender bias, there is a lack of research investigating factors that might attenuate it. A number of unanswered questions are relevant and important. For instance, is the bias lower among referees who are higher in expertise? Do high-level referees possess the same stereotypes as lower-level referees? In addition, does the gender bias occur only in male referees, or is it also apparent in female referees? Can the bias be reduced when there is high motivation to control the bias or when referees have more time to make their decisions? The present research sought to address these questions.

## How Does Gender Bias Arise?

To address these questions, we need to consider how gender biases may arise. If gender stereotypes play a crucial mediating role, then there are several relevant considerations. In general, stereotypes can affect our judgments automatically without our knowledge of how they are operating (Banaji, Hardin, & Rothman, 1993; Devine, 1989). For this reason, a person who is an expert in making particular judgments, such as refereeing, might be as prone to stereotype effects as a person who is less of an expert, because expertise in making a particular type of judgment does not necessarily require knowledge of personal stereotypes and how they are operating. As long as we are ignorant of these stereotypes, they can continue to exert powerful nonconscious effects. Particularly relevant here is evidence that gender stereotypes are pervasive (e.g., Kiefer & Sekaquaptewa, 2007; Rudman & Kilianski, 2000), and observational studies in real match settings have revealed that the tendency for male referees to apply sporting sanctions more frequently to female players emerges at all levels of competition (Souchon et al., 2004; Souchon, Cabagno, Rascle, et al., 2009).

When it comes to the effect of referee gender on gender bias in decisions, the extensive literature on ingroup bias has found a tendency to favor members of our own group over members of another group (Hewstone, Rubin, & Willis, 2002; Tajfel, Billig, Bundy, & Flament, 1971). In relation to sports, for example, Australian 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 & Larsen, 1998). In the Olympic Games, research has found that judges tend to give skaters of their own nationality the maximum score (among the judges), and their overall rank for these skaters tends to be higher than the skaters' final Olympic standings (Whissell, Lyons, Wilkinson, & Whissel, 1993).

However, other research that focuses more specifically on how perceived gender influences gender bias in stereotypically masculine domains, like team contact sports, suggests a different pattern. Research into the "glass cliff" phenomenon has found that both males and females preferred female over male job candidates for precarious leadership positions—in other words, both males and females tend to set women up for a fall in a masculine domain (Haslam & Ryan, 2008). Importantly, this preference is often justified by participants in terms of the stereotypical characteristics of women. Elsewhere, studies of the "queen bee syndrome" have highlighted how successful women working in traditionally masculine domains such as business and academia actually show *more* stereotype-consistent bias against female colleagues (e.g., Derks, Ellemers, Van Laar, & De Groot, 2011). Particularly in view of the wide availability and automaticity of stereotype effects, these findings highlight that stereotypes are powerful determinants of judgments of our *own* groups as well as other groups (e.g., research on stereotype threat: Stone, Lynch, Sjomeling, & Darley, 1999). Thus, if gender stereotypes are automatically driving gender bias in refereeing decisions, then this effect is likely to occur for both male and female referees.

Finally, research on stereotyping processes has revealed that individuals are able to control their stereotypes only under conditions in which they know that a judgmental bias does exist, are strongly motivated to control the bias, and have sufficient cognitive resources available (e.g., Fiske & Neuberg, 1990; Pendry, 1998; Strack & Deutsch, 2004). Motivation to control bias can be high when people hold egalitarian values that oppose prejudice and are made aware of any bias they might hold (e.g., Moskowitz, Gollwitzer, Wasel, & Schaal, 1999).

The availability of cognitive resources is higher when people are less distracted, have more time to make their decisions, and have greater cognitive abilities per se (e.g., Chaiken, Liberman, & Eagly, 1989). The effect of time available for decision making is particularly relevant to the contexts in which referees make their decisions. In team contact sports, these decisions must be made quickly, often within a few seconds of an incident (e.g., Jones et al., 2002; Nevill et al., 2002). This time constraint, together with the presence of distractions (e.g.,

crowd noise, player behavior) could greatly limit referees' ability to deliberatively process the information they have. We can speculate that referees' decisional biases against women should be attenuated only when they are informed about the potential for bias and motivated to control the bias and have sufficient time to control the bias in their judgments. Level of expertise and gender were not expected to moderate the degree of bias.

## *Experiment 1*

Although research has shown that expert referees tend to take more accurate decisions than novice or low-level referees when fouls are clear or obvious (i.e., clearly identified with a high level of agreement between a panel of experts; e.g., Hancock & Ste-Marie, 2013; MacMahon, Helsen, Starkes, Cuypers, & Weston, 2007), it would be very likely that, when faced with ambiguous situations, expert or high-level referees may be influenced by gender stereotypes because the influence of these stereotypes is automatic and unconscious (see, for example, Rudman & Kilianski, 2000). This reasoning is reinforced by Souchon, Cabagno, Rascle, and colleagues' (2009) observation that male referees tend to show bias against female players in the highest-level handball championship in France. Furthermore, Jones and colleagues (2002) found that a team's aggressive reputation affected disciplinary decisions among soccer referees with a high level of expertise (i.e., those who typically officiated in semiprofessional games or have officiated in the English Football League). The first aim of Experiment 1 was therefore to test the hypothesis that referees will be influenced by gender stereotypes regardless of their level of expertise, at both intermediate and national (high) levels of competition. The second aim was to analyze the content of referees' player gender stereotypes. Indeed, referees' player gender stereotypes have never been directly assessed in studies of team contact sports. It therefore remains to be shown whether referees with different levels of expertise have similar gender stereotypes and whether these stereotypes are congruent with referees' decisional biases.

To test these hypotheses, we examined the operation of stereotyping processes in refereeing in a controlled experimental design that was as similar as possible to a real match setting. We chose to conduct this research in relation to the sport of handball because female handball is very well developed in many European countries, making it easier to videotape the required range of female handball games than would have been the case, for example, for female soccer games or female rugby games. Referees had only a few seconds in which to make their decisions without being aware that the aim of the study was to analyze the influence of player gender on their decisions. The situations shown to referees concerned both intermediate and national handball competition level, and referees who varied in expertise were sampled. In the second part of the study, the extent to which referees explain players' gender differences in decision making in terms of specific gender stereotypes was also assessed.

# Method

## Participants and Design

One hundred forty-five male handball referees ($M_{age}$ = 28.20, $SD$ = 12.74; $M_{experience}$ = 7.57, $SD$ = 5.88), including 50 young referees ($M_{age}$ = 15.48, $SD$ = 1.44; $M_{experience}$ = 2.62, $SD$ = 1.49), 48 intermediate referees ($M_{age}$ = 34.85, $SD$ = 10.79; $M_{experience}$ = 10.77, $SD$ = 6.26), and 47 national[1] referees ($M_{age}$ = 34.95, $SD$ = 10.89; $M_{experience}$ = 9.59, $SD$ = 4.92), participated in the study under the supervision of the French Handball Federation. Twelve national-level referees ($M_{age}$ = 34.08, $SD$ = 8.67; $M_{experience}$ = 10.15, $SD$ = 4.76) also completed a pilot study to assess the perceived similarity of situations involving male players and female players. The main study consisted of two stages: a decision-making task based on video stimuli, and assessment of referees' stereotypes and decision-making explanations. The research and the pilot study were approved by an ethics committee, and participants gave their informed consent.

## Pilot Study

Twenty-seven matches (18 matches from the intermediate level and 9 matches from the national level) were recorded using digital video equipment. The camera, placed on the lowest level of the spectator gallery, focused only on players in possession of the ball (with a ×2 zoom) and the results of situations (e.g., missed shot). The first author then carefully selected pairs of situations that were judged to be equivalent between male players and female players. There were 32 pairs at intermediate level and 40 at national level. In each situation, the attacking player missed his or her pass or shot after a defensive transgression.

Only one instance of contact between opposing players was visible in each situation. The performance of the player in possession after the defensive contact was also apparent (i.e., missed pass or missed shot), but neither the actual match referees' decision nor players' reactions after the incidents were visible (the appearance of referees was blurred when they were in the picture). Importantly, the first and last images of each situation were frozen for 1 s to allow participants to make their decisions.

Twelve national referees then completed—for each of the 72 pairs of situations—at home, individually, and within a two-month sampling period the following two questions on a scale ranging from 1 (absolutely not) to 5 (strongly similar): (a) "In your opinion, are the two situations identical or strictly comparable as regards the decision-making processes that are involved?"; (b) "Is the intensity of contact really identical/strictly comparable from your point of view?" Situations were considered to be similar when the means to questions a and b were between 4 and 5: average intraclass correlation coefficient = .75 for question a and .76 for question b. Moreover, referees had to indicate if they would apply sporting sanctions in each situation.

As part of this process, six pairs of equivalent situations at intermediate level and six pairs of equivalent situations at the national level were selected. These situations were ambiguous because referees in the pilot study were in disagreement concerning the sporting and the disciplinary decision to make for each situation (i.e., in terms of whether to apply sporting sanctions $M_{agreement}$ = .64, at the intermediate level and $M_{agreement}$ = .66, at the national level). These six ambiguous situations included two missed-pass situations and four missed-shot situations.

## Materials and Procedure

Two different video files were developed with Adobe Premier Software (sixth version): one containing only situations involving male players (60 situations) and one containing only situations involving female players (62 situations). Within these two video files, situations were grouped into three different competition levels: highest local level (20 situations each involving men and women), intermediate level (23 situations involving men and 26 involving women), and second-highest national level (17 situations involving men and 16 involving women). Within both the intermediate level and the second-highest national level, six male situations were matched in terms of similarity with six female situations, producing a 6 (situation) × 2 (player gender) design within both the intermediate level and the national level. All other situations were neutral and added to maximize the ecological validity of the study and to disguise the specific focus on ambiguous situations.

Before each video mounting, participants had to complete a short introductory task in which they watched a random selection of six game situations (lasting between 20 and 40 s each), which came in reality from games at the local level (two situations), intermediate level (two situations), and national level (two situations). They were asked to gauge the level of competition and quickly write down one, two, or three reasons to justify their choice (e.g., a lot of spectators indicating a high level of competition).

Digital video files were projected onto a 1.42 × 1.88 m display (Jones et al., 2002). To familiarize themselves with the task and apparatus, participants were presented with four example situations. The situations in each video file were numbered from 1 to 60 (male video) and 1 to 62 (female video), and the number appeared for one second before each situation. A 5-s countdown at the end of each situation constituted the window in which participants had to make the refereeing decision. An audio effect at the end of the 5 s informed participants that a new situation was about to be presented. Between each section of the video (e.g., between the introductory task and the first block of situations), participants were informed during a 20 s interval as to the competition level of the upcoming block of situations. For both video files, the intermediate-level situations appeared first, the local-level situations second, and the national-level situations third. They then had a 20 s countdown before the first situation of the block was presented, giving a total of 40 s recuperation time between blocks. Participants had 2-min recuperation periods between each video file. While each participant viewed all situations in both videos in the main study, the order of presentation was varied to minimize order effects, such that 60 referees viewed the male video file first, while 85 viewed the female file first.

Testing was carried out during four referees' meetings. The experimenter explained that the aim of the study was to better understand referees' decision making. To cover a larger variety of situations, the experimenter explained that participants were to judge handball game situations both from different competition levels and from the male and the female championships. They were to make decisions relating to one set of male player situations and subsequently one set of female player situations (or the inverse in the other condition). Referees received oral and written instructions that the players that they had to judge in the video situations were highly typical of their competition level and player gender and that all the situations took place in the first 15 min of the game (i.e., toward the beginning of the match).

The referees made their decisions individually (without communication with other people) using a questionnaire for the male video file and a separate questionnaire for the female video file. For each situation, referees had to make one sporting and one disciplinary decision. They had to put a tick in one of four different boxes for sporting decisions: no intervention, 9-m throw, 7-m throw, or attacking foul. For disciplinary decisions, they had to put a tick in one of four different boxes: no intervention, yellow card, 2-min suspension, or red card (in order of increasing severity).

To gauge the extent to which they explained and justified their decisions in terms of gender-stereotypic beliefs, referees were then given three questions asking them to explain why refereeing may be different according to the sex of players (see Appendix). Questions were derived from observations displayed from the lowest to the highest competition level (Local, Intermediate and National) within both the male and the female handball championships in France (Souchon, Livingstone, Rascle, Cabagno, & Maio, submitted). At the end, they were asked to place the questionnaire in a box to guarantee anonymity. After data collection, which took approximately 40 min per participant, participants were asked about what they believed to be the purpose of the study. During debriefing, only one referee asked why they had to judge female and male players. All of the referees thought that the study focused on the influence of players' level of competition on their decision making.

# Results

The dependent variables were the sporting sanctions and the disciplinary sanctions applied by participants. Sporting decisions were coded as 1 if the referee opted to return the ball to the attacking player (i.e., a 7-m or

9-m throw) and 0 if the referee decided to give the ball to the defensive players (i.e., no intervention or attacking foul). Disciplinary decisions were also coded as 1 (yellow, 2-min suspension, or red card) or 0 (no punishment).

A mixed-model ANOVA (2 × 3 × 2 × 2 × 2) on decisions was conducted. Video order (video mounting with male situations first vs. video mounting with female situations first) and level of expertise (young referees vs. intermediate referees vs. national referees) served as between-subject factors, while player sex (women vs. men players), type of sanction (sporting vs. disciplinary) and competition level (intermediate test situations vs. national test situations) served as within-subject factors.

Results indicated a significant main effect of player gender, $F(1, 139) = 104.01$, $p < .001$, $\eta^2_p = .43$; a significant main effect of expertise, $F(1, 139) = 42.04$, $p < .001$, $\eta^2_p = .37$; a significant main effect of type of sanction, $F(1, 139) = 680.41$, $p < .001$, $\eta^2_p = .83$; and a significant main effect of competition level $F(1, 139) = 170.09$, $p < .001$, $\eta^2_p = .55$. Results also indicated a significant two-way interaction between player gender and expertise, $F(2, 139) = 8.23$, $p < .001$, $\eta^2_p = .10$ and a significant three-way interaction between player gender, type of sanction, and competition level, $F(1, 139) = 20.35$, $p < .01$, $\eta^2_p = .12$.

Table 1 depicts main effects and interactions effects. The main effect of player gender indicated that participants were more likely to apply sanctions to female players than to male players ($p < .001$). The main effect of level of expertise indicated that junior referees tended to apply more sanctions than intermediate and national referees ($ps < .001$). The difference between intermediate- and national-level referees was not significant ($p > .05$). The main effect of the type of sanction indicated that referees were more likely to apply sporting sanctions than disciplinary sanctions ($p < .001$). The main effect of competition level indicated that referees were more likely to intervene for the national-level test situations than for the intermediate-level test situations ($p < .001$).

The two-way interaction between player gender and expertise indicated that junior referees, intermediate referees, and national referees ($ps < .001$) were more likely to intervene with female players than with male

**Table 1 Means (*SD*) Concerning Main and Interactions Effects in Experiment 1**

| | Mean (*SD*) | | | $\eta^2_p$ |
|---|---|---|---|---|
| Player Gender (PG) | Female Players | Male Players | | .43 |
| | .59 (.15)** | .46 (.13) | | |
| Expertise (E) | Junior R | Intermediate R | National R | .37 |
| | .62 (.11)** | .46 (.08) | .49 (.08) | |
| Type of Sanction (TS) | Sporting | Disciplinary | | .83 |
| | .67 (.11)** | .37 (.15) | | |
| Competition Level (CL) | Intermediate | National | | .55 |
| | .45 (.15)** | .60 (.12) | | |
| PG × E | Female Players | Male Players | | .10 |
| Junior R | .71 (.11)** | .52 (.15) | | |
| Intermediate R | .51 (.10)** | .42 (.10) | | |
| National R | .54 (.12)** | .43 (.11) | | |
| PG × TS × CL | Sporting | Disciplinary | | .12 |
| | Female / Male | Female / Male | | |
| Intermediate Level | .65 (.19)** / .55 (.22)** | .37 (.22)** / .21 (.20) | | |
| National Level | .82 (.15)** / .67 (.18)** | .49 (.22)** / .40 (.19) | | |

*Intercategory difference is significant $p < .05$ and **$p < .001$ (from the category on the left to the category on the right).

players. Junior referees applied more sanctions to female players than did intermediate or national referees ($ps < .001$). Junior referees also applied more sanctions to male players than did intermediate or national referees ($ps < .001$). Intermediate and national referees did not differ in their application of sanctions to male players or to female players ($ps > .05$).

The three-way interaction between player gender, type of sanction, and competition level indicated that referees were more likely to apply sporting sanctions to female players than to male players at the intermediate competition level and at the national competition level ($ps < .001$). The interaction also indicated that referees were more likely to apply disciplinary sanctions to female players than to male players at the intermediate competition level and at the national competition level ($ps < .001$). Referees tended to apply sporting sanctions more frequently to female players at the national competition level than in the three other conditions ($ps < .05$). Female players at the national competition level received more disciplinary sanctions than in the three other conditions ($ps < .05$). Whatever the competition level, for both the male and the female situations, referees applied more sporting sanctions than disciplinary sanctions.

### Stereotypes

The data relating to participants' gender stereotypes were content analyzed. Two analysts initially identified and coded 634 individual units in different inductively created categories (Kippendorf, 2013) from the referees' responses to the two questions (agreement and explanation) concerning the three different observations made (see Appendix). After agreement and corrections in conjunction with an external researcher, these individual units were organized into 21 categories. The rate of agreement between the two main coders was 89.8%, and the rate of agreement between the first coders was 95.4%. Participants broadly used three types of explanations to justify their decision: (a) gender stereotypes and explanations related to the influence of stereotypes (17 categories, 457 units), (b) characteristic or context of the game (36 units; e.g., it is more difficult to notice incidents in the masculine game than in the feminine game due to the speed of play), and (c) other categories (141 units) related to "referees not being explicit/understandable" (72 units), "referee disagreement" (46 units), and "avoiding the issue of player gender completely" (23 units). Table 2 depicts how the specific content of referees' gender stereotypes and explanations related to the influence of stereotypes (in terms of subjective explanations) at each level of expertise.

## Discussion

Experiment 1 tested the hypothesis that referees will be influenced by gender stereotypes regardless of their level of expertise at both intermediate and national levels. The second aim was to analyze the extent to which referees employ gender stereotypes to explain and justify gender bias in refereeing decisions.

Consistent with our hypothesis, the effect of gender in our analyses indicated that, across levels of expertise, the referees applied sanctions more frequently to female players than to male players in each video order condition and at each level of competition. Although the interaction effects indicated that the magnitude of the effect of gender was moderated by referee expertise and competition level, the simple effect of gender was still significant in all cases, indicating that gender bias is evident even among the most expert referees and at the highest levels of competition. These results are consistent with other findings that player gender can influence male referees' application of sporting sanctions at both an intermediate competition level (Coulomb-Cabagno et al., 2005; Souchon et al., 2004, Souchon et al., 2010) and a national competition level (Souchon, Cabagno, Rascle et al., 2009). The effect of player gender is consistent with studies showing that men generally tend to perceive women to be less competent than men in stereotypically masculine domains (e.g., Eagly, Karau, & Makhijani, 1995) and with the fact that male referees questioned in this study tended to believe that female players are less competent than male players. These negative gender stereotypes might then make referees more likely to interpret similar transgressions differently. For example, despite the similarity of the situations and the fouls, referees might infer that a female player will lose her balance after the transgression but that a male player will not. They therefore apply sporting sanctions more frequently (i.e., apply the advantage rule less frequently) to female victims of a defensive transgression to compensate for their inability to carry on (i.e., gain an advantage) after the transgression.

Also consistent with our hypothesis was the finding that male referees at all levels of expertise applied disciplinary sanctions more often to female transgressors than male ones in each video order condition and at each level of competition. These results are consistent with those of Souchon et al. (2004), who found experimentally that male referees penalize female transgressors more severely than males ones, but are in disagreement with those of Souchon et al. (2010), who observed that male players were under some conditions punished more severely than female players. This disagreement may be explained by the fact that some game situations involving male and female players may be not totally comparable (e.g., more intensity, or more multiple transgressions among male players) in ecologically valid observations, while in this research the conditions under which referees made their decisions were rigorously controlled. Referees could become more likely to punish female transgressors more severely because the foul may be perceived to be more dangerous for the female victim player or that female players need more protection than male players (Glick & Fiske, 1996). Consistent with this explanation, referees in this research tended to indicate that it was more important to protect female players than male players.

**Table 2 Percentage (Number of Units) of Referees Who Stated at Least One Stereotypical Explanation Related to Gender**

| | Young | Regional | National | Together |
|---|---|---|---|---|
| **Gender Stereotype-Based Explanations for Bias** | | | | |
| **Attacking player abilities**—male players tend to have more skills | | | | |
| (%) Ability to resist defensive transgression/ retain ball | 12 (7) | 12.5 (8) | 23.4 (12) | 15.8 (27) |
| (%) Physical abilities | 16 (9) | 12.5 (6) | 31.9 (15) | 20 (30) |
| (%) Technical abilities | 2 (1) | 31.2 (20) | 40.4 (25) | 24.1 (46) |
| *Total abilities* | **24 (17)** | **37.5 (34)** | **65.9 (52)** | **42.1 (103)** |
| **Physical involvement and risk taken**—male players tend to be more physically involved | | | | |
| Attacking player physical involvement | | | | |
| (%) Greater physical involvement | 8 (5) | 18.8 (11) | 29.8 (16) | 18.6 (32) |
| (%) Men more determined to continue their attack after a defensive foul | 2 (1) | 4.1 (2) | 12.8 (8) | 6.2 (11) |
| Defensive player physical involvement | | | | |
| (%) Men are more aggressive | 14 (7) | 2.1 (1) | 12.5 (6) | 9.6 (14) |
| In general | | | | |
| (%) Men's game is tougher | 26 (16) | 14.6 (7) | 10.6 (5) | 17.2 (28) |
| *Total physical involvement* | **44 (29)** | **35.4 (21)** | **55.3 (35)** | **44.8 (85)** |
| **Collective performance**—male players tend to have more collective skill | | | | |
| (%) Better collective level | 0 (0) | 4.2 (2) | 8.5 (4) | 4.1 (6) |
| **Other** | | | | |
| (%) Men are better in general | 10 (5) | 8.3 (5) | 14.8 (7) | 11 (17) |
| (%) Better training | 0 | 8.3 (4) | 2.1 (1) | 3.4 (5) |
| (%) Better rules understanding | 4 (2) | 0 (0) | 2.1 (1) | 2 (3) |
| **Total Explanation** | **64 (53)** | **68.7 (66)** | **87.2 (100)** | **73.1 (219)** |
| Subcategory: High-level competition players | | | | |
| (%) Women tend to play like men | 16 (8) | 14.6 (9) | 31.9 (17) | 20.7 (34) |
| **Gender stereotype-based justifications for bias** | | | | |
| (%) Contact can be more severe before need to return ball to men, so less benevolence than with females | 10 (5) | 14.6 (7) | 25.5 (24) | 16.5 (36) |
| (%) Women need to be more protected | 18 (10) | 20.1 (10) | 23.4 (14) | 20 (34) |
| (%) More "shocked" by female fouls | 2 (1) | 0 (0) | 2.1 (1) | 1.3 (2) |
| **Total Justification** | 26 (16) | 33.3 (17) | 44.6 (39) | 35.2 (72) |
| **Counter gender stereotype-based explanations for bias** | | | | |
| (%) Better female players collective level | 0 | 2.1 (1) | 2.1 (1) | 1.4 (2) |
| (%) Defensive female players are more aggressive | 24 (14) | 10.4 (6) | 12.7 (7) | 15.8 (27) |
| **Total Counter gender stereotype** | **24 (14)** | **12.5 (7)** | **14.9 (8)** | **17.2 (29)** |

Correspondingly, other research has revealed that aggressive behaviors are more shocking when they are displayed by a woman than by a man because female aggression violates traditional gender stereotypes not to be aggressive (Burgess & Borgida, 1999; Eagly & Karau, 2002).

Finally, the content analysis of referees' explanations for gender bias indicates not only the wide availability of gender stereotypes but also that these stereotypes are used both to *explain* the bias, and also to *justify* it (e.g., Hoffman & Hurst, 1990; Tajfel, 1981). Thus, while referees explain bias in decisions as being due to stereotypic expectations of player competences (e.g., that female players are less skillful and aggressive); they also use these stereotypic expectations to *justify* the bias through, for example, the suggestion that female players therefore need greater protection. This active justification suggests that the influence of stereotypes is important not just as a post hoc way of explaining bias, but as a way of perpetuating future bias. We will return to this point in the General Discussion.

## *Experiment 2*

The first study revealed that male referees, whatever their level of expertise, applied sporting and disciplinary sanctions more frequently to female players than to male players. This raises the question of whether these processes depend on referees' gender. In other words, does the gender bias emerge among female referees?

### Referees' Gender and Biases Toward Female Players

Experiment 2 had two goals. In addition to replicating the results of the first experiment, we tested the effects of a motivation to control gender bias and of time constraints before the decisions. Motivation was manipulated by giving (vs. not giving) participants information about gender bias in refereeing. Given egalitarian norms, we expected that the information about bias would elicit a motivation to reduce it. Time constraints were manipulated by giving participants 3 s or 10 s in which to make their decisions.

## Method

### Participants and Design

One hundred fifteen sport science students ($M_{age}$ = 20.76, $SD$ = 5.04; $M_{experience}$ = 8.61, $SD$ = 4.66), including 57 female players ($M_{age}$ = 19.82, $SD$ = 1.31; $M_{experience}$ = 7.15, $SD$ = 3.57) and 58 male players ($M_{age}$ = 21.69, $SD$ = 6.89; $M_{experience}$ = 10.03, $SD$ = 5.17), participated in the experiment. These participants specialized in team sports (7 volleyball players, $M_{experience}$ = 6.00, $SD$ = 3.46), semicontact team sports (65 handball players, $M_{experience}$ = 8.76, $SD$ = 4.74), 18 soccer players ($M_{experience}$ = 10.02, $SD$ = 4.33), or 7 basketball players ($M_{experience}$ = 10.14, $SD$ = 3.13) or contact team sports (18 rugby players, $M_{experience}$ = 7.05, $SD$ = 5.95) at three competition levels (41 local, $M_{age}$ = 20.02, $SD$ = 1.28, $M_{experience}$ = 7.09, $SD$ = 4.37; 62 intermediate, $M_{age}$ = 20.17, $SD$ = 2.15, $M_{experience}$ = 8.69, $SD$ = 3.65; 12 national, $M_{age}$ = 26.33, $SD$ = 3.95, $M_{experience}$ = 13.33, $SD$ = 6.98). Participants made decisions in one of four conditions: (a) no information on biases and 3 s to make decisions ($N$ = 21), (b) no information on biases and 10 s to make decisions ($N$ = 31), (c) information on biases and 3 s to make decisions ($N$ = 18), or (d) information on biases and 10 s to make decisions ($N$ = 45). The research was approved by an ethics committee, and participants gave their informed consent.

### Materials

The video files used in this experiment were identical to the files used in the first experiment, except that the local situations were deleted to make the video files shorter. In the two video files (male players and female players), situations were then grouped into two competition levels: intermediate level (23 situations involving men and 24 involving women) and second-highest national level (17 situations involving men and 16 involving women).

### Procedure

Testing was carried out during meetings with students specialized in team sports (with a preference for semicontact team sports or contact team sports) in different sport science departments. The experimenter explained that the aim of the study was to better understand referees' decision making. In the *no bias information* conditions, referees received no special oral or written instructions. In the *bias information* conditions, referees received orally the information that (a) the procedure was designed to study whether referees made different decisions as a function of player gender, and (b) previous research has shown that referees tend to be more severe toward female players than to male players both with sporting and disciplinary sanctions. We then stated that participants should make egalitarian decisions between female players and male players because the French Handball Federation states that there is no reason to make different refereeing decisions according to player gender.

The referees made their decision individually (without communication with other people) using a questionnaire for the male video file and a separate questionnaire for the female video file. For each situation, referees had to make one sporting and one disciplinary decision. They responded on scales ranging from −3 (let the attacking players recover the ball despite the defensive transgression and let the game continue without intervening at all) to +3 (intervene to return the ball to the victim of a defensive transgression and severely punish the transgressor with a disciplinary punishment) for the sporting decision and the disciplinary decision.

Digital video files were again projected onto a 1.42 × 1.88 m display (Jones et al., 2002). To familiarize

themselves with the task and apparatus, participants were presented with four example situations. Data collection took approximately 25 min per participant, and they were then debriefed.

## Results

A mixed-model ANOVA ($2 \times 2 \times 2 \times 2 \times 2 \times 2$) on decisions was conducted. Participants' gender, motivation (spontaneous vs. controlled), and time constraint condition (3 s vs. 10 s) served as between-subject factors, while player gender (female vs. male players), type of sanction (sporting vs. disciplinary), and competition level (intermediate situations vs. national situations) served as within-subject factors.

Results indicated a significant main effect of player gender, $F(1, 107) = 127.29$, $p < .001$, $\eta^2_p = .54$; a significant main effect of type of sanction, $F(1, 107) = 51.28$, $p < .001$, $\eta^2_p = .32$; a significant main effect of competition level, $F(1, 107) = 89.88$, $p < .001$, $\eta^2_p = .45$; a significant two-way interaction between type of sanction and participants' gender, $F(1, 107) = 10.16$, $p < .05$, $\eta^2_p = .09$; a significant three-way interaction between player gender, type of sanction, and competition level, $F(1, 107) = 10.87$, $p < .01$, $\eta^2_p = .08$; a significant three-way interaction between type of sanction, competition level, and time constraint, $F(1, 107) = 8.29$, $p < .01$, $\eta^2_p = .07$; a significant three-way interaction between type of sanction, motivation, and time constraint, $F(1, 107) = 4.57$, $p < .05$, $\eta^2_p = .04$; and a four-way interaction between player gender, motivation, time constraint, and participants' gender, $F(3, 93) = 5.21$, $p < .03$, $\eta^2_p = .05$.

Table 3 depicts main effects and the three-way interaction between player gender, type of sanction, and competition level and the four-way interaction between player gender, motivation, time constraint, and participants' gender. The main effect of player gender indicated that participants were more likely to apply sanctions to female players than to male players ($p < .001$). The main effect of the type of sanction indicated that participants were more likely to apply a sporting sanction than a disciplinary sanction ($p < .001$). The main effect of competition level indicated that participants were more likely to apply sanctions within national-level test situations than within intermediate-level test situations ($p < .001$). The interaction between player gender, type of sanction, and competition level indicated that participants were more likely to apply sporting sanctions to female players than to male players at both the intermediate and the national levels ($ps < .001$). They were also more likely to punish female players than male players with a disciplinary sanction at both the intermediate level ($p < .001$) and the national level ($p < .01$), with a stronger gender bias at the intermediate level than at the national level.

The significant interaction between type of sanction and participants' gender; the significant interaction between type of sanction, competition level, and time constraint; and the significant interaction between type of sanction, motivation, and time constraint were qualified by the significant four-way interaction between player gender, motivation, time constraint, and participants' gender. Table 3 indicates that participants applied more sanctions to female players than to male players under spontaneous (unmotivated) conditions whatever the time constraints and the participants' gender. In contrast, participants tended to apply sanctions equally to female players and male players in the controlled (motivated) condition whatever the time constraint condition. The exception to this was that female participants in the motivated condition with a 3-s time constraint applied more sanctions to female players than male players.

## Discussion

The two goals of the second experiment were to replicate the results of the first experiment among both male and female judges and to test the conditions under which referees may control gender bias in their decisions. Concerning the first goal, results indicated strongly that both male and female participants made more severe sporting and disciplinary decisions toward female players than male players. This finding replicates the strong biases evident in the first experiment and is consistent with observations made at intermediate (Souchon et al. 2010) and national competition levels (Souchon, Cabagno, Rascle, et al. 2009). These results go one step further by revealing that both male and female participants tend to be biased against female players. This bias is consistent with the idea that female participants may be automatically influenced by the stereotype of their own group, as has been shown in the stereotype threat literature (e.g., Stone et al., 1999).

Concerning the second goal, results revealed that participants in the spontaneous condition displayed gender bias whatever the time constraints and the participants' gender. In contrast, and with the exception of female participants in the controlled condition with 3 s to make decisions, participants in the controlled conditions did not display gender bias. These results are consistent with our hypothesis, in that while participants may have applied their gender stereotypes and expectations in the spontaneous condition, this was overridden by an egalitarian norm in the controlled condition. Nevertheless, research has shown that controlling stereotypes or expectation effects tends to be very difficult under time pressure or high cognitive load despite a strong motivation to control biases (Fiske & Neuberg, 1990; Pendry, 1998). This may explain why female participants have failed to control the gender bias in the condition in which they get only 3 s to make their decisions.

## *General Discussion*

The aims of this research were to address several important but unanswered questions regarding gender bias in referees' decisions. Specifically, is gender bias lower among referees who are higher in expertise? Do referees

**Table 3 Means (*SD*) Concerning Main and Interactions Effects in Experiment 2**

| | Mean (*SD*) | | | | $\eta^2_p$ |
|---|---|---|---|---|---|
| Player Gender (PG) | Female Players | | Male Players | | .54 |
| | .46 (.63)** | | −.26 (.62) | | |
| Type of Sanction (TS) | Sporting | | Disciplinary | | .32 |
| | .58 (.88)** | | −.38 (.83) | | |
| Competition Level (CL) | Intermediate | | National | | .45 |
| | −.21 (.58)** | | .40 (.65) | | |
| PG × TS × CL | Sporting | | Disciplinary | | .08 |
| | Female | Male | Female | Male | |
| Intermediate Level | .63 (1.09)** | −.16 (1.08) | −.18 (.93)** | −1.12 (1.06) | |
| National Level | 1.32 (1.11)** | .52 (1.10)** | .07 (1.20)* | −.31 (1.01) | |
| PG × M × TC × G | | | | | .05 |
| Male Participant | 3 s | | 10 s | | |
| | Female | Male | Female | Male | |
| Spontaneous (Unmotivated) | .55 (.59)** | −.30 (.56)* | .54 (.84)* | −.18 (.60) | |
| Controlled (Motivated) | .51 (.54) | -.04 (.95) | .30 (.72) | −.19 (.69) | |
| Female Participant | 3 s | | 10 s | | |
| | Female | Male | Female | Male | |
| Spontaneous (Unmotivated) | .78 (.46)* | .15 (.37) | .50 (.47)** | −.54 (.44) | |
| Controlled (Motivated) | .67 (.76)** | −.55 (.57)* | .29 (.67) | −.23 (.55) | |

*Intercategory difference is significant $p < .05$ and **$p < .001$ (from the category on the left to the category on the right).

*Note*. M is motivation, TC is time constraint, and G is participants' gender.

report stereotypic beliefs regarding gender and use these to explain gender bias in decisions? Does gender bias occur only in male referees' decisions and not in female referees' decisions? Can the bias be reduced when there is high motivation to control the bias or when referees have more time to make their decisions?

Concerning the first question, results indicated that referees from all levels of expertise tend to apply sanctions differently to female players. While other research has shown that expert referees tend to take more accurate decisions than lower-expertise referees concerning clear situations (Hancock & Ste-Marie, 2013; MacMahon et al., 2007), gender bias was, as expected, evident in the present research among expert referees when judging ambiguous situations. This finding is congruent with research showing that national referees in handball applied more severe sporting sanctions against female players at the highest national competition level (Souchon, Cabagno, Rascle et al., 2009) and with research showing that high competition level soccer referees tend to be influenced by a team's aggressive reputation (Jones et al., 2002), and that soccer referees of all levels of experience tend to be influenced by the noise of the crowd (Nevill et al., 2002), and that elite standard assistant referees in soccer show difficulties judging offside situations (e.g., Catteeuw, Gilis, Wagemans, et al., 2010; Catteeuw, Gilis, Jaspers, et al., 2010).

Although the present experiments did not directly test the impact of referees' gender stereotypes as a mediator in explaining the gender biases, previous research has shown that gender stereotypes can have a pervasive effect without people being aware of the process (e.g., Banaji et al., 1993; Kiefer & Sekaquaptewa, 2007). More

generally, automatic cognitive associations learned in memory through conditioning processes tend to better explain spontaneous decisions and behaviors than explicit or deliberated cognition or evaluation (e.g., Devine, 1989; Strack & Deutsch, 2004). It is therefore plausible that, through cultural and personal socialization, referees from all levels of expertise could develop explicit or deliberated negative stereotypes of female players as less competent than male players and thus unconsciously associate in memory female players with a lower level of sporting skill (see, for example, Blair, Ma, & Lenton, 2001 for the implicit strong and weak gender stereotype). Implicit stereotyping processes therefore have the potential to influence expert judges as powerfully as nonexpert judges under conditions in which they are not as conscious of the influence of player gender on their decision making.

In relation to the explanatory role of gender stereotypes when it comes to the gender bias in referees' decisions, we found on the one hand that referees with different levels of expertise tended to report structurally similar and shared explicit gender stereotypes as explanations for gender bias in decisions. Referees mainly described female players as being less competent than male players. More specifically, female players were characterized as being less physically and technically able to resist defensive transgressions and retain the ball, as taking fewer physical risks, and as being less physically involved than male players. The clear availability of—and consensus in—such stereotypes indicates the potential for referees to implicitly associate female players with a lesser level of sporting skill, particularly given the general correspondence between implicit and explicit stereotypes (Nosek et al., 2007). An interesting future research question could be to test whether implicit and explicit gender stereotypes differentially predict gender biases and if referee expertise moderates any effect.

Notwithstanding the outcomes of such research, the gender stereotypes spontaneously invoked by referees were used as a basis for *justifying* differential treatment of male and female players. In particular, referees at all levels of expertise suggested that female players would need more protection than male players, while others suggested that equivalent aggressive behaviors could be more shocking if perpetrated by female players. This resonates with more general findings that stereotypes are invoked not only as in situ or post hoc explanations for aspects of the social world but are also used to rationalize and justify present and future treatment of members of different social categories (Hoffman & Hurst, 1990; Tajfel, 1981). This extends to ostensibly benevolent and helpful behavior that is based on the supposed low competence of an outgroup (Fiske, Cuddy, Glick, & Xu, 2002). For example, ambivalent sexism theory suggests that gender stereotypes lead some individuals to be more benevolent (e.g., paternalist) toward women when they conform to the traditional gender stereotype (i.e., being sweet, feminine), but to be harsher or hostile toward women if they transgress their gender role (e.g., by being aggressive or competitive) (Glick & Fiske, 1996). In terms of gender bias in refereeing decisions, this approach would suggest that gender stereotypes lead to more benevolent or protective interventions on behalf of female victims of defensive transgressions along with more severe punishment of "aggressive" transgressions by female players. Because referees' responses concerning refereeing biases could be prone to social desirability concerns, it is also possible that both explicit and implicit processes contribute to explaining these gender biases (see also Souchon, Cabagno, Rascle, et al. 2009 for a discussion of these processes).

Crucially, as Study 2 indicates, it also appears that female and male referees hold similar gender stereotypes and attitudes. Indeed, female participants in Study 2 generally displayed as much gender bias as male participants. These findings indicate that gender biases against female players were highly pervasive and were evident across different levels of players' competition level, referees' level of expertise, and participants' gender. This is a particularly notable finding because most of the previous research on gender bias in refereeing decisions has involved a less powerful or controlled methodology, such as observations made mainly at the intermediate competition level and only with male referees (e.g., Souchon, Cabagno, Rascle, et al., 2009; Souchon et al., 2010).

Our final research question concerned the controllability of gender bias. Our findings indicated that participants made more egalitarian decisions (i.e., showed no gender bias) when they were made explicitly aware of the gender bias and the fact that this contravenes referee guidelines. This result is congruent with other studies showing that individuals may be concerned with not appearing sexist for fear of the social disapproval that sexist responses or behavior might elicit (Klonis, Plant, & Devine, 2005). This result is also congruent with research showing that stereotype activation can be controlled by having chronic egalitarian goals (Moskowitz, et al., 1999). Nevertheless, in our research it is not possible to discern whether it was external pressure (i.e., social norms of not being or appearing sexist), internal pressure (i.e., egalitarian goals or values), or both that led our participants to control gender bias. It is also possible of course that the process may vary across individuals. In any case, the finding that referees may mitigate against gender bias clearly extends previous research on referees' decision making. Several judgmental heuristics have been shown to be used in refereeing (e.g., Frank & Gillovich, 1988; Jones et al., 2002), but no research has shown that referees may control their use under certain conditions, such as when the egalitarian norms of refereeing are made salient. That said, suppressing stereotypes requires a great deal of effort and cognitive capacity: for example, in this research, female participants still displayed a gender bias when they had only 3 s to make their decisions.

One limitation of this research concerns the representativeness of our sample and how we could generalize our findings to other team contact sports. In particular, we focused quite specifically on gender stereotypes and explanations related to player gender differences in hand-

ball. This raises the question of whether these stereotypes and explanations apply in other team contact sports. Moreover, faced with a difficulty in finding real female referees, we used sport science students—although the ability to compare a female and a male cohort and the general consistency in the gender bias findings across Studies 1 and 2 increase our confidence in the validity of the findings.

Another possible limitation concerns a potential confound between referees' age, experience, and expertise in Study 1. Specifically, we tested the influence of refereeing expertise using three groups, one of which was markedly younger and less experienced (junior referees) than the other two (intermediate and national referees). In contrast, other studies in the area of referee expertise tend to use three groups with the same age and experience to isolate the unique influence of expertise (e.g., Williams & Davids, 1995). Future research could apply such designs. It is also important to acknowledge that the finding in Study 1 that expert referees were as prone to gender biases as referees lower in expertise may have been influenced by the methodology we used. Specifically, the situations shown to referees were videotaped from the spectators' gallery and not from the field of play, which is arguably a less realistic way of assessing referees' responses (Ste-Marie, 2003). Having said that, the use of videotaped situations from real matches allowed external validity concerns to be balanced with the highly controlled nature of the studies, which allowed for a greater degree of certainty and precision regarding the nature of the gender bias in decisions. Another possible concern is that the instructions received by referees that they would view situations from three levels of competition for each video mounting (one for male players and one for female players) could allow category calibration to be present explicitly due to the instruction (e.g., Unkelbach & Memmert, 2008). Finally, we did not directly test the direct influence of *variation* in the strength and content of referees' gender stereotype on gender biases. It may take several studies before researchers can pin down all of the "hows, whens, and whys" of the influence of player gender on referees' decision making.

Thus, despite the ambitious nature of our multifactorial research designs, there is clearly scope to expand on it in future research. One possibility worthy of future investigation is that spontaneous decision making in judging a foul may be more closely related to *implicit* than *explicit* cognitions and evaluations. Future studies using implicit measures like the implicit association task (Greenwald, McGhee, & Schwartz, 1998) should explore the role of implicit and explicit gender stereotypes, implicit and explicit attitudes toward gender (e.g., use of the Ambivalent Sexism Inventory), and implicit and explicit motivations to control gender bias (e.g., equality or power goal or values, motivation to control prejudice scale) on referees' decision making. There would also be value in testing whether it is possible through conditioning or imagery to associate female players with high sporting abilities and low need for protection and to examine whether it would subsequently be possible for referees to limit gender bias under more cognitively loaded conditions. Finally, it could be that external factors provide less of a motivation to control gender bias than do internal factors. For example, Maio, Olson, Allen, and Bernard (2001) showed that when individuals reflect by themselves about why equality is important, they are more egalitarian in a discriminatory task than when only the value of equality is activated or researchers gave reasons to participants explaining why equality value is important. For all of these perspectives, a psychology software tool could be useful in generating response formats that reflect more closely the modes and time of response in real team collective sport situations like handball.

## Note

1. In French handball there are four groups of national referees (from Group 4 to Group 1) that can officiate at five national competition levels (from National 3 to first Division). In our national-level sample, we had three Group 1 referees (officiating in First Division), 31 Group 3 referees (officiating in National 1—third division), and 7 Group 4 referees (officiating in National 3 or 2, below the third division). Seven referees explained that they were national referees without identifying their national group. Turning to our intermediate-level sample, there are three referee groups (from R3 to R1) that can officiate at four intermediate (or provincial) competition levels. In our sample, we had 26 R1 referees who can officiate at the highest intermediate (or provincial) competition level. Other referees explained only that they were regional (intermediate or provincial) referees without identifying their specific groups.

## References

Banaji, M.R., Hardin, C., & Rothman, A.J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology*, *65*, 272–281. doi:10.1037/0022-3514.65.2.272

Blair, I.V., Ma, J.E., & Lenton, A.P. (2001). Imagining stereotypes away: The moderation of implicit stereotype through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841. doi:10.1037/0022-3514.81.5.828 PubMed

Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, *5*, 665–692. doi:10.1037/1076-8971.5.3.665

Catteeuw, P., Gilis, B., Jaspers, A., Wagemans, J., & Helsen, W. (2010). Training of perceptual-cognitive skills in offside decision making. *Journal of Sport & Exercise Psychology*, *32*, 845–861. PubMed

Catteeuw, P., Gilis, B., Wagemans, J., & Helsen, W. (2010). Perceptual-cognitive skills in offside decision making: Expertise and training effects. *Journal of Sport & Exercise Psychology*, *32*, 828–844. PubMed

Chaiken, S., Liberman, A., & Eagly, A.H. (1989). Heuristic and systematic information processing within and beyond the

persuasion context. In J.S. Uleman & J.A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

Coulomb-Cabagno, G., Rascle, O., & Souchon, N. (2005). Players' gender and male referees' decisions about aggression in French soccer: A preliminary study. *Sex Roles*, *52*, 547–553. doi:10.1007/s11199-005-3720-z

Deaux, K., & Lafrance, M. (1998). Gender. In D.T. Gilbert, S.T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th ed., pp. 788–827). Boston: McGraw-Hill.

Derks, B., Ellemers, N., Van Laar, C., & De Groot, K. (2011). Do sexist organizational cultures create the Queen Bee? *The British Journal of Social Psychology*, *50*, 519–535. doi:10.1348/014466610X525280 PubMed

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18. doi:10.1037/0022-3514.56.1.5

Eagly, A.H., & Karau, S.J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*, 573–598. doi:10.1037/0033-295X.109.3.573 PubMed

Eagly, A.H., Karau, S.J., & Makhijani, M.G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, *117*, 125–145. doi:10.1037/0033-2909.117.1.125 PubMed

Fiske, S.T., Cuddy, A.J.C., Glick, P., & Xu, J. (2002). A model of (often) mixed stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902. doi:10.1037/0022-3514.82.6.878 PubMed

Fiske, S.T., & Neuberg, S.L. (1990). In M.P. Zanna (Ed.), *Advances in experimental social psychology: Vol. 23. A continuum of impression formation, from category based to individuating processes: Influences of information and motivation on attention and interpretation* (pp. 1–74). San Diego: Academic Press.

Frank, M.G., & Gillovich, 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. doi:10.1037/0022-3514.54.1.74 PubMed

Glick, P., & Fiske, S.T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512. doi:10.1037/0022-3514.70.3.491

Greenwald, A.G., McGhee, D.E., & Schwartz, J.K.L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. doi:10.1037/0022-3514.74.6.1464 PubMed

Hamilton, D.L., & Sherman, J.W. (1994). Stereotypes. In R.S. Wyer, Jr. & T.K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 59–82). Hillsdale, NJ: Erlbaum.

Hancock, D.J., & Ste-Marie, D. (2013). Gaze behaviors and decision making accuracy of higher- and lower-level ice hockey referees. *Psychology of Sport and Exercise*, *14*, 66–71. doi:10.1016/j.psychsport.2012.08.002

Haslam, S.A., & Ryan, M.K. (2008). The road to the glass cliff: Differences in the perceived suitability of men and women for leadership positions in succeeding and failing organizations. *The Leadership Quarterly*, *19*, 530–546. doi:10.1016/j.leaqua.2008.07.011

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604. doi:10.1146/annurev.psych.53.100901.135109 PubMed

Hoffman, C., & Hurst, N. (1990). Gender stereotypes: Perception or rationalization? *Journal of Personality and Social Psychology*, *58*, 197–208. doi:10.1037/0022-3514.58.2.197

Jones, M.V., Paull, G.C., & Erskine, J. (2002). The impact of a team's aggressive reputation on the decisions of association football referees. *Journal of Sports Sciences*, *20*, 991–1000. doi:10.1080/026404102321011751 PubMed

Kiefer, A.K., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes: A prospective study of female college students. *Psychological Science*, *18*, 13–18. doi:10.1111/j.1467-9280.2007.01841.x PubMed

Kippendorf, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). London: Sage.

Klonis, S.C., Plant, A.E., & Devine, P.G. (2005). Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin*, *31*, 1237–1249. doi:10.1177/0146167205275304 PubMed

MacMahon, C., Helsen, W., Starkes, J., Cuypers, K., & Weston, M. (2007). Decision-making skills and deliberate practice in elite association football referees. *Journal of Sports Sciences*, *25*, 65–78. doi:10.1080/02640410600718640 PubMed

Maio, G.R., Olson, J.M., Allen, L., & Bernard, M.M. (2001). Addressing discrepancies between values and behavior: The motivating effect of reasons. *Journal of Experimental Social Psychology*, *37*, 104–117. doi:10.1006/jesp.2000.1436

Mohr, P.B., & Larsen, K. (1998). Ingroup favoritism in umpiring decisions in Australian Football. *The Journal of Social Psychology*, *138*, 495–504. doi:10.1080/00224549809600403

Moskowitz, G.B., Gollwitzer, P.M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology*, *77*, 167–184. doi:10.1037/0022-3514.77.1.167

Nevill, A.M., Balmer, N.J., & Williams, A.M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, *3*, 261–272. doi:10.1016/S1469-0292(01)00033-4

Nosek, B.A., Smyth, F.L., Hansen, J.J., Devos, T., Lindner, N.M., Ranganath, K.A., . . . Banaji, M.R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88. doi:10.1080/10463280701489053

Pendry, L.F. (1998). When the mind is otherwise engaged: Resource depletion and social stereotyping. *European Journal of Social Psychology*, *28*, 293–299. doi:10.1002/(SICI)1099-0992(199803/04)28:2<293::AID-EJSP869>3.0.CO;2-7

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognitive perspective. *Psychology*

*of Sport and Exercise*, *7*, 555–575. doi:10.1016/j.psychsport.2006.03.007

Plessner, H., Schweizer, G., Brand, R., & O'Hare, D. (2009). A multiple-cue learning approach as the basis for understanding and improving soccer referees' decision making. In M. Raab, J. Johnson, & H. Heekeren (Eds.), *Progress in brain research: Mind and motion: The bidirectional link between thought and action* (pp. 151–158). Amsterdam: Elsevier Press.

Rudman, L.A., & Kilianski, S.E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, *26*, 1315–1328. doi:10.1177/0146167200263001

Souchon, N., Cabagno, G., Rascle, O., Traclet, A., Dosseville, F., & Maio, G.R. (2009). Referees' decision making about transgressions: The influence of player gender at the highest national level. *Psychology of Women Quarterly*, *33*, 445–452. doi:10.1111/j.1471-6402.2009.01522.x

Souchon, N., Cabagno, G., Traclet, A., Dosseville, F., Livingstone, A.G., Jones, M., & Maio, G.R. (2010). Referees' decision making and player gender: The moderating role of the type of situation. *Journal of Applied Sport Psychology*, *22*, 1–16. doi:10.1080/10413200903250476

Souchon, N., Cabagno, G., Traclet, A., Trouilloud, D., & Maio, G.R. (2009). Referees' use of heuristics: The moderating impact of standard of competition. *Journal of Sports Sciences*, *27*, 695–700. doi:10.1080/02640410902874729 PubMed

Souchon, N., Coulomb-Cabagno, G., Traclet, A., & Rascle, O. (2004). Referees' decision making in handball and transgressive behaviors: Influence of stereotypes about gender of players? *Sex Roles*, *51*, 445–453. doi:10.1023/B:SERS.0000049233.28353.f0

Souchon, N., Livingstone, A.G., Rascle, O., Cabagno, G., & Maio, G.R. (submitted). Influence of player gender, competition level and game situations on referees' decision making in handball.

Ste-Marie, D.M. (2003). Expertise in sport judges and referees: Circumventing information processing limitations. In J.L. Starkes & A. Ericsson (Eds.), *Expert performance in sport: Advances in research on sport expertise* (pp. 169–190). Champaign, IL: Human Kinetics.

Stone, J., Lynch, C.I., Sjomeling, M., & Darley, J.M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, *77*, 1213–1227. doi:10.1037/0022-3514.77.6.1213

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247. doi:10.1207/s15327957pspr0803_1 PubMed

Tajfel, H. (1981). *Human groups and social categories*. New York: Cambridge University Press.

Tajfel, H., Billig, M., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178. doi:10.1002/ejsp.2420010202

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. doi:10.1126/science.185.4157.1124 PubMed

Unkelbach, C., & Memmert, D. (2008). Game management, context effects, and calibration: The case of yellow cards in soccer. *Journal of Sport & Exercise Psychology*, *30*, 95–109. PubMed

Van Quaquebeke, N., & Giessner, S.R. (2010). How embodied cognitions affect judgments: Height-related attribution bias in football fouls calls. *Journal of Sport & Exercise Psychology*, *32*, 3–22. PubMed

Whissell, R., Lyons, S., Wilkinson, D., & Whissel, C. (1993). National bias in judgments of olympic-level skating. *Perceptual and Motor Skills, 77,* 355–358 10.2466/pms.1993.77.2.355.

Williams, A.M., & Davids, K. (1995). Declarative knowledge in sport: A byproduct of experience or a characteristic of expertise? *Journal of Sport & Exercise Psychology*, *17*, 259–275.

*Manuscript submitted: January 21, 2013*

*Revision accepted: July 12, 2013*

# Appendix

## Question 1:

Referees proportionally return back the ball more frequently to female victims of a defensive transgression than male players, in relation to the number of defensive transgression male player and female player can display in a match.

Female player Highest national level > Male player highest national level

Female player highest regional level > Male player highest intermediate level

Female player highest departmental level > Male player highest local level

Do you agree with these results?

How do you explain them?

## Question 2:

The proportion of "neutralisation" (referees intervene immediately after the defensive transgression on the player in possession of the ball) is more important in the female game than in the male game, in relation to the number of defensive transgression male player and female player can display. Nevertheless, there is no gender difference on this dimension at the highest national level.

Female player Highest national level = Male player highest national level

Female player highest intermediate level > Male player highest intermediate level

Female player highest local level > Male player highest intermediate level

Do you agree with these results?

How do you explain them?

## Question 3:

When players miss their pass after being victim of a defensive transgression, referees tend to return back the ball more frequently to female player than males ones whatever the competition level.

Female player highest national level > Male player highest national level

Female player highest intermediate level > Male player highest intermediate level

Female player highest local level > Male player highest local level

This is always true for miss shot situations.

Female player Highest national level > Male player highest national level

Female player highest intermediate level > Male player highest intermediate level

Female player highest local level > Male player highest local level

Do you agree with these results?

How do you explain them?