

Running head: RACISM IN SOCCER

RACISM IN SOCCER? PERCEPTION OF CHALLENGES OF BLACK AND WHITE
PLAYERS BY WHITE REFEREES, SOCCER PLAYERS, AND FANS¹

PASCAL WAGNER-EGGER, PASCAL GYGAX, AND FARFALLA RIBORDY

University of Fribourg, Switzerland

Wagner-Egger, P., Gygax, P., & Ribordy, F. (2012). Racism in soccer? perception of challenges of black and white players by white referees, soccer players, and fans. *Perceptual & Motor Skills, 114*, 275-289.

¹ Address correspondence to Pascal Wagner-Egger (pascal.wagner@unifr.ch).

Summary.— This experiment investigated challenge evaluations in soccer and their relation to prejudice: more precisely, whether skin colour may influence judgements of soccer tackles. Three groups of participants (soccer players, referees and soccer fans) were asked to evaluate challenges, featuring Black and White players as aggressors and victims in a mixed-design study. Results showed that participants made some differentiations between Black and White players in a challenge evaluation task. Participants were more likely to consider within-group challenges as fouls and were faster to consider challenges made by Black players as fouls. On the other hand, fouls made by White players were seen as more severe. There were no major differences between the participating groups, suggesting that the observed effects were independent of how good players were or whether they were referees or not.

In present times, racism is undoubtedly a major concern in professional sports, and especially in soccer: in European championships, African players suffered regular racist acts committed by spectators as well as other players. Among the most prominent examples are those of Samuel Eto'o in Spain¹, Marc Zoro in Italy², Pascal Chimbona³. The president of the *Fédération Internationale de Football Association* (FIFA), Sepp Blatter, has often voiced his concern about racism in soccer, and has affirmed his willingness to fight against this plague to professional sport.

Racism in soccer is a multifaceted problem that encompasses a broad range of social phenomena such as the sociological and political background of soccer fans, soccer culture (e.g., Back, Crabbe & Solomos, 1999), economic discrimination of players (Szymanski, 2000), crowd dynamics, or social judgement of play. The focus of this present study was on the latter. Several studies in social psychology have addressed social judgement of performance by athletes, coaches, referees, or spectators (for a review, see Plessner & Haar, 2006), but only a few have focused on judgements of aggressive behaviours in soccer (e.g., fouls). The present study primarily investigated the extent to which Black players were discriminated against by fans, referees, and players, when judging challenges. A challenge is a short sequence of play in which a player tries to take the ball of another player. In a soccer game, a challenge may be judged by the referee as an offense (therefore sanctioned) or as a correct intervention.

Social Cognition and Sport

The prototypical intergroup nature of team sports has often been considered as appropriate for the study of social psychological phenomena. Most importantly, evaluation and judgement processes in sport have often been under scrutiny, especially in regard to the different biases that may act upon those processes. For example, Hastorf and Cantril (1954) showed that American soccer fans of two competing teams evaluated infractions of the rules, while watching the same match, in favour of their own team, exemplifying the process by

which people favour the behaviours and perceptions of the members of their own group (i.e., the so-called *ingroup bias*). Snibbe, Kitamaya, Markus, and Suzuki (2003) also found that European and American students, but not Japanese students, evaluated their local university baseball team more positively than the opposing team on personality traits. Numerous other studies have investigated ingroup bias in sport judgements at various levels of social categorization (e.g., home team, nation), on various targets (e.g., referees, judges, spectators), in real competition (home advantage effect, e.g., Nevill & Holder, 1999) or laboratory experiments (e.g., Ansorge & Scheer, 1988; de Fiore & Kramer, 1982; Markman & Hirt, 2002; Mohr & Larsen, 1998; Plessner, 1999; Ste-Marie, 1996; Seltzer & Glass, 1991; Smith, 2003; Sumner & Mobley, 1981; Whissel, Lyons, Wilkinson, & Whissel, 1993).

Stereotypes are another critical source of biased perception in sport judgements. Racial stereotypes, most relevant to this study, have been found to bias judgement of basketball performance (Stone, Perry, & Darley, 1997). Participants listening to a college basketball game judged the same player as having more athletic abilities when he was presented as a Black player, but as displaying more basketball intelligence when he was presented as a White player.

In the present study, although the primary interest was on challenge and not performance judgements made by White participants (i.e., of Black and White soccer players), these judgements were expected to reveal both processes of stereotype based judgement and ingroup bias.

Other studies have specifically focused on the perception of hostile behaviours in sport as well as external factors that may bias such a perception. Frank and Gilovich (1988), for example, found that wearing black uniforms rather than any other colour in team sports leads players to be perceived as behaving in a more aggressive manner. Official statistics of championship games also supported these findings; collective sport teams wearing black uniforms ranked near the top of their leagues in penalties. Others researchers have suggested

that colour perception can be mediated by cultural factors. However, Tiryaki (2005) failed to demonstrate this bias in soccer referees from Turkey, rejecting a linear influence of black uniforms. In a different line of research, others have analysed the influence of the gender of team sports players on referees' decisions about transgressive game behaviours. They consistently found that although men actually displayed more aggressive behaviours than women, the latter were more penalized (Coulomb-Cabagno, Rascle, & Souchon, 2005; Souchon, Coulomb-Cabagno, Traclet, & Rascle, 2004). It is beyond the scope of this paper to explain these findings. However, they document, in conjunction with the other studies presented so far, that the evaluation of performances and challenges is mediated by factors external to the actual behaviour.

As Black individuals repeatedly suffer from being judged as more aggressive than White individuals in violence situations (Harrison & Willis Esqueda, 2001), as well as in their communication style (Ogawa, 1971), and tend to be associated with threat (Correll, Wittenbrink, Park, Judd, & Goyle, 2011), it was expected those prejudices to be likely to affect challenge evaluation in soccer. In addition to these prejudices, seeing Black athletes as more athletic and as having more physical abilities for team sports (Rhode & Butler, 1975; Stone et al., 1997) could lead to an exaggerated aggressive perception of their behaviours. Strong discriminatory biases were hypothesized when White participants evaluate challenges made by Black players, but also when the challenges are made *on* Black players.

In light of recent and increasing expressions of racism in soccer, as well as past literature on discriminations based on skin colour, discriminatory biases were expected towards Black players in different forms. First, it was hypothesized that participants would evaluate more challenges as fouls when made by Black players (i.e., an aggressor effect), and more challenges as fouls when made on White players (i.e., a victim effect). Second, it was hypothesized that the time to make those discriminatory decisions would be faster (e.g., smaller response latencies when deciding that a challenge made by a Black player is a foul).

The rationale for this hypothesis lies in the well established phenomenon in social psychology according to which priming a social category facilitates reactions to stimuli that are congruent with the category's stereotype (e.g., Bargh, 2006). As hostility and danger are part of the Black stereotype (Devine, 1989), shorter latencies to detect hostile behavior (i.e., fault) from Black players were predicted: faster response times are traditionally considered as mirroring the ease by which a process is activated (Correll, Park, Judd, & Wittenbrink, 2007; Wittenbrink, Judd & Park, 1997). Third and finally, challenges made by Black players and those made on White players were expected to be evaluated as more severe than others.

Method

Participants

Forty-three White soccer players (18 players from the Swiss fifth Division, 3 players from the Swiss fourth Division, 2 players from the Swiss third Division, 9 players from the Swiss second Division, 11 players from the Swiss first Division; age $M = 24.7$ years, $SD = 4.20$), 17 White referees of regional level (from the Swiss fifth Division to first Division; age $M = 41.9$ years, $SD = 13.00$), and 22 White soccer fans (10 sports and 12 psychology students interested in soccer but not involved in competitive soccer; age $M = 23.6$ years, $SD = 3.59$) took part in this experiment. The whole sample was therefore composed of 82 participants. Soccer players were mostly recruited by writing letters or e-mails addressed to their soccer club president or coach. Referees were contacted personally by mail or by local referee associations, and students from the Universities of Fribourg, Bern, Geneva and the Swiss Federal School for Sports of Macolin were asked to participate by their professors.

Measures

Challenge Evaluation Task (CET). Four hundred and seventy-five soccer video game sequences, each representing a game situation in which one player attempted to get the ball from another player, were created using a video game console (Xbox 360 ©) and the soccer game FIFA 2005 ©. This game conveniently allowed us to create realistic sequences of game

situations from various camera viewpoints. Recently, Kozlov and Johansen (2010) advocated for the use of video-game-based virtual environments in social psychological research, as a good compromise between experimental control and ecological validity. The sequence duration varied between 1'24" and 1'96". In order to control the colour of the players' skin and shirts, two unrecognizable teams with two shirt colour options (green and white) were chosen. These shirts were chosen to avoid, as much as possible, that participants would associate them to well-known European teams (e.g., blue for Italians) and to avoid any known perceptual bias (e.g., red colour bias, as in Hagemann, Strauss & Leißing, 2008; or black colour bias, as in Frank & Gilovich, 1988).

From the initial 475 sequences, 258 sequences were selected with (1) appropriate image quality, (2) the ethnicity of both players visible (two independent judges assessed this variable), (3) no recognisable areas of the field and (4) the absence of distracters (e.g., other players, fans, adverts, etc...). A subsequent evaluation of these 258 remaining sequences was performed by two independent judges (i.e., expert soccer players). Both of the judges viewed the sequences in the presence of the experimenter and had the possibility to view each sequence several times if deemed necessary. Their task was to rate the quality of each sequence in terms of both the ecological validity of each challenge (i.e., "the situation seemed real") and the relative ambiguity of the challenge (i.e., "the challenge could be sanctioned") on a five-point scale ranging from 0 (the sequence should be excluded) to 4 (the sequence is both realistic and ambiguous).

Sixty-four sequences, all having been evaluated as highly realistic and ambiguous were chosen and grouped into four conditions: 16 sequences of a Black player challenging another Black player, 16 sequences of a Black player challenging a White player, 16 sequences of a White player challenging another White player and 16 sequences of a White player challenging a Black player. Half of the White players had dark hair and half had blond hair. Shirt colours (i.e., white or green) were counterbalanced across all sequences.

Although all sequences were evaluated as ambiguous by independent judges, these judges could have provided biased evaluation (i.e., already assessed some sequences as more severe, depending on the players' skin colour). To address this issue, for each experimental sequence, an *objective severity* score was computed. The purpose of this score was to ensure that possible severity differences in the sequences would be accounted for in the data analyses and thus would not have an impact upon potential condition differences. The detailed procedures of the way that this score was accounted for will be presented below.

The first criterion of the objective severity score was the *primary point of contact*. A score was attributed to each sequence in accordance to the primary contact made by the challenger, the worse the challenge, the higher the score (i.e., 0 if the contact is made to the ball, 1 if both a leg and the ball are touched at the same time, and 2 if the contact is made to the leg alone). The second criterion was the *angle of attack* (i.e., 0 if the tackle was sideways, 1 if the tackle was made at a 45 degree angle, either from ahead of or behind the player, and 2 if the tackle was made from straight ahead or from behind, at no angle). The third criterion was the *nature of the tackle* (i.e., a standing tackle was given 0, whereas a sliding tackle was given 1). The fourth and final criterion was the *resulting position of the challenged player*, somehow less important than the first three criteria, hence the value sequence was smaller (i.e., 0 if the player is still standing after the challenge, 0.5 if the player is struggling to stand after the challenge, and 1 if the player falls on the ground). To compute an *objective severity* score for each sequence, the criteria values were added. Resulting scores therefore ranged from 0 to 6.

The CET was presented on a Macintosh MacBook Pro using Psyscope X Software (Cohen, MacWhinney, Flatt, & Provost, 1993). Responses were collected using the keyboard of the computer, which permits 17 milliseconds accuracy.

Procedure

Participants were told that the study investigated the way people perceive challenges

in soccer. Participants were presented with 64 sequences of challenges, presented one after the other, and, for each sequence, they were instructed to carefully watch the challenge and decide, as quickly and accurately as possible, if they thought the challenge was an actual foul or not. Responses were given by pressing one of two keys (i.e., labelled *yes* or *no*) on the keyboard. Participants' dominant hand was always on the *yes* key. Once a decision was made, a four-point severity scale (1=*weak severity* coloured white, 2=*mild severity* coloured light grey, 3=*strong severity* coloured dark grey and 4=*extreme severity* coloured black) automatically appeared on the screen. Participants had to decide to what extent the challenge was a severe foul. When participants answered *no* in the first task, they just pressed the space bar (labelled *no foul*) on the severity rating. The detailed timing of the task is given in Figure 1. As a control measure, participants were asked after completion of the task whether they noticed that soccer players differ in skin colour in the CET.

[Insert Figure 1 about here]

Analysis

Three general 3 (Group: Players vs. Referees vs. Fans) X 2 (Aggressor: Black vs. White) X 2 (Victim: Black vs. White) mixed ANOVA, considering *Aggressor* and *Victim* as within-factors and *Group* as a between-factor, were carried on the three DVs. According to the first hypothesis, challenges made by Black aggressors should be more often considered as fouls than when committed by White aggressors, and challenges made on White victims should be more often considered as fouls than when committed on Black victims. The same main effects of *Aggressor* and *Victim* were predicted on the severity ratings of the fouls (more severity when the aggressor is Black, and when the victim is White), as stated by hypothesis 3. A main effect of aggressor on response latencies was expected following hypothesis 2: response time should be shorter when the aggressor is Black than when it is White.

Results

All analyses were conducted by controlling for heterogeneity in sequence *objective*

severity, in addition to the fact that the selection of sequences was based on similarity as evaluated by independent judges. Although the way this control was achieved was different for each measurement, it was always based on the *objective severity* score.

Foul Evaluations

Before analysing the results, responses were transformed as followed. Positive responses (i.e., *yes it is a foul*) were coded as 1 and negative responses were coded as 0. For each sequence, positive responses were weighed by dividing them by the *objective severity* score. Each participant's resulting *foul evaluations* score was computed by simply taking the mean score (see Table 1 for overall means). Higher scores meant that participants were more likely to evaluate challenges as fouls, despite their low *objective severity*. Lower scores indicated a tendency to evaluate few challenges as fouls, even when the challenge is *objectively severe*. Hence, the lower the score, the more participants underestimated challenges as fouls, and the higher the score, the more participants overestimated challenges as fouls.

[Insert Table 1 about here]

A general 3 (Group: Players vs. Referees vs. Fans) X 2 (Aggressor: Black vs. White) X 2 (Victim: Black vs. White) mixed ANOVA, considering *Aggressor* and *Victim* as within-factors and *Group* as a between-factor, revealed a main effect of *Aggressor*, $F(1, 79) = 16.08$, $p < .05$, $\eta^2 = .17$, suggesting a greater likelihood for challenges made by White players to be evaluated as fouls than challenges made by Black players. There was also an *Aggressor* by *Group* interaction effect, $F(2, 79) = 3.82$, $p < .05$, $\eta^2 = .04$, suggesting that the previous effect was stronger in the referee group than in the fan group, but absent in the player group (see Table 1). Most interestingly, there was a strong *Aggressor* by *Victim* interaction effect, $F(1, 79) = 30.01$, $p < .05$, $\eta^2 = .27$, indicating that challenges including an aggressor and a victim of the same skin colour were more likely to be considered as a foul than challenges including players of different skin colour. Participants may have considered the latter challenges as

sources of intergroup conflict exacerbation and hence may have treated them more cautiously. Finally, and perhaps logically, a main effect of *Group*, $F(1, 79) = 10.20, p < .05, \eta^2 = .12$, showed that referees were generally more likely to evaluate challenges as fouls ($M = .19, SE = .01$) than did fans ($M = .16, SE = .01$), who in turn were more likely to evaluate challenges as fouls than players did ($M = .14, SE = .01$; all post-hoc comparisons at $p < .05$; see Table 1). There were no other main or interaction effects.

Contrary to our first hypothesis, there were no signs of discriminatory processes towards Black players in the evaluation of challenges. If anything, fans and players were more lenient towards intergroup conflicts, and referees were harsher on challenges made by White players.

Positive Response Latencies

To account for heterogeneity of sequence severity, a somewhat different procedure than in the previous analyses was adopted. The procedure here, inspired by psycholinguistics transformation procedures (such as proposed by Trueswell, Tanenhaus and Garnsey, 1994), accounts for individual difference in response latencies (i.e., irrelevant of any experimental condition) and for the fact that sequences were of different *objective severity* scores. In these analyses, it was also decided to only analyse positive response latencies, as the study was mostly interested in the automatic activation of discriminatory processes, as opposed to the controlled inhibition of those processes, more relevant in our other measures. The transformation procedure was conducted as followed: For each participant, a regression equation of *positive response latencies* against *objective severity* was produced. Computing the slope and the intercept of the regression enabled the calculation of residual response latencies for each participant by subtracting the response latencies predicted by the regression equation from the actual response latencies. Statistical analyses were conducted on these residual positive response latencies. Negative residual latencies meant that responses were faster than expected, and positive ones meant that they were longer. Although relatively

complex, this data transformation addressed differences in sequence severity, as well as variations between participants. Twelve extreme values (3.7%) were replaced by their group mean to ensure normal distributions (Kolmogorov-Smirnov tests). Mean residual latencies are shown in Table 1.

A general 3 (Group: Players vs. Referees vs. Fans) X 2 (Aggressor: Black vs. White) X 2 (Victim: Black vs. White) mixed ANOVA, considering *Aggressor* and *Victim* as within-factors and *Group* as a between-factor, on the residual *positive response latencies* revealed that participants were faster to evaluate challenges as being fouls when the aggressor was Black ($M_{residual} = -63.99$, $SE = 16.10$) than when the aggressor was White ($M_{residual} = 59.21$, $SE = 14.69$), $F(1, 79) = 22.21$, $p < .05$, $\eta^2 = .22$. Although this effect confirmed hypothesis 2 on the automatic discriminatory processes towards Black players, it was qualified by an *Aggressor* by *Victim* interaction effect, $F(1, 68) = 6.98$, $p < .05$, $\eta^2 = .10$, suggesting that the observed *Aggressor* effect was stronger when the victim was White rather than when the victim was Black. In all, participants were slowest when evaluating a foul made by a White player on another White player ($M_{residual} = 113.83$, $SE = 31.32$) but fastest when evaluating a foul made by a Black player on a White player ($M_{residual} = -92.95$, $SE = 30.50$). This tendency would also support a discriminatory process towards Black players (i.e., not only towards aggressors, but also towards victims).

Severity Ratings

Severity ratings were transformed using the same procedure as in the foul evaluations analyses, weighting each severity rating by dividing it by the *objective severity* score. Higher scores thus indicated that participants overestimated the severity of fouls, and lower scores that they underestimated this severity (see Table 1).

A general 3 (Group: Players vs. Referees vs. Fans) X 2 (Aggressor: Black vs. White) X 2 (Victim: Black vs. White) mixed ANOVA, considering *Aggressor* and *Victim* as within-factors and *Group* as a between-factor, on the foul severity ratings (taking into account each

sequence severity score) indicated a main effect of *Aggressor*, $F(1, 79) = 98.16, p < .05$, $\eta^2 = .55$, where fouls made by White players rated as more severe ($M = .62, SE = .01$) than those made by Black players ($M = .53, SE = .01$). This intriguing result, contradicting hypothesis 3, may express a possible reverse prejudice attitude (outgroup favouritism) at a more explicit level. However, there was also a main effect of *Victim*, $F(1, 79) = 16.82, p < .05$, $\eta^2 = .18$, where fouls made on White players rated as more severe ($M = .60, SE = .01$) than those on Black players ($M = .56, SE = .01$), hinting at ingroup favouritism. One could argue that participants, although not rating Black and White players' challenges in the same way, attempted to compensate by giving advantages to both groups, either when victims or aggressors. As in the *foul evaluations* analyses, there was a main effect of *Group*, $F(1, 79) = 6.04, p < .05$, $\eta^2 = .07$, which showed that referees were generally more likely to evaluate challenges as severe ($M = .61, SE = .02$) than did fans ($M = .59, SE = .02$), and that, in turn, Fans were more likely to evaluate challenges as severe than did players ($M = .53, SE = .01$; all post-hoc comparisons at $p < .05$). There were no other main or interaction effects⁴.

The three different CET scores (*foul evaluations*, *residual latencies* and *severity ratings*) displayed very low inter-correlation (two significant correlations out of 48), demonstrating on a cognitive level that the three different tasks were accessing different levels of cognitive processing and on a methodological level that the effects found were independent of a sequence difference artefact.

Discussion

In this study, participants (i.e., soccer referees, players and fans) were shown different situation sequences, in which either a Black player or a White player made a challenge on a Black or a White player. Participants had to decide, as fast as possible, if they thought the challenge was an actual foul, and if so, they had to rate the severity of the foul. Overall, although strong discriminatory processes towards Black players were expected, these were not always apparent. Nevertheless, irrelevant of the group they were in, participants

seemed to (implicitly) distinguish between Black and White players when evaluating the challenges.

First, although to a lesser extent by referees, participants were applying stricter rules when the aggressor and the victim were of the same skin colour. This might be considered as a sign of explicit caution, as in real soccer situations, and it may be wise not to exacerbate possible inter-ethnic conflicts by constantly judging inter-ethnic challenges as being against the rules. Referees specifically seemed to express signs of caution when challenges were made by Black players on White victims, but also, more generally and similarly to the other groups, when Black players were making the challenges (i.e., referees were less likely to evaluate challenges made by Black players as fouls). At an explicit level, this general effect may mirror the desire not to appear discriminatory towards Black players. This is only a tentative explanation, but overall, we do not believe these results reflect any true discriminatory process. Second, and at a more implicit level (i.e., response latencies are more difficult to control), our participants were generally faster when deciding that a challenge by a Black player was a foul and even more so when the victim was a White player.

In a series of studies, Correll and colleagues designed a laboratory task aiming at simulating real life and crucial decisions of police officers, the decision to shoot or not potentially dangerous human targets. The aim of the task was to decide as rapidly as possible to shoot by pressing on a key when a human target that appeared on a computer screen was armed, and not doing so by pressing another key if the target is unarmed. The main independent variable in Correll and colleagues studies was the skin colour of the target, either Black or White. Crucially, they did not constantly found a race-based discriminatory bias, which would consist in more hits and false alarms for Black targets (Correll, Park, Judd, & Wittenbrink, 2002; Correll, Urland, & Ito, 2006; Correll, Park, Judd, Wittenbrink, Sadler, & Keese, 2007; however, Correll, Park, Judd, & Wittenbrink, 2007, using a strong priming procedure, and more recently, Correll, Wittenbrink Park, Judd, & Goyle, 2011, and Ma &

Correll, 2011 did find such discriminatory bias). By contrast, they repeatedly show regarding reaction times that their participants were faster to shoot an armed Black target than an armed White target. As in the present study, discriminatory biases were more apparent in response latencies than in explicit behavior.

When considering *severity ratings*, a somehow different pattern from the two other CET measures was apparent. An initial argument would be that as challenges on Black players were always rated as less severe than those on White players, it signalled some level of discriminatory processes. However, as fouls by White players were always rated as more severe than those made by Black players, it also signalled a reverse discriminatory process. In this task, two processes could be at stake: A discriminatory process, rather difficult to control, as reflected in the *response latencies* and in the *severity ratings* (of fouls made on Black players), by which Black players are discriminated against, and a second process, compensatory in nature, by which fouls by White players were rated as more severe. In essence, participants have conflicting sources of information which result in differential treatments of White and Black players, at times discriminatory to Black players, and at times to White players. At least, contrary to the *response latencies*, which are more difficult to control, participants in both explicit measures of the CET showed some level of control, hinting that even if discriminatory processes may be activated, they can be compensated.

Although this latter explanation might be suited to *severity ratings*, an alternative explanation could also be given. If, stereotypically, Black players are generally perceived as being more aggressive (e.g., Black people generally perceived as more *aggressive* and *violent* when involved in interpersonal violence, Harrison & Willis Esqueda, 2001), participants may, most of the time, *expect* them to foul. As a result, any perceived foul involving a Black player may be seen as *common* and as such, less severe. As they would be unexpected, fouls by and on White players may therefore be perceived as more severe. This might even explain some differences in *response latencies*. If a foul is expected, it should normally be assessed as such

more rapidly. This alternative explanation raises an interesting issue of expectancy and would certainly need further investigation.

Still, interestingly, there were differences between the three participating groups. Overall, referees quite naturally judged more challenges as being fouls and evaluated them as being more severe than did the two other groups, and fans judged more challenges as being fouls and evaluated them as being more severe than players did. This may be seen as an effect of closeness to action, as players are themselves regularly engaged in challenges during soccer matches. Overall, the pattern of results were very similar among the groups, indicating that the mechanisms under scrutiny in our CET measures are very stable and are shared by most people closely interested in soccer. The only differences were that referees showed more. For practical reasons (i.e., security) extreme soccer fans (i.e., hooligans) were not tested.

Although the results raise some interesting issues, several aspects of the material have to be critically considered. First, video game sequences were chosen so as to avoid (1) the cumbersome endeavour of having players acting out the situations and (2) using material in which players may have been too easily recognisable. Although these seem like fair arguments, video game situations may not entirely capture the true feeling of a real game situation. A second issue that may be considered as problematic in the material is the fact that the study did not use a fully repeated measure design. The sequences, although highly comparable, were different across conditions. In one or more conditions, the *true severity* (i.e., that someone could argue to be different from *objective severity*) of the sequences may have unexpectedly biased participants' responses. In other words, results may mirror true sequence differences, and not top-down decision processes, as is suggested. In light of the results that were obtained, this possibility is highly unlikely. For the sake of the argument, consider each of the three dependent variables as an actual mirror of uneven choice of sequences. In this line of thinking, the results from the *foul evaluations* analyses would mainly suggest that sequences in which challenges included an aggressor and a victim of the same skin colour

were more severe (i.e., interaction effect). Taken alone, it could actually be argued that these results might be due to a confounding *true severity* factor. *Positive response latencies* analyses suggest a completely different pattern, namely that sequences in which challenges are made by Black players were responded to faster, a result that was interpreted as discriminatory towards Black players (i.e., main effect of aggressor), and even more so for those made on White players. A confounding *true severity* factor argument becomes even more difficult to sustain with the evidence from the *Severity ratings* analyses suggesting that only sequences in which challenges are made by White players are of higher severity (i.e., main effect of aggressor). In all, the results indicated that the sequences were processed mostly in a top-down fashion, and that a possible confounding *true severity* factor is quite inconceivable in this present study.

The study was also only exploratory in nature, in the sense that (1) it did not provide any data on the possible circumstances in which those effects could be attenuated, and (2) it did not take any other real behavioural measure than pressing a button to say whether a challenge was or was not a foul (i.e., similarly to blowing a whistle). It may be interesting to assess whether the fact that after seeing a challenge, players act differently towards the other team, as a function of the skin colour of the aggressor. It may be fruitful investigate different ways to attenuate any categorisation in soccer based on skin colour – on adult players as well as junior players. In conclusion, although the study found some evidence of discrimination signals in the quasi-experimental task (as participants were not randomly assigned in groups), those were not always against Black players: Thus, differentiation judgements in soccer based on skin colour may not be a black or white judgement.

References

- Ansorge, C. J., & Scheer, J. K. (1988) International bias detected in judging gymnastic competition at the 1984 Olympic Games. *Research Quarterly for Exercise and Sport*, 59, 103–107.
- Back, L., Crabbe, T. & Solomos, J. (1999) Beyond the racist/hooligan couplet: Race, social theory and football culture. *British Journal of Sociology*, 50, 419-442.
- Bargh, J. A. (2006) What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology*, 36, 147-168.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993) PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257-271.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002) The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83, 1314-1329.
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007) The influence of stereotype on decisions to shoot. *European Journal of Social Psychology*, 37, 1102-1117.
- Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007) Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, 92(6), 1006-1023.
- Correll, J., Urland, G.R., & Ito, T.A. (2006) Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, 42, 120-128.
- Correll, J., Wittenbrink, B., Park, B., Judd, C. M., & Goyle, A. (2011). Dangerous enough: Moderating racial bias with contextual threat cues. *Journal of Experimental Social Psychology*, 47, 184–189.

- 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.
- De Fiore, R., & Kramer, T. J. (1982) The effect of team affiliation on perception in sports. *International Journal of Sport Psychology, 13*, 43-49.
- Devine, P. G. (1989) Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5-18.
- Frank, M. G., & Gilovich, T. (1988) The dark side of self- and social perception: Black uniforms and aggression in professional sports. *Journal of Personality and Social Psychology, 54*, 74-85.
- Hagemann, N., Strauß, B. & Leißing, J. (2008) When the referee sees red... *Psychological Science, 19*, 769-771.
- Harrison, L. A., & Willis Esqueda, C. (2001) Race stereotypes and perceptions about Black males involved in interpersonal violence. *Journal of African American Men, 5*, 81-92.
- Hastorf, A. H., & Cantril, H. (1954) They saw a game: A case study. *Journal of Abnormal and Social Psychology, 49*, 129-134.
- Kozlov, M. D., & Johansen, M. K. (2010). Real behavior in virtual environments: Psychology experiments in a simple virtual-reality paradigm using video games. *Cyberpsychology, Behavior, and Social Networking, 13*, 711-714.
- Ma, D. S. & Correll, J. (2011). Target prototypicality moderates racial bias in the decision to shoot. *Journal of Experimental Social Psychology, 47*, 391-396.
- Markman, K. D., & Hirt, E. R. (2002) Social prediction and the "allegiance bias". *Social Cognition, 20*, 58-86.
- Mohr, P. B., & Larsen, K. (1998) Ingroup favoritism in umpiring decisions in Australian Football. *The Journal of Social Psychology, 138*, 495-504.
- Nevill, A. M., & Holder, R. L. (1999) Home advantage in sport: An overview of studies on

- the advantage of playing at home. *Sports Medicine*, 28, 221-236.
- Ogawa, D. M. (1971) Small-group communication stereotypes of Black Americans. *Journal of Black Studies*, 1, 273-281.
- Plessner, H. (1999) Expectation biases in gymnastics judging. *Journal of Sport and Exercise Psychology*, 21, 131-144.
- Plessner, H., & Haar, T. (2006) Sports performance judgements from a social cognitive perspective. *Psychology of Sport and Exercise*, 7, 555-575.
- Rhode, L., & Butler, J. S. (1975) Sport and racism: A contribution to theory building in race relations. *Social Science Quarterly*, 55, 919-925.
- Seltzer, R., & Glass, W. (1991) International politics and judging in Olympic skating events. *Journal of Sport Behavior*, 14, 189-200.
- Smith, D. R. (2003) The home advantage revisited. *Journal of Sport & Social Issues*, 27, 346-371.
- Snibbe, A. C., Kitayama, S., Markus, H. R., & Suzuki, T. (2003) They saw a game: A Japanese and American (football) field study. *Journal of Cross-Cultural Psychology*, 34, 581-595.
- 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.
- Ste-Marie, D. (1996) International bias in gymnastic judging: Conscious or unconscious influences ? *Perceptual and Motor Skills*, 83, 963-975.
- Stone, J., Perry, Z. W., & Darley, J. M. (1997) "White Men Can't Jump": Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology*, 19, 291-306.
- Sumner, J., & Mobley, M. (1981) Are cricket umpires biased ? *New Scientist*, 91, 29-31.
- Szymanski, S. (2000) A market test for discrimination in the English professional leagues.

Journal of Political Economy, 108, 590-603.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994) Semantic influences on parsing - use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285-318.

Tiryaki, M. S. (2005) Assessing whether black uniforms affect the decisions of Turkish soccer referees: Is finding of Frank and Gilovich's study valid for Turkish culture ? *Perceptual and Motor Skills*, 100, 51-57.

Whissel, R., Lyons, S., Wilkinson, D., & Whissel, C. (1993) National bias in judgements of Olympic-level skating. *Perceptual and Motor Skills*, 77, 355-358.

Wittenbrink, B., Judd, C. M., & Park, B. (1997) Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, 72, 262-274.

Author's Note

This research was supported by a research grant from the Swiss Federal Council of Sports (FCS) to the first and second author.

Footnotes

¹BBC Sport 26 February 2006, retrieved from
<http://news.bbc.co.uk/sport2/hi/football/europe/4751876.stm>

²BBC Sport 27 November 2005, retrieved from
<http://news.bbc.co.uk/sport2/hi/football/africa/4476412.stm>

³L'Equipe, 15 november 2004, retrieved from
http://www.lequipe.fr/Football/breves2004/20041115_155946Dev.html

⁴The control variable which measured whether participants had noticed that challenges involved Black and White players was entered in the ANOVAs. Overall, the same results were obtained. Concerning the *mean residual latencies*, an additional interaction effect between this *control* variable and the *aggressor* and *victim* variables was obtained, indicating that the observed effect was stronger when participants reported that they had identified Black and White players. A second additional interaction effect *control* x *group* x *aggressor* x *victim* for the number of fouls identified, showed that the observed "caution" effect still held in most subgroups, except for the referees who did not notice ethnic differences and players who did notice these differences. Both subgroups identified more fouls when the victim was Black.

Table 1. Mean scores (Standard Errors) for the Challenge Evaluation Task.

DV	Group	Black Aggressor		White Aggressor	
		Black Victim	White Victim	Black Victim	White Victim
Foul evaluation score	Players	.16 (.006)	.13 (.007)	.14 (.007)	.15 (.007)
	Referees	.19 (.009)	.16 (.011)	.20 (.011)	.20 (.011)
	Fans	.16 (.008)	.14 (.010)	.15 (.010)	.18 (.010)
Mean residual times of positive responses (ms)	Players	-46.3 (22.1)	-63.3 (39.1)	-8.6 (34.4)	70.6 (40.1)
	Referees	-19.7 (35.1)	-134.2 (62.1)	-29.0 (54.7)	142.5 (63.8)
	Fans	-39.1 (30.8)	-81.4 (54.6)	51.3 (48.1)	128.4 (56.1)
Severity	Players	.48 (.014)	.51 (.018)	.56 (.017)	.58 (.019)
	Referees	.53 (.022)	.59 (.028)	.64 (.027)	.69 (.031)
	Fans	.52 (.020)	.57 (.025)	.61 (.024)	.66 (.027)

Figure 1. Detailed timing of challenge evaluation task (CET).

