Journal of Sport & Exercise Psychology, 2010, 32, 483-498

Crowd Noise as a Cue in Referee Decisions Contributes to the Home Advantage

Christian Unkelbach[1] and Daniel Memmert[2]
[1]University of Heidelberg; [2]German Sport University Cologne

The home advantage is one of the best established phenomena in sports (Courneya & Carron, 1992), and crowd noise has been suggested as one of its determinants (Nevill & Holder, 1999). However, the psychological processes that mediate crowd noise influence and its contribution to the home advantage are still unclear. We propose that crowd noise correlates with the criteria referees have to judge. As crowd noise is a valid cue, referee decisions are strongly influenced by crowd noise. Yet, when audiences are not impartial, a home advantage arises. Using soccer as an exemplar, we show the relevance of this influence in predicting outcomes of real games via a database analysis. Then we experimentally demonstrate the influence of crowd noise on referees' yellow cards decisions in soccer. Finally, we discuss why the focus on referee decisions is useful, and how more experimental research could benefit investigations of the home advantage.

Keywords: sport psychology, home advantage, referee decisions, cue learning

Sport teams as well as individual athletes performing "at home" have higher success rates than teams or athletes performing "away." This home advantage is one of the best established phenomena in sports and offers a highly interesting research field for psychology in general and sport psychology in particular (for reviews, see Courneya & Carron, 1992; Carron, Loughhead, & Bray, 2005). Next to other factors (e.g., familiarity, territoriality, travel fatigue), crowd noise has been suggested as one of the major determinants of the home advantage (e.g., Clarke & Norman, 1995; Nevill & Holder, 1999; Pollard, 1986). However, the psychological processes that mediate the impact of crowd noise on performance outcomes are still unclear and sometimes contradictory (e.g., compare Wallace, Baumeister, & Vohs, 2005; Pollard & Pollard, 2005; and Holder & Nevill, 1997).

While many studies investigated the crowd influences on individual athletes (see the review by Carron et al., 2005), we will focus on crowd influences on judges and referees (Mascarenhas, O'Hare, & Plessner, 2006). We believe this focus is useful for two reasons: First, judges and referees have a pivotal role in determining

Christian Unkelbach is with the Psychologisches Institut, Universität Heidelberg, Heidelberg, Germany. Daniel Memmert is with the Institut für Bewegungswissenschaft, Deutsche Sporthochschule Köln, Köln, Germany.

performance and competition outcomes; if they are influenced by the presence of crowds and crowd noise, we have identified a strong contribution to the variance in the home advantage. Second, previous studies showed that home advantages and thereby possible crowd noise effects are reduced or absent when performances are scored objectively rather than subjectively (e.g., weight lifting vs. figure skating; see below, Nevill & Holder, 1999), again demonstrating the importance of referees and judges. In other words, we believe that crowd noise contributes to the home advantage by influencing referees' decisions and judges' verdicts.

How might crowd noise influence judgments and decisions in sport performances? Building on a Brunswikian perspective (Brunswik, 1957), we proposed that referees use crowd noise as a proximal cue to judge distal criteria (e.g., How hard was a foul? or How aesthetically pleasing was a performance?). Brunswik's approach is best illustrated with depth perception: Although images on the retina are strictly two dimensional, humans learn to infer the distal criterion of depth from proximal cues, such as feature overlap or disparities between the images on both eyes. For example, overlap correlates perfectly with depth, because objects that are closer to the perceiver will hide more distant objects. Similarly, referees might use crowd noise because they learn the correlation between crowd noise and specific criteria in a given situation. For judgments in sports, this cue-learning hypothesis is easily illustrated: For example, when aesthetic performance is judged, good performances will elicit more cheering from the audience than bad performances. Similarly, in team sports, clear fouls elicit more distinct responses from the audience than less clear fouls. Even more, any foul elicits more crowd responses than no foul. This leads to a substantial correlation between crowd noise and the criterion to be judged. And just as people learn to use depth cues in perception to judge depth, referees might learn to use crowd noise in referring to judge the respective criteria (performance, fouls, etc.). Because home performers in figure skating receive greater overall responses from the crowd, they should receive higher ratings, and because the home crowd reacts more strongly toward fouls committed against the home team, more free kicks are awarded to the home team, and the away team is punished more frequently for playing rough. Consequently, a distinct and predictable home advantage arises. We will discuss this cue-learning hypothesis in more depth below, but first we examine some of the available evidence on the contribution of crowd noise on the home advantage.

A Cursory Literature Review

Consistent with the suggested importance of referee decisions for the home advantage, there is a greater home advantage for sports in which subjective referee decisions influence a competition's outcome (e.g., scoring in boxing, figure skating; Balmer, Nevill, & Lane, 2005), compared with sports in which performance is measured objectively (e.g., weightlifting, short-track speed skating; Balmer, Nevill, & Williams, 2003). The importance of subjectivity is strongly highlighted in the study by Balmer and colleagues (2005): When boxing competition between evenly matched fighters ended in a knockout (an objective criterion), the probability of a "home" win was only .57; however, when the match was decided by technical knockouts or points decisions (i.e., subjective criteria), the probability of a home win increased to .66 and .74, respectively. Thus, they showed greater home advantages within a sport due to increased importance of subjective judgments. Similarly, Balmer, Nevill, and Williams (2001) investigated the home advantage for the Olympic Winter games from 1908 to 1998; across all events, they found a home advantage. This advantage, however, was moderated by disciplines. Disciplines with objective performance measures (e.g., speed skating) showed less of a home advantage. These data support Holder and Nevill's (1997) conclusion that the home advantage is absent when performance is scored objectively, without subjective input from judges or referees.

Consistent with the suggested impact of the crowd, Schwartz and Barsky's (1977) classic study found a correlation between the home advantage and crowd density (i.e., percentage used of available venue spectator capacity). Similar results were reported by Clarke and Norman (1995), as well as Agnew and Carron (1994). Nevill, Newell, and Gale (1996) even found an absolute crowd effect on the home advantage in English and Scottish soccer. However, there are also many researchers who report failures to find crowd effects on the home advantage (e.g., Strauss, 2002), or even performance impairments due to crowd influences (Baumeister & Steinhilber, 1984).

These are only a few prominent flashlights of research on how crowd noise and refereeing decisions impact home advantages. However, all of the studies cited above use archival data to test the impact of crowd noise on performance and referee decisions. Although statistical models and methods have improved considerably and have become much more refined over the years (e.g., from Schwartz & Barsky, 1977, to Pollard & Pollard, 2005), archival analyses can almost never show causal relationships. To show cause-and-effect relations, experiments are necessary; yet, to our knowledge, there are only two experimental studies that investigated the impact of crowd noise on refereeing decisions.

First, Nevill, Balmer, and Williams (1999) presented 52 challenge scenes from one European Champions League soccer match to eleven experienced referees in the game's chronological sequence. Half of the challenges were initiated by a home team player and half were initiated by an away team player. The referees' task was to call a foul or not. Six referees judged the scenes without sound, and five with the original sound from the game. When sound and thus, crowd noise was present, referees called more challenges by the away team players; this effect disappeared when no crowd noise was present. The authors concluded that when in doubt, judges "refereed to the crowd for guidance" (p. 1416).

Second, Nevill, Balmer, and Williams (2002) presented 47 challenge scenes from one English Premier League soccer match to 40 referees of varying expertise level; 18 referees watched the challenges without sound and 22 with the original sound from the stadium. Their task was to classify the challenge as a home team player foul, an away team player foul, no foul at all, or uncertain. When referees could hear the crowd noise, they called *fewer* fouls on the home team compared with the away team. This effect was moderated by referees' expertise; more experienced referees were less influenced by crowd noise.

Although both experiments are highly informative, they have some limitations; first, both studies used a sound/no sound manipulation, which allows the reinterpretation of watching a game under "natural conditions" (i.e., with sound) and "unnatural conditions" (i.e., without sound). This is a proper way to maximize the variation on the independent variable, but it is a slight weakness of both designs and it prohibits the direct conclusion that crowd noise influences referee decisions in favor of the home team (i.e., by the way of *more* noise in favor of the home team). Rather, it tests the hypothesis that home crowd noise in comparison with no home crowd noise biases decision toward the home team. Second, and more importantly, both studies used only challenge incidents from one game each (for example, Liverpool vs. Leicester City, in Nevill et al., 2002). In terms of stimulus sampling, this could be seen as a serious flaw, as it might be something about these particular games or these particular crowds that causes the observed effects (cf. Wells & Windschitl, 1999). Sutter and Kocher (2004, p. 463) already noted that it is impossible in this design to distinguish a "home bias" from a "Liverpool bias." The problem is highlighted by the differential outcomes of otherwise very similar experiments. The first study shows an increase of fouls called on the away team, while the latter shows an decrease of fouls called on the home team. Nevill and colleagues (2002) tried to reconcile their contradictory findings by introducing the idea that referees tried avoiding a "bad call" against the home team, or in general, they try to avoid unpopular decisions.

The hypothesis presented by Nevill and colleagues (2002) is a motivational one because referees *want* to avoid displeasing the crowd, and more so when there is crowd noise; thus, referees call fewer fouls on the home team when crowd noise is present. This idea is substantiated by data from Balmer and colleagues (2007), who showed that crowd noise is associated with increased anxiety and mental effort in referees; therefore, referees tended to cope with such situations with more popular decisions for the home team. Sutter and Kocher (2004) present a similar motivational account, based on the idea that referees are informed agents who want to balance two different goals: being impartial to their employers (i.e., the soccer governing institutions) and pleasing the crowd. As the crowd has the more immediate influence, a home bias arises.

The problem with such motivational explanations is that they also allow the opposite outcome. Just because referees feel bullied by the home crowd, they might deliberately act against home teams/athletes (as predicted by reactance theory, Miron & Brehm, 2006). Further, motivational influences have difficulties explaining results in experimental laboratory settings. For example, in the experiment by Nevill and colleagues (2002), the question arises why referees should try to appease a crowd that is not present at all—one needs the additional assumption that referees indeed act "as if in the game."

In contrast, the present cue hypothesis is a purely cognitive concept that explains biased decisions due to crowd noise as a simple case of cue learning (i.e., it only matters whether the cue is present or not), which allows testable predictions independent of the specific settings (i.e., laboratory vs. real games). However, motivational explanations are still a necessary construct because there are findings and results that our pure cue-learning hypothesis cannot explain; for example, the data by Nevill et al. (2002), or effects of crowds on biased allowed extra time in soccer (e.g., Dohmen, 2008). We will return to this issue in the general discussion.

Theoretical Background: The Case of Cue Learning

Our hypothesis is built on a Brunswikian perspective; according to Brunswik (1957), there are many properties to which people have no direct sensory access.

To assess these *distal* properties, people rely on *proximal* cues. Again, the most prominent case is depth perception. People's visual perception does not assess depth directly; rather, people infer depth from cues provided by the two two-dimensional images on the retina. The most important cues to infer depth are overlap (objects that hide other objects are closer to the perceiver), motion parallax (when the perceiver moves, objects that are closer seem to move faster than faraway objects), and texture gradients (distant objects have denser textures than close objects). And infants learn by haptic feedback, while trying to reach for objects, such that if one object overlaps another object, the former is closer than the latter. In other words, people learn the correlation between a cue (here: overlap) and a criterion (here: distance). This cue usage is often so over-learned that in most cases, people are not even aware of using these cues at all.

This Brunswikian perspective, the need to use proximal cues to assess otherwise inaccessible properties, is by no means restricted to basic perceptual processes. For example, recipients of persuasive speeches are more convinced if the message is accompanied by favorable audience reactions (Axsom, Yates, & Chaiken, 1987). These authors also framed their results as such that people use audience reactions as a cue in their judgments. Similarly, people laugh longer at a funny video clips when another laughing person is present, again showing how people use an available cue to evaluate a distal criterion (Devereux & Ginsburg, 2001).

The implications for judgment and decision making in sports are immediately obvious: If referees have to judge aesthetic performance in figure skating, they cannot access this criterion directly, but must infer it from multiple cues that are integrated into a judgment. If referees have to judge the severity of a challenge in soccer, they also have to integrate multiple cues into a final judgment (Mascarenhas et al., 2006; Plessner, Schweizer, Brand, & O'Hare, 2009). And because they learn that crowd noise correlates with the severity of fouls (or the excellence of performance), they use crowd noise as an additional cue in their judgment processes, similar to cue learning in perception (e.g., Jacobs, 2002), memory judgments (Unkelbach, 2006), or decision making (e.g., Evans, Clibbens, Cattani, Harris, & Dennis, 2003).

This Brunswikian cue-learning approach shares some properties with the idea that referees use crowd noise as a "judgmental heuristic." However, the notion of a heuristic carries three problems: First, as defined in the judgment and decision literature, heuristics are normally employed when people are not motivated or accountable for their decision, as a mental effort–saving device (e.g., Chaiken, 1980; Chaiken & Trope, 1999). This does not fit with most referees' situations, although the time pressures that referees experience are often used as manipulations to force people into a heuristic judgment mode. Second, heuristics are conceptualized as available "rules of thumb"; in contrast, many authors attribute the influence on referees to processes that are not accessible to awareness (e.g., Nevill and Holder, 1999, p. 232; Sutter & Kocher, 2004, p. 468). And third, heuristics do not explain how they are acquired, whereas the cue-learning approach specifies a perception-like process, based on feedback learning (e.g., Jacobs, 2002). Thus, we believe the idea of a cue-learning process is superior to the idea of a heuristic judgment process; it builds on well-established models in perception and psychophysics, and is more parsimonious in its assumptions. However, more recently, the use of cues has been called a "heuristic" as well (e.g., Hertwig, Herzog, Schooler, & Reimer, 2008). In that sense, our idea that referees use crowd noise as a cue is indeed a judgmental heuristic, but then the label no longer has implications for the underlying processes.

In the remainder of this section, we investigate the Brunswikian cue approach in German soccer, but the present hypothesis applies to all sports in which subjective referee judgments and decisions influence competition outcomes. Our dependent variable of interest is awarding a yellow card for committed fouls. It is important to keep in mind that the case of yellow cards is only one example of a way to learn how crowd noise contributes to the home advantage via referee decisions—the same logic is applicable to virtually all subjective judgments and decisions in the sports domain.

Yellow cards have been implemented in soccer since the 1970 World Cup as an official warning sign for rough and dangerous fouls and unsportsmanlike behavior; as such, they have become one of the most important instruments to regulate soccer games. They are ideally suited to study judgment and decision processes because referees have a great deal of freedom with when to award such warnings or not (e.g., Unkelbach & Memmert, 2008). The probability to award a yellow card should increase with the roughness of a given foul. And as crowd noise should also be a direct function of how rough a foul was, referees should learn the correlation between crowd noise and foul roughness. Hence, referees should award yellow cards with higher probability in the given high crowd noise compared with low crowd noise.

In Study 1, we investigate the impact of crowd noise on yellow card decisions in soccer, using data from the Bundesliga, the highest soccer league in Germany. We will show that referees award more yellow cards to the away team, an effect that is amplified when crowd density is high (see above; Schwartz & Barsky, 1977), and when games are played in "pure" soccer stadiums compared with stadiums with a running track (i.e., when the crowd is closer to the field, and thus, the referee). Yet, as we have argued, archival data does not allow causal conclusions. Thus, Study 2 will present an experiment that avoids some of the design problems of the experimental studies so far and corroborate the effects we found in the databank analysis.

Study 1: A Databank Analysis Testing the Influence of Crowd Noise on Yellow Cards in Soccer

Our hypothesis predicts that more crowd noise should lead to more yellow cards against the away team, and consequently, a home advantage. To test this, we analyzed five seasons of the German Bundesliga, the highest soccer league in Germany (1997/98 through 2001/2002).

Materials and Indices

The German Bundesliga has 18 teams, which play a first and second leg per season; thus, the data set included 1530 games. We defined the home advantage as the differences in goals scored by the home and away team (i.e., goals scored by the home minus goals scored by the away team; for example, as done by Boyko, Boyko, and Boyko, 2007). Similarly, we defined a yellow card effect as the difference between yellow cards awarded against the home and away team (i.e., yellow cards awarded against the away minus yellow cards awarded against the home team). We included only first yellow cards in the analysis; second yellow cards against a player (i.e., yellow-red cards and thereby send-offs) were not included (cf. Downward & Jones, 2007). When these indices are positive, they testify to a home advantage. For crowd noise, we used crowd density as an index, that is, the percentage used of a stadium's absolute visitor capacity.[1] Additionally, the architecture of the different venues in which all these games took place allows for a more refined test of the hypothesis that larger crowds lead to more yellow cards against the away team. In Germany, there are two kinds of stadiums: First, there are all-purpose venues that are characterized by a track and field lane that separates the pitch from the audience. Second, there are "pure" stadiums, in which the crowd is not separated from the pitch. Consequently, games in pure stadiums should show amplified crowd effects because crowd noise is more directly transmitted to referees and players.

Study 1 Results

First, we checked whether a home advantage exists at all in this data set. On average, home teams scored 1.72 (SD = 1.36) goals and away teams only 1.17 (SD = 1.13) goals, $t(1529) = 11.98$, $p < .001$.[2] This result is consistent with the data reported by Clarke and Norman (1995), who also found a home advantage of about half a goal in English soccer. However, this home advantage did not correlate with crowd density (again, visitors number relative to stadium size), $r(1530) = -.006$, *ns*, whereas other authors found correlations of crowd density with outcomes (e.g., Schwartz & Barsky, 1977; Agnew & Carron, 1994).

Note, however, that we did not predict a direct influence of crowd noise on competition outcomes. To support our hypothesis, there should be a difference in yellow card frequency for the home and the away team. Overall, 6489 yellow cards were awarded. On average, home teams were awarded 1.89 (SD = 1.19) yellow cards, while the away teams were awarded 2.35 (SD = 1.27) yellow cards, $t(1529) = 11.10$, $p < .001$. The effect size is astonishing given the mean of yellow cards awarded per game (i.e., 4.24). Most importantly, this difference in yellow cards correlated significantly with crowd density, $r(1530) = .134$, $p < .001$. And finally, the difference in yellow cards correlated significantly with the difference in goals, our index for the home advantage, $r(1530) = .095$, $p < .001$.

Pure Soccer vs. Track and Field Stadiums

As discussed, the effects of crowd density and therefore crowd noise should be amplified in pure soccer stadiums due to the closer proximity of crowd and referees. Our sample included 543 games in pure and 987 games in nonpure stadiums, which allowed testing this hypothesis. Indeed, in pure stadiums, the difference in yellow cards awarded in favor of the home time was $M_{pure} = 0.66$, whereas in stadiums with a track-and-field lane, the difference was reduced $M_{nonpure} = 0.359$, $t(1084) = 3.42$, $p < .001$ (degrees of freedom are corrected for unequal variances). The correlation between crowd density and yellow cards also shows the expected effect. In pure stadiums, the correlation is larger, $r(543) = .140$, $p < .001$, than in nonpure stadiums, $r(987) = .093$, $p < .005$; the difference in correlations, however, is not significant. Further, the distinction between pure and nonpure stadiums also explains the lack of a direct relation between crowd density and goal differences for the home and away teams. For games in pure stadiums, crowd density correlated significantly with goal difference, $r(543) = .120$, $p < .005$. For games in stadiums with a track-and-field lane, no such correlation existed, $r(987) = -.025$, *ns*. This leads to the nonsignificant overall correlation of crowd density and home advantage reported above.

Differential Effects for Home and Away Teams

Nevill and colleagues (2002) argued, based on their experimental results, that the dominant effect of crowd noise is to reduce the number of fouls called against the home team. This contradicts the cue-learning approach and the data presented by Nevill and colleagues (1999). The present archival data analysis allows testing these competing predictions; if Nevill and colleagues (2002) are correct, then crowd density should *reduce* the number of yellow cards awarded to the home team (i.e., a negative correlation). The cue-learning approach predicts that crowd density should only *increase* the number of yellow cards awarded to the away team (i.e., a positive correlation). In the present data, the answer is clear cut: Across games, crowd density did not correlate with the number of *home* team yellow cards, $r(1530) = -.009$, *ns*, but did correlate with the number of *away* team yellow cards, $r(1530) = .164$, $p < .001$.

Study 1 Discussion

Based on our hypothesis, we found that crowd density predicts the amount of yellow cards awarded against the away team, which in turn correlates with game outcome. Further, this relationship was strengthened when referees and crowds were in close proximity, which supposedly amplifies the impact of crowd noise as a cue. This is an important insight because some authors failed to find crowd effects on outcomes (e.g., Strauss, 2002); if referees and venues were considered, this inconsistency disappears and allows new and testable hypotheses (at least in our data set). Crowd noise does not directly influence outcomes, but influences referee decisions; additionally, this effect partially depends on the strength of the cue (i.e., proximity of referees and crowd). In turn, only if these referee decisions or judgments are influential enough for the outcome should crowd size, and in the present case, the kind of stadium, influence competition outcomes.

In addition, the separate analysis of yellow cards awarded to the home and away teams clearly supported the Brunswikian approach; higher home crowd density leads to more cards against the away team, but does not influence card decisions toward the home team. While this is highly consistent with a cue-learning approach, it contradicts the idea that referees want to appease the home crowd by calling fewer fouls and awarding fewer cards toward the home team.

One might argue that the observed correlations, albeit highly significant, are not impressive. Yet, in soccer, there are many other decisions that correlate with crowd noise that heavily influence a game's outcome (e.g., free kicks, red cards, and penalties; see Boyko, Boyko, & Boyko, 2007). We did not consider them here because the following experiment used yellow cards as the main dependent variable. As argued above, without experimental evidence to back up archival data, there are many simple counter-arguments: For example, losing teams could start to play more aggressive and the causality is actually reversed; because away teams lose more frequently, they receive more yellow cards. Similarly, better teams could simply have larger or better attended venues and consequently, this third variable could create the correlation. A crowd density index partly avoids this problem (see Note 2), but one would need the additional control that better clubs do not simply have higher crowd density. In addition, crowd density could also cause more revenue for teams and clubs, allowing them to attract better players, leading to better performances. Again, the causality would be reversed.

Given this caveats, the data are highly consistent with earlier reports on the relation between crowd size and referee decisions in soccer. For example, Downward and Jones (2007) reported a similar trend for crowd size and number of yellow cards awarded against the away team. They analyzed 857 games of the Football Association Cup in England and found that 1.71 first yellow cards were awarded toward away teams, whereas only 1.35 cards were awarded toward home teams; again, a highly significant difference (but also see Note 3).

However, our analysis did not include many other variables authors have suggested that might moderate the relation between crowd noise and the home advantage, for example, nonlinear relationships between crowd size and home bias (e.g., Downward & Jones, 2007),[3] crowd composition (i.e., the ratio of home and away spectators; Garicano, Palacios-Huerta, & Prendergast, 2005), or personality variables of the referees (e.g., Page & Page, in press). The point of the present databank analysis was to demonstrate that there is a systematic home bias in awarding yellow cards and this bias is related to crowd density, our proxy for crowd noise. Having established the basic phenomenon, we can now turn to a direct experimental test of the hypothesis that referees use crowd noise as a cue.

Study 2: An Experiment Testing the Influence of Crowd Noise on Yellow Cards in Soccer

The cue-learning hypothesis predicts that louder crow noise should influence referee decisions in the direction of the cue's correlation with the criterion (e.g., noise and foul severity). Hence, they should award yellow cards with higher probability in the same scene given high crowd noise compared with low crowd noise. The following experiment presents a direct test of this prediction.

Study 2 Method

Participants, Design, and Materials

Twenty male referees of the German Football Association (DFB) participated (M_{age} = 22.5; SD = 8.04), and they were recruited during an educational workshop at a DFB training center. They had refereed across various levels of DFB leagues (depending on their age), with a minimum experience of two years. The main independent variable, whether a scene was presented with high or low crowd noise volume, was manipulated within participants.

We used 56 digitally available foul scenes, from 56 different soccer games, that were successfully employed in previous experiments on referee decision making

(Memmert, Unkelbach, Ertmer, & Rechner, 2008; Unkelbach & Memmert, 2008). Twenty-eight of these scenes led to a factual yellow card on the pitch and 28 did not. Naturally, the scenes did not include any hint to the actual decisions and they were rated to be of equal roughness (see Unkelbach & Memmert, 2008). For crowd noise, we used four different sound files containing crowd noise from other soccer stadiums when a foul just happened (i.e., the wavelike increase and decrease when something of importance happens in a stadium). We randomly combined these sound files with the video clips of the fouls scenes (i.e., each sound file was used 14 times). The volume for each scene was set to high and low randomly during the actual presentation (see below). This procedure avoids confounding crowd noise with other factors inherent in the scenes, and crowd noise was thus fully independent of the home and away status of a team. This design also ameliorates the problem of stimulus sampling, as we used scenes from 56 different games. In addition, each scene would be played with high and low volume instead of sound vs. no sound, which avoids the problem that watching a scene without sound is a somewhat unnatural condition for referees. Thus, in this design, any difference between a scene presented with low volume and high volume can only be due to the volume level of crowd noise.

Procedure

The experimenter informed the referees that they would participate in an experiment to determine how and when referees award yellow cards. All referees agreed to participate. Then, the experimenter led a first group of ten referees into a seminar room of the training center, which was equipped with a video projector and a sound system. Referees sat around the projector screen. They received a 56-page questionnaire with two boxes on each page to indicate whether they would award a yellow card for a presented foul or not. The experimenter instructed them that they should make their decision individually and "as if in an actual game." Then the foul scenes presentation started. A Microsoft Visual Basic computer program controlled this presentation and the computer's output volume level, but the volume level of the room's sound system was fixed. The program randomly selected half of the scenes from the factual yellow card and no yellow card categories as high volume scenes (presented at 90% sound output) and the remaining as low volume scenes (presented at 10% sound output), resulting in an approximately 50 dB difference in the volume in terms of the physically equivalence. Each scene lasted 4–7 s, and participants made their decisions immediately afterward. After each decision round, the experimenter prompted the program to continue. Following the 56 scenes, the referees answered demographic questions and were informed that the study was now over. Then, the second group was called into the room. Everything was identical to the first group, with one exception: The presentation order of the scenes was not randomized, but yoked with the first group, and each scene that was selected as a high volume scene was now presented as a low volume scene (and vice versa). After the second group finished, the experimenter debriefed all participants about the hypothesized effects and thanked them for their cooperation

Study 2 Results

In postexperimental funneled questioning, no referee reported suspicion about the sound volume and only one mentioned the variations in sound. We first analyzed participants' mean probability to award yellow cards as a function of high and low volume and actual decision (yellow card or not). As predicted, referees awarded yellow cards with higher probability when scenes were presented with high volume ($M = .589$, $SD = .088$) compared with when the same scenes were presented with low volume ($M = .486$, $SD = .126$), $t(19) = 3.88$, $p < .001$, $d = 1.78$. As could be expected, referees also awarded yellow cards with higher probability when an actual yellow card had been awarded ($M = .723$, $SD = .164$) compared with when no actual yellow card had been awarded ($M = .351$, $SD = .138$), $t(19) = 14.01$, $p < .001$, $d = 4.27$. However, there was no interaction of these two variables, $F(1, 19) = 0.06$, *ns*, indicating that the influence of crowd noise was independent from the factual decision on the pitch.

The previous analysis is collapsed across groups; Table 1 presents the full design (actual decision on the pitch: yellow cards vs. no yellow cards × volume: high vs. low × group: first vs. second group), corrected for referees' mean propensity to award a yellow card, making the comparison of cells equivalent to an overall repeated measurement analysis. This table allows the direct comparison of the same scenes under high and low volume conditions. As can be seen, for all comparisons involving the same scenes, high volume led to higher probabilities of a yellow card. A mixed ANOVA[4] with all three factors (group, factual decisions, and volume, with repeated measures on the latter two factors) delivers the same results as the simple *t* tests, demonstrating a highly significant effect of volume $F(1, 18) = 14.96$, $p < .001$, $d = 1.82$. When scenes were presented with high volume referees were more likely to award yellow cards compared with when the same scenes were presented with low volume. More importantly, this analysis also shows that the effect of volume did not interact with group or factual decision on the pitch, *F*s < 1, *ns*.

Study 2 Discussion

High volume crowd noise led to substantially more yellow cards than low volume crowd noise. Presented with high volume crowd noise, referees had an approximately .10 higher probability to award a yellow card than when the *identical* scene

Table 1 Mean Decision Proportions to Award a Yellow Card as a Function of Factual Decision, Group, and High vs. Low Volume (*SD* in Parentheses). Comparisons of High and Low Volume Are Based on Identical Scenes.

Factual Decision	High Volume (Group 1)	Low Volume (Group 2)	High Volume (Group 2)	Low Volume (Group 1)
Yellow Card	.826 (.084)	.684 (.151)	.720 (.105)	.669 (.171)
No Yellow Card	.332 (.124)	.261 (.097)	.485 (.078)	.332 (.102)

was presented with low volume crowd noise. As argued, we believe this effect is due to referees learning the correlation between foul severity and the crowd noise.

In this vein, the present results differ from the data by Nevill, Balmer, and Williams (2002), who found *fewer* challenges awarded for the home team when crowd noise was present. Yet, the data converge with the results from Nevill, Balmer, and Williams (1999). Given this convergence, together with Sutter and Kocher's critique (2004) and the data presented in our archival analysis, it seems possible that there is something specific about the game used in Nevill and colleagues' 2002 experiment.

One limitation of the current study, and probably of most experimental studies, is the laboratory situation. Referees who judge scenes on a video screen are hardly in the same situation as referees on the pitch. At present, we see now better and feasible way to investigate the effects with clear manipulations of the independent variable. Yet, in our opinion, the combination of archival data analyses—that is, factual judgments and decisions, with somewhat artificial, but highly controlled laboratory experiments—provides a useful compromise. For example, Hill and Barton (2005) argued, based on archival data, that wearing red in a combat sport such as wrestling is beneficial for the athlete. However, more recently, Hagemann, Strauss, and Leissing (2008) found in an experiment that it is indeed a refereeing bias that creates the advantage for wearing red. Given identical performances, judges awarded more points in Tae Kwon Do to fighters who wore red compared with fighters who wore blue. Thus, neither pure archival data nor laboratory data alone can answer the most interesting questions in sport psychology.

Overall Conclusions

The psychological demands of refereeing and performance judgments have instigated increasingly more research within the last few years (Mascarenhas, O'Hare, & Plessner, 2006; Plessner & Haar, 2006). Building on this prior research, we demonstrated the impact of crowd noise on referees' decisions when they have to judge the severity of foul scenes; thereby, we showed the possible contribution of crowd noise to the home advantage via referee decisions. The underlying model is the assumed correlation between crowd noise, a proximal cue, and the severity of the foul, the distal criterion. The cue should be context specific, so that the same loud audience reaction leads to more positive evaluations in figure skating, but to a higher probability of a yellow card for fouls in soccer. Thus, although we have only used soccer as an exemplar, the model is applicable to all sports in which subjective judgments influence a competition's outcome or the rating of a performance; this guiding model is implicitly present in many other conceptions of crowd noise influence (e.g., Nevill et al., 1999), and even explicitly spelled out in models of referee decision making (e.g., Plessner et al., 2009).

Yet, the present studies did not test the underlying cognitive process directly; at present, the advantage of our cue-learning approach over other explanations (e.g., motivational approaches, and crowd noise as a heuristic) rests in its foundation in well-established cognitive models and the clearly testable predictions. The direct test of the cue-learning hypotheses is a line of research we are pursuing right now. Based on other successful cases on learning and relearning cues (e.g., Unkelbach, 2006, 2007), we are creating situations in which crowd noise is no longer a valid cue or even correlates inversely with the criterion to be judged; that is, referees should learn that crowd noise is actually louder when there is no foul and judge following incidents accordingly. If this relearning is successful, it will provide strong evidence for the case of cue learning.

The Brunswikian approach is a pure cognitive model for the influence of crowd noise on home advantages. However, we do not want to preclude motivational accounts on the home bias. For example, Dohmen (2008) as well as Garicano and colleagues (2005) presented evidence that soccer referees allow differential amounts of extra time, depending on the score: Referees allowed more extra time when home teams needed one more goal to win or even the score, whereas they allowed less extra time when home teams were leading. Such effects cannot be explained by our cue-learning hypothesis; awarding extra time is a deliberate decision, and motivational factors have greater explanatory power in this domain. It will be a challenge for future research to reconcile and integrate deliberate and motivational processes with basic cognitive effects such as cue learning in order to achieve a more comprehensive model of the home advantage. As it is, we believe the presented experiment clearly establishes the influence of crowd noise on referee decisions. In combination with the present archival data, we are confident about the evidence that crowd noise influences referees' yellow cards decisions, which in turn contributes to a home advantage.

Notes

1. The use of crowd density as a proxy for crowd noise instead of absolute visitor numbers has an additional advantage; in the German Bundesliga, the size of the stadium correlates with the strength of the team, as better teams have larger stadiums. This problem does not exist with percentages.

2. We do not report effect sizes for the databank analysis; all reported effects (i.e., mean difference in goals scored, mean differences in yellow cards awarded) refer to scales with an inherent meaning, and can therefore be judged according to "practical" significance (Kirk, 1996; Thompson, 2002). In addition, correlation coefficients are effect size indicators by themselves (Rosenthal & Rubin, 2003).

3. The nonlinear trend reported by Downward and Jones (2007) fits nicely with well-established models of social impact (e.g., Latané, 1981). However, the data reported by Downward and Jones suggests some problems with the analysis. It is not clear whether variables were centered before using them in the logistic regression analysis and the rather weak effects in terms of significance suggest multicollinearity in the regression. Thus, these results should be treated with caution.

4. Many researchers advocate the use of logistic regression for a binary dependent variable. However, standard regression models, which factually underlie the presented ANOVA results, have no trouble dealing with binary dependent variables, given sufficient cell frequencies and no extreme values (<.05 and >.95; Lunney, 1970). We believe the ANOVA presentation is easier to understand than logistic regression results.

Acknowledgments

The present research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG UN 273/1-1), awarded to the first author.

References

Agnew, G., & Carron, A. (1994). Crowd effects and the home advantage. *International Journal of Sport Psychology, 25,* 53–62.

Axsom, D., Yates, S.M., & Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. *Journal of Personality and Social Psychology, 53,* 30–40.

Balmer, N.J., Nevill, A.M., & Lane, A.M. (2005). Do judges enhance home advantage in European championship boxing? *Journal of Sports Sciences, 23,* 409–416.

Balmer, N.J., Nevill, A.M., & Williams, A.M. (2001). Home advantage in the Winter Olympics (1908–1998). *Journal of Sports Sciences, 19*, 129–139.

Balmer, N.J., Nevill, A.M., & Williams, A.M. (2003). Modelling home advantage in the Summer Olympic Games. *Journal of Sports Sciences, 21,* 469–478.

Balmer, N.J., Nevill, A.M., Lane, A.M., Ward, P., Williams, A.M., & Fairclough, S. (2007). Influence of crowd noise on soccer refereeing consistency in soccer. *Journal of Sport Behavior, 30,* 130–145.

Baumeister, R.F., & Steinhilber, A. (1984). Paradoxical effects of supportive audiences on performance under pressure: The home field disadvantage in sports championships. *Journal of Personality and Social Psychology, 47,* 85–93.

Boyko, R.H., Boyko, A.R., & Boyko, M.G. (2007). Referee bias contributes to home advantage in English Premiership football. *Journal of Sports Sciences, 25*(11), 1185–1194.

Brunswik, E. (1957). Scope and aspects of the cognitive problem. In H. Gruber, K.R. Hammond, & R. Jessor (Eds.), *Contemporary approaches to cognition* (pp. 5–31). Cambridge: Harvard University Press.

Carron, A.V., Loughhead, T.M., & Bray, S. (2005). The home advantage in sport competitions: Courneya and Carron (1992) conceptual framework a decade later. *Journal of Sports Sciences, 23,* 395–407.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752–766.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.

Clarke, S.R., & Norman, J.M. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician, 44,* 509–521.

Courneya, K.S., & Carron, A.V. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport & Exercise Psychology, 14,* 13–27.

Devereux, P., & Ginsburg, G. (2001). Sociality effects on the production of laughter. *The Journal of General Psychology, 128,* 227–240.

Dohmen, T.J. (2008). The influence of social forces: Evidence from the behavior of football referees. *Economic Inquiry, 46,* 411–424.

Downward, P., & Jones, M. (2007). Effects of crowd size on referee decisions: Analysis of the FA Cup. *Journal of Sports Sciences, 25,* 1541–1545.

Evans, J.S., Clibbens, J., Cattani, A., Harris, A., & Dennis, I. (2003). Explicit and implicit processes in multicue judgment. *Memory & Cognition, 31,* 608–618.

Garicano, L., Palacios-Huerta, I., & Prendergast, C. (2005). Favoritism under social pressure. *The Review of Economics and Statistics, 87,* 208–216.

Hagemann, N., Strauss, B., & Leissing, J. (2008). When the referee sees red.... *Psychological Science, 19,* 769–771.

Hertwig, R., Herzog, S., Schooler, L., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34,* 1191–1206.

Hill, R., & Barton, R. (2005). Red enhances human performance in contests. *Nature, 435*(7040), 293.

Holder, R.L., & Nevill, A.M. (1997). Modelling performance at international tennis and golf tournaments: Is their a home advantage? *Journal of the Royal Statistical Society. Series A, (Statistics in Society), 46,* 551–559.

Jacobs, R.A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences, 6,* 345–350.

Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56,* 746–759.

Latané, B. (1981). The psychology of social impact. *The American Psychologist, 36,* 343–356.

Lunney, G.H. (1970). Using analysis of variance with a dichotomous dependent variable: an empirical study. *Journal of Educational Measurement, 7,* 263–269.

Mascarenhas, D., O'Hare, D., & Plessner, H. (2006). The psychological and performance demands of association football refereeing. *International Journal of Sport Psychology, 37,* 99–120.

Memmert, D., Unkelbach, C., Ertmer, J., & Rechner, M. (2008). Gelb oder kein Gelb? Persönliche Verwarnungen im Fußball als Kalibrierungsproblem [Yellow card or no yellow card? Soccer cautioning as a calibration problem]. *Zeitschrift für Sportpsychologie, 15,* 1–11.

Miron, A.M., & Brehm, J.W. (2006). Reactance theory: 40 years later. *Social Psychology, 27,* 9–18.

Nevill, A.M., & Holder, R.L. (1999). Home advantage in sport: An overview of studies on the advantage of playing at home. *Sports Medicine (Auckland, N.Z.), 28,* 221–236.

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.

Nevill, A.M., Balmer, N.J., & Williams, A.M. (1999). Crowd influence on decisions in association football. Lancet, *353*(9162):1416.

Nevill, A.M., Newell, S.M., & Gale, S. (1996). Factors associated with home advantage in English and Scottish soccer. *Journal of Sports Sciences, 14,* 181–186.

Page, K., & Page, L. (2010). Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, *31*, 192–199.

Plessner, H., & Haar, T. (2006). Sports performance judgments from a social cognition perspective. *Psychology of Sport and Exercise, 7,* 555–575.

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.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences, 4,* 237–248.

Pollard, R., & Pollard, G. (2005). Long-term trend in home advantage in professional team sports in North America and England (1876-2003). *Journal of Sports Sciences, 23,* 337–350.

Rosenthal, R., & Rubin, D. (2003). r-sub(equivalent): A Simple Effect Size Indicator. *Psychological Methods, 8,* 492–496.

Schwartz, B., & Barsky, S.F. (1977). The home advantage. *Social Forces, 55,* 641–661.

Strauss, B. (2002). The impact of supportive sports spectators on performance in team sports. *International Journal of Sport Psychology, 33,* 372–390.

Sutter, M., & Kocher, M.G. (2004). Favoritism of agents – The case of referees' home bias. *Journal of Economic Psychology, 25,* 461–469.

Thompson, B. (2002). 'Statistical,' 'practical', and 'clinical': How many kinds of significance do counselors need to consider? *Journal of Counseling and Development, 80,* 64–71.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science, 17,* 339–345.

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33,* 219–230.

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.

Wallace, H.M., Baumeister, R.F., & Vohs, K.D. (2005). Audience support and choking under pressure: A home disadvantage. *Journal of Sports Sciences, 23,* 429–438.

Wells, G., & Windschitl, P. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125.

Manuscript received: October 5, 2009

Revision accepted: April 1, 2010