

Running head: Fair judgment

How the motivation to judge fair influences memory for in- and out-group behavior

Martin Scarabis

University of Muenster, Germany

Arnd Florack

University of Basel, Switzerland

(Swiss Journal of Psychology, IN PRESS)

Martin Scarabis

Psychologisches Institut IV

Universität Münster

Fliednerstraße 21

48149 Münster

Germany

Email: scarabis@psy.uni-muenster.de

Phone: ++49/251/83-39149

Fax: ++49/251/83-31351

Key words: Stereotypes, Prejudice, Suppression, Rebound

Abstract

The present study investigated the subtle effects of an instruction to judge fairly and to avoid the influence of stereotypes on judgments. Previous studies have shown that instructions to disregard stereotypes may lead to an enhanced accessibility of the respective stereotype, which, in turn, facilitates encoding and identification of stereotype-consistent information, but not of stereotype-irrelevant information. Extending this research, the present study examined the effects of a fairness instruction on the encoding of stereotype-inconsistent information compared to stereotype-consistent information. It was found that participants who were instructed to avoid stereotypic judgments showed enhanced memory for stereotype-consistent as well as for stereotype-inconsistent information. Thus, fairness instructions have the unwanted effect of enhancing memory for stereotype-consistent information, but also the desired effect of enhancing memory for information contradicting the stereotype.

How the motivation to judge fair influences memory for in- and out-group behavior

"In the truest sense of the word, fair play was trampled underfoot," FIFA president Sepp Blatter said at a news conference after a world cup play-off between the national soccer teams of Switzerland and Turkey on November 16, 2005. The game was accompanied by mutual accusations of unfair behavior, leading to a scuffle in the tunnel right after the end of the game. Players of both teams were punished later by the FIFA. This incident attracted our attention, but not only because of the players' behavior, but also because of the biased perception in both countries. Indeed, in the view of many Swiss soccer fans, the Turkish players and team personnel behaved rudely and aggressively, while many Turks might have thought the same about the Swiss team members. Since these differences in perceptions became obvious, some people argued that such incidents often lead to overgeneralization and may affect the relations of the involved groups negatively. Thus, it seems very reasonable to call for more fairness on as well as outside the soccer field.

In this paper, we examine the effects of motivated fairness on the encoding of stereotype-consistent and stereotype-inconsistent incidents. Indeed, recent research has shown that the motivation to form a fair judgment may have the negative effect of enhancing memory for stereotype-consistent information (Macrae, Bodenhausen, Milne, & Wheeler, 1996; Sherman, Stroessner, Loftus, & Deguzman, 1997). However, this research left open whether the motivation to form fair judgments might also enhance memory for stereotype-inconsistent information. We put forward that fairness motivations enhance the encoding of all stereotype relevant information and we report a study demonstrating that people who were instructed to judge fairly

showed better memory for stereotype-consistent *and* stereotype-inconsistent information than people who did not receive this instruction.

Theoretical Background

The effects of group affiliations on the judgment of out-group members we mentioned in the introduction are well documented. Hastorf and Cantrill (1954), for instance, found that the fairness and unfairness of behaviors during a football match was affected by the group membership of the judges. In addition, other research has shown that stereotypes may bias judgments and impressions (e.g., Shaffer & Collier, 2002; Stone, Perry, & Darley, 1997), lead to biased decisions (e.g., Gray & Ashmore, 1976; Kite et al., 2001), and - obviously - impair intergroup relations. Thus, there are good reasons why it would be desirable for people to avoid influences from stereotypes on their judgments and behavior.

A common strategy for addressing the problem of stereotypical judgment is to warn people about potential biases and to ask them to judge fairly. A central question concerns the consequences of such instructions. Are these efforts successful at removing the unwanted consequences of stereotypes and at reducing stereotypes in the long term, or are there undesired side effects when people intentionally try to make judgments that are not affected by stereotypes?

Stereotype Correction

Research over the last two decades or more has indeed demonstrated that fairness instructions or warnings that stereotypes might influence a judgment can have positive and negative effects. While some studies have shown that people are able to correct their judgments for stereotypical influence (e.g., Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002), others have provided evidence that the intentional suppression of stereotypes may

also have unwanted effects (e.g., Macrae, Bodenhausen, Milne, & Jetten 1994; Sherman et al., 1997). It seems that people are able to correct their explicit judgments for stereotypic bias if some conditions hold true: a) they know about the possible influence on their behavior (Aberson & Ettl, 2004), b) they have a naïve theory of how much they are influenced and how to correct the bias (Wegener & Petty, 1997; Wegener, Dunn, & Tokusato, 2001), and c) the affected behavior can be deliberately controlled, or, if it is automatic, individuals are trained how to establish automatically controlled responses (Kawakami et al., 2000). However, even if all these conditions apply, and people are informed about the possible influence of a certain stereotype and are explicitly asked to provide fair judgments, subtle negative effects may result.

Rebound Effects

One possible undesirable consequence of the conscious avoidance of stereotype influence are rebound effects, the phenomenon that people may suppress the influence of stereotypes on explicit judgments for a certain time, but then rely on stereotypes even more strongly if the reason for suppression is removed. For example, Macrae et al. (1994) found in a series of three experiments that the induced goal of suppressing stereotypic thinking about skinheads led to an increased stereotypicality of verbal expressions, behavioral effects, and an enhanced accessibility of the stereotype when the goal of stereotype suppression was no longer active. Furthermore, there are several studies showing the impact of the motivation to suppress stereotypes on memory processes (Macrae et al., 1996; Macrae, Bodenhausen, Milne, & Ford, 1997; Sherman et al., 1997). Macrae et al. (1996), for instance, demonstrated that people were better at recalling stereotypical information about an elderly man when they were instructed not to rely on the

stereotype while forming an impression about the target. Similarly, Sherman et al. (1997) found that people who listened to an audiotape recording of an Asian female describing her activities during a typical day could better recognize stereotypical information than non-stereotypical information, but only when they were asked to avoid stereotypes during impression formation and not when they received no such instruction. Thus, there is some evidence that the suppression of stereotypes leads to improved memory for stereotypic information.

A prominent explanation for such unwanted subtle consequences of stereotype suppression is derived from the analyses of the processes that are involved when people try to suppress unwanted thoughts. Wegner (1994) suggested that two processes are activated when people try to inhibit thoughts. One process is assumed to be concerned with the detection of unwanted thoughts. If unwanted thoughts are detected by this automatic monitoring process, a second, intentional process is supposed to suppress the use of the unwanted thoughts and to concentrate on distractors that replace the to be suppressed thoughts. As the first process implies a search for unwanted thoughts in memory such material should become more accessible. Furthermore, if the monitoring process finds suspicious information, particular attention from the individual is assumed to be necessary to avoid the influence of the information on impression formation and judgment. According to the suppression of stereotypes model (Macrae et al., 1994), this goes along with an increased accessibility of the stereotype in memory and, therefore, leads to an improved encoding for stereotype-consistent information that is provided in the context of impression formation. In support of this reasoning, it has been found that people who received a suppression instruction are better at recalling (Macrae et al., 1996) and recognizing stereotype-consistent information (Sherman et al., 1997).

The experiments of Sherman et al. (1997) and Macrae et al. (1996) share some features restricting their relevance for situations of higher complexity. First, participants received information about single targets only. Hence, the studies mainly demonstrate the influence of the motivation not to stereotype on the impression formation of a single person. In the introduction, we referred to an example in which behavior by multiple in-group and out-group persons was observed. Outside the lab, individuals often perceive behavior of several different in-group and out-group members. Second, the available information about the target person was either *stereotype-consistent* or *stereotype-irrelevant*. In the studies of Macrae et al. and Sherman et al., *stereotype-inconsistent* information was not included. For example, the information in the study of Sherman et al. that someone prepares dinner at home is not relevant to the stereotype of Asian-Americans and is, therefore, not *stereotype-inconsistent*. With a few exceptions, previous studies examining the effects of suppression or fairness instructions contained only *stereotype-irrelevant*, but not *stereotype-inconsistent* information.

One exception is a study by Dumont et al. (2003). Participants of this study who were asked to suppress the stereotype were more affected in their impression by *stereotype-consistent* as well as by *stereotype-inconsistent* information than participants who did not receive this instruction. In addition, Galinsky and Moskowitz (2007) demonstrated in a study that the instruction to suppress a certain stereotype makes *stereotype-consistent* and *stereotype-inconsistent* traits more accessible, while the accessibility of *stereotype-irrelevant* traits remains unaffected.

The Present Study

In the present study, we examined whether an instruction to form a fair judgment enhances the encoding of *stereotype-consistent* and *stereotype-*

inconsistent information. First, we assume that fairness instructions direct the attention to stereotype-consistent incidents, because individuals have to control their judgments of these incidents. Second, we hypothesize that fairness instructions direct also the attention to stereotype-inconsistent incidents, because individuals can use these incidents to replace their stereotype-consistent thoughts. Furthermore, we assumed that fairness instructions mainly direct the attention to the behavior of out-group members while the attention to in-group behaviors will be reduced. We suppose that individuals have a naïve theory that unfair judgments result mainly from a biased judgment of out-group members. Thus, compared to a condition without a fairness instruction a condition with fairness instruction should lead to enhanced memory for stereotype-consistent and stereotype-inconsistent out-group behaviors while the memory for in-group behaviors should be reduced.

We tested our hypotheses in an applied setting involving the judgment of behaviors of a great number of in-group and out-group members. We asked participants of our study to judge the behavior of German and Turkish players of German amateur leagues. In the German amateur leagues, the problems of foreign players and unfair playing behavior are openly discussed. There is a prevailing stereotype that Turkish players, who represent a great percentage of players in German amateur leagues, behave unfairly and rudely more often than German players. This surely holds for German players and German viewers of soccer games, but it seems not to be limited to these groups. For example, Scarabis, Florack, and Gosejohann (2003) found that German pupils and students showed a bias to automatically associate the category of Turks more strongly with unfair behavior than the category of Germans. Pilz (2000) analyzed archival data of sentences handed

down by German athletic disciplinary courts. He found that non-German players received significantly longer suspensions than Germans for the same type of foul.

The present study was based on the research paradigm of Sherman et al. (1997). Half of the participants were instructed to judge fair, and recognition sensitivity for the presented information was assessed. However, our study differed from the study of Sherman et al. in two main points. First, participants in our study received information about stereotype-consistent and stereotype-inconsistent behaviors. Second, we did not use an impression formation task in which participants were asked to form an impression of a single target. We applied a procedure in which participants had to make a great number of judgments including multiple targets. This task was difficult for the participants, and it was also similar to the task of referees in sports who have to make a lot of single judgments.

Method

Participants and design

Ninety-two participants (35 female, 57 male) were randomly assigned to one of two conditions (fairness instruction: yes or no). The mean age was 26.97 years (ranging between 20 and 56 years). Participants were recruited by a cover story titled "soccer expertise". As compensation for taking part in the experiment, student participants received course credit. Four participants indicated a citizenship other than German and were excluded from the following analysis.

Procedure

After arriving at the laboratory, participants were seated in front of a personal computer. All instructions and experimental tasks were presented on the computer screen. First, participants read that they would have to judge several soccer scenes from German amateur leagues according to whether a player showed fair or unfair

behavior. Then, in one condition, an instruction was given to disregard the ethnic group membership of the players ($n = 43$). In this fairness instruction condition, participants were informed that there was a tendency to judge foreign players more severely when they play foul, and that they should not base their judgment of fairness on the nationality of the players, but only on the real fairness of the situation. In addition, all participants in the fairness instruction condition received a mock magazine essay that was concerned with violence and soccer. One paragraph was titled "Foreign players receive higher penalties," another "Importance of fair judgments." At the end of the article, participants received the instruction to judge fairly on the following task. We decided to include the mock magazine in order to strengthen participants' motivation to comply to the experimental task by providing them with a credible and face-valid background story. In the no-fairness instruction condition ($n = 45$), participants did not read the article and were not asked for an unbiased judgment, but to simply read the scenes and judge them according to the targets' fairness on a six point-scale.

Following the initial instructions, all participants were presented with 32 scenes from different soccer matches in the German amateur league and the instruction to read the scenes and to judge the fairness of the player (whose name was capitalized) on a six-point scale (1 = *very unfair*, 6 = *very fair*). Names of the target players were of unambiguously German (e.g., Holger Westhoff, Christian Peters) or Turkish origin (e.g., Abdullah Buruk, Mustafa Aksu). Additionally, photographs of the respective targets were shown. These pictures were taken from a larger pool and only those were selected that were clearly identified in a pretest as showing a German or a Turk. The presented scenes contained eight scenes depicting clearly fair behavior by a Turkish player and eight scenes depicting fair play by a German

player, and eight scenes describing unambiguously unfair behavior by a Turkish player and eight describing unfair behavior by a German player. A fair scene, for example, described a member of the losing team who congratulates each player of the winning rival team. An unfair scene, for instance, showed the target player shoving another player. Examples for each type of vignettes can be found in the Appendix. The order of presentation of the scenes was randomized. To control for influences of the behavioral descriptions, the group membership of the target players was reversed for half of the participants. This means that half of the participants saw a fair or unfair scene with a Turkish target, and the other half saw the identical scene with a German target. *Fairness judgments* for the scenes from the encoding phase were averaged separately for fair behaviors by German targets, unfair behaviors by German targets, fair behaviors by Turkish targets, and unfair behaviors by Turkish targets.

The central dependent measure of the present study was *recognition sensitivity* assessed by a recognition task. After participants completed their ratings, they were informed that they would see scenes from soccer matches again, and this time the names and pictures of the players would be removed. Furthermore, they were asked to indicate whether the scene had appeared in the first part of the study and, if so, whether the target was German or Turkish. Participants were further instructed to guess the nationality if they could not remember the target's nationality in the original scenes. Altogether, each subject saw 64 scenes which contained the 32 descriptions presented before and, additionally, 32 distractor scenes that had not been presented before. As in the study of Sherman et al. (1997) and others (e.g., Koomen & Dijkster, 1997; Sherman & Frost, 2000), we computed Grier's (1971) index A' as a nonparametric measure of recognition sensitivity on the basis of participants

old/new judgments.¹ This was done separately for four types of information (fair behavior by Turkish/German players and unfair behavior by Turkish/German players).

Finally, questions regarding demographic data were asked and some scales unrelated to the present study were applied.

Results

Fairness Judgments

An analysis of variance with the fairness manipulation (no fairness instruction vs. fairness instruction) as independent factor and the targets' group membership (Turkish vs. German) and the type of behavior (fair vs. unfair) as within-subject factors revealed a main effect for type of behavior, $F(1, 86) = 1605.56$, $p = .000$, $\eta^2 = .95$. Participants rated fair behavior ($M = 5.11$; $SD = .49$) as fairer than unfair behavior ($M = 2.07$; $SD = .42$). Thus, participants clearly distinguished between fair and unfair behaviors. In addition, the analysis revealed a marginal significant interaction between the targets' group membership, the type of behavior, and the fairness manipulation, $F(1, 86) = 3.33$, $p = .07$, $\eta^2 = .04$. When participants received a fairness instruction, they rated the fair behavior by Turkish players ($M = 5.22$; $SD = .53$) as fairer than fair behavior by German players ($M = 5.03$; $SD = .38$), $t(42) = 2.28$, $p = .03$, $d = .70$, but not when they received no such instruction, $t(44) < 1$, *ns*. No differences between the judgment of Turkish and German players as regards the unfair behavior were found either in the fairness instruction condition, $t(42) < 1.3$, *ns*, or in the no-fairness instruction condition, $t(44) < 1$, *ns*. Furthermore, the interaction between the targets' group membership and the type of behavior was significant, $F(1, 86) = 4.23$, $p = .04$, $\eta^2 = .05$. If the experimental conditions are not taken into account, the effects observed in the fairness instruction condition are also apparent.

There was a tendency that the participants rated the fair behavior by Turkish players ($M = 5.12$; $SD = .57$) as fairer than fair behavior by German players ($M = 5.07$; $SD = .51$), $t(87) = 1.85$, $p = .06$, $d = .40$. The difference between the unfair behaviors of German and Turkish players was not significant, $t(87) < 1$, *ns*.

Recognition Sensitivity

We hypothesized that participants would be more sensitive to recognizing fair and unfair behaviors by Turkish targets when they were instructed to make fair judgments than when they received no fairness instruction. Furthermore, we assumed that a fairness instruction would weaken the memory for in-group behaviors. In line with these hypotheses, an analysis of variance with the fairness instruction (no fairness instruction vs. fairness instruction) as independent factor and the targets' group membership (Turkish vs. German) and the type of behavior (fair vs. unfair) as within-subject factor yielded a significant interaction between the fairness instruction and the targets' group membership, $F(1, 86) = 28.74$, $p = .000$, $\eta^2 = .05$, while the three-way interaction between the fairness instruction, the targets' group membership, and the type of behavior was not significant, $F(1, 86) < 1$, *ns*. The means for the interaction between the fairness instruction and the targets' group membership are depicted in Table 1.

Participants were better at recognizing behavior by Turkish targets when they received a fairness instruction ($M = .61$; $SD = .07$) than when they received no fairness instruction ($M = .56$; $SD = .07$), $t(86) = 3.77$, $p = .000$, $d = .81$. But they were worse at recognizing behavior by German targets when they received a fairness instruction ($M = .55$; $SD = .08$) than when they received no fairness instruction ($M = .61$; $SD = .08$), $t(86) = 3.80$, $p = .000$, $d = .82$. Since the three-way interaction was not significant, it can be concluded that participants had a better memory for

behaviors of Turkish targets when they received a fairness instruction irrespective of whether the behavior was fair or unfair. Participants were better at recognizing behaviors by Turkish targets ($M = .61$; $SD = .07$) than behaviors by German targets ($M = .55$; $SD = .08$) when they received a fairness instruction, $t(42) = 4.13$, $p = .000$, $d = 1.28$, while they were better at recognizing behaviors by German targets ($M = .61$; $SD = .08$) than behaviors by Turkish targets when they received no fairness instruction ($M = .56$; $SD = .07$), $t(44) = 3.46$, $p = .001$, $d = 1.04$. These effects are consistent with the prediction that a fairness instruction increases the relevance of out-group targets behavior in general. In addition to the reported effects, the analysis of variance yielded a marginally significant effect of the type of behavior, $F(1, 86) = 3.29$, $p = .07$, $\eta^2 = .04$. The recognition for unfair behavior ($M = .59$, $SD = .08$) was higher than the recognition for fair behavior ($M = .57$, $SD = .06$). All other effects were not significant, $F_s(1, 86) < 2.04$, $ps > .15$, $\eta^2 < .03$.

A detailed analysis of the contrasts on the level of the different types of behaviors underlines that the reported effects hold true for unfair as well as for fair behavior. When participants received a fairness instruction, they were better at recognizing fair ($M = .62$; $SD = .09$) as well as unfair behavior ($M = .61$; $SD = .09$) of the Turkish targets than when they received no fairness instruction (fair behavior: $M = .55$; $SD = .07$; unfair behavior: $M = .57$; $SD = .11$), $ts(86) > 1.73$, $ps < .05$, $ds > .37$, one-sided. As regards the German targets, participants were worse in recognizing fair ($M = .54$; $SD = .07$) and unfair behavior ($M = .56$; $SD = .13$) when they received a fairness instruction than when they received no fairness instruction (fair behavior: $M = .59$; $SD = .07$; unfair behavior: $M = .63$; $SD = .11$), $ts(86) > 2.89$, $ps < .01$, $ds > .62$.

Discussion

The present study replicates previous findings (Macrae et al., 1996; Sherman et al., 1997) that the instruction to judge fairly and avoid stereotypic influences leads to an enhanced memory performance for stereotype-consistent information. Extending previous research, the current study additionally shows that a fairness instruction also enhances memory for stereotype-inconsistent behavior. In previous studies of Macrae et al. (1996) and Sherman et al. (1997), participants observed only stereotype-consistent or stereotype-irrelevant behavior, but not behavior that contradicted stereotypical expectations. In the current study, we presented participants stereotype-consistent as well as stereotype-inconsistent behaviors of soccer players and found that the memory for stereotype-consistent *and* stereotype-inconsistent behaviors was stronger when participants received an instruction to form fair judgments than when they did not receive this instruction. Thus, the current study provides a less pessimistic view on the effects of fairness instruction showing that, in addition to memory for stereotype-consistent information, at least the memory for stereotype-consistent information is enhanced when individuals try to make fair judgments.

The findings of the current study fit our basic assumption that individuals who try to avoid stereotypical influence focus on incidents in which they are in danger to apply the stereotype (unfair behaviors in the current study) and, additionally, they focus on stereotype-inconsistent incidents (fair behaviors in the current study) which might help them to replace their stereotype-consistent thoughts. The first process should enhance encoding of stereotype-consistent information. The second process should enhance encoding of stereotype-inconsistent information. The results of the present study also complement findings by Galinsky and Moskowitz (2007). The researchers demonstrated that an induced motivation not to stereotype heightens not

only the accessibility of stereotype-consistent traits, but also the accessibility of stereotype-inconsistent traits as well. We did not assess the accessibility of information, but the memory for stereotype-consistent and stereotype-inconsistent behaviors and found an enhanced recognition sensitivity for both kinds of behaviors after participants were instructed to judge fairly. The finding that, as regards the fair scenes, participants who received a fairness instruction evaluated the out-group players more positively than the in-group players also shows that participants focus particularly on stereotype-inconsistent incidents. Indeed, the difference in evaluations of in-group and out-group players was not significant for the unfair incidents. Thus, the results suggest that deliberate control processes regard particularly the observation of stereotype-inconsistent behaviors. Indeed, an obvious way to avoid a stereotype is to highlight stereotype inconsistent behaviors.

In support of our further predictions, participants of the current study showed lower recognition sensitivity for in-group behavior when they read a fairness instruction than when they read no such instruction. We suppose that individuals have a naïve theory that they can avoid biased judgments most if they control the judgments of out-group behaviors. Since such control processes require cognitive resources (Galinsky & Moskowitz, 2007), it seems reasonable that fairness-instructed individuals have fewer cognitive capacities left to encode in-group behavior. Interestingly, we found a memory advantage for in-group behavior compared to out-group behavior if participants did not read a fairness instruction. This finding may be a result of an out-group homogeneity effect (Simon, 1992). Several studies have revealed that people often perceive less differences between out-group members than between in-group members especially in natural groups (Ostrom & Sedikides, 1992) and that it is more difficult for individuals to distinguish different out-group

members than to distinguish different in-group members (Sporer, 2001). We assume that this process is predominant if individuals are not particularly motivated to avoid stereotypic influences on their judgments.

Limitations of the Present Study

In the present study, participants worked on a complex task and recognition was difficult. Since previous research has shown higher recognition sensitivity for expectancy-incongruent information in tasks with greater processing demands (cf. Stangor & McMillan, 1992; Rojan & Pettigrew, 1992), it cannot be ruled out that the memory effects of the fairness instructions might be weaker for less complex tasks, for example, for single target impression formation.

Another limitation of the present study might result from the kind of incidents we presented. In previous studies (Macrae et al., 1996; Sherman et al., 1997), researchers presented stereotype-consistent and stereotype-irrelevant information. We presented participants incidents with in-group and out-group members showing fair or unfair behavior in soccer games, but we did not present behaviors that were completely irrelevant for the stereotype (e.g., a player passing the ball to another player). We conducted the study without the presentation of irrelevant behaviors, because previous studies have already shown the memory advantage for stereotype-consistent behavior compared to stereotype-irrelevant behavior (Macrae et al., 1996; Sherman et al., 1997). However, future studies might rely on a more realistic form of presentation including all kinds of behaviors. For example, recognition sensitivity could be assessed after participants watched complete soccer matches.

Implications

The results of the current paper imply that one could be a little bit more optimistic about the effectiveness of instructions to avoid stereotypical biases than

results of previous studies suggest (e.g., Macrae et al., 1996; Sherman et al., 1997). If individuals who are forced to suppress stereotypes tend to memorize stereotype-consistent and stereotype-inconsistent information better than uninstructed persons, they should at least develop inconsistent or more differentiated out-group representations in memory over time. However, it should be taken into account that the naturally occurring frequency of behaviors might be relevant in considering the implications of the present study for stereotype maintenance or change. In real life, behavior that directly contradicts stereotypes might be rather rare. This is especially true for competitive situations (Sherif & Sherif, 1953). Let us consider the example of watching a soccer game. People watching an average soccer game may have a greater chance of observing rude or unfair behaviors than outstandingly fair behaviors. The results of the present study suggest that for individuals who try to make fair judgments, memory for rude behavior by a stereotyped group should be better than for rude behavior by players of a non-stereotyped group. If strikingly fair behavior is indeed rare, such rude behavior might have stronger effects on the maintenance of the stereotype than the improved memory might have in the opposite direction.

In addition, it is reasonable to assume that the controlled focus on stereotype-inconsistent behaviors needs cognitive resources which are not always available. In many real-life situations, people are distracted by the flow of stimuli for their environment. Consider again the example of soccer fans watching an exciting match. Under such circumstances, automatic memory processes which lead to an enhanced accessibility of stereotype-consistent information might be still active while controlled processes related to the focus on stereotype-inconsistent incidents might work only to a limited degree.

Thus, altogether the current paper suggests that instructing people to judge fairly might have the unwanted side effect that people can better remember stereotype-consistent behavior. This negative effect might be leveled out by an improved memory for stereotype-inconsistent information. However, one must take into account that this requires at least an equal frequency of stereotype-consistent and stereotype-inconsistent events as well as ample cognitive resources during encoding. Furthermore, research on stereotype change shows that the perception of inconsistent behavior has positive effects only a) when it is perceived repeatedly, b) when it is perceived as performed by many different out-group members (Weber & Crocker, 1983), and c) when it is performed by persons that are typical for the out-group (Wilder, 1984). Thus, possible positive effects of an enhanced memory for stereotype-inconsistent behaviors might be restricted to these conditions, or at least to the frequency of such behaviors. To draw firm conclusions about the benefit of fairness instructions, one must consider in addition different cognitive processes and behaviors (cf. Kawakami, Dovidio, & Kamp, 2005). Such analyses might include encoding (Sherman et al., 1997), recall of information (Macrae et al., 1996), behavior during interaction (Dumont et al., 2003), or behavior after a “relaxation” period (Macrae et al., 1994). Indeed, well-intentioned efforts to reduce stereotyping and discrimination may backfire on some of these levels under certain circumstances, and it remains a challenge to develop strategies that circumvent this kind of backlash.

The importance of fairness instruction is well recognised in practice. For example, soccer associations offer referees with special trainings to make them aware of the possible influence of stereotypes and to motivate them to avoid such biases (Halm, Metag, & Gosejohann, 2004). However, research has shown that such

instructions may backfire. Further research is needed to identify the mechanisms to support the positive effects of such instructions which were besides the negative effects observed in the current study.

References

- Aberson, C. L., & Ettlin, T. (2004). Meta analytic evidence of two types of favoritism: African American favoring responses and aversive racism. *Social Justice Research, 17*, 25-46.
- Dumont, M., Yzerbyt, V. Y., Snyder, M., Mathieu, B., Comblain, C., & Scaillet, N. (2003). Suppression and hypothesis testing: Does suppressing stereotypes during interactions help to avoid confirmation biases? *European Journal of Social Psychology, 33*, 659-677.
- Galinsky, A. D., & Moskowitz, G. B. (2007). Further ironies of suppression: Stereotype and counterstereotype accessibility following suppression. *Journal of Experimental Social Psychology, 43*, 833-841.
- Gray, D. B. & Ashmore, R. D. (1976). Biasing influence of defendants' characteristics on simulated sentencing. *Psychological Reports, 38*, 727-738.
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424-429.
- Halm, D., Metag, H., & Gosejohann, S. (2004). Förderung von Kommunikations- und Handlungstechniken bei Fußballschiedsrichtern in spielbedingten Konfliktsituationen [Promotion of communication and action techniques for soccer referees in conflict situations during play]. *Journal für Konflikt- und Gewaltforschung, 6*, 19-40.
- Hastorf, A. H. & Cantrill, H. (1954). They saw a game: A case study, *Journal of Abnormal and Social Psychology, 49*, 129-134.
- Kawakami, K., Dovidio, J.F., & van Kamp, S. (2005). Kicking the habit: Effects of nonstereotypic association training on the application of stereotypes. *Journal of Experimental Social Psychology, 41*, 68-75.

Kawakami, K., Dovidio, J.F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, 78, 871-888.

Kite, M. E., Russo, N. F., Brehm, S. S., Fouad, N. A., Hall, C. I., Hyde, J. S., & Keita, G. P. (2001). Women psychologists in academe: Mixed progress, unwarranted complacency. *American Psychologist*, 56, 1080-1098.

Koomen, W. & Dijker, A. J. (1997). In-group and out-group stereotypes and selective processing. *European Journal of Social Psychology*, 27, 589-601.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Ford, R. L. (1997). On the regulation of recollection: The intentional forgetting of stereotypical memories. *Journal of Personality and Social Psychology*, 72, 709-719.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, 67, 808-817.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition*, 14, 1-20.

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, 83, 1029-1050.

Ostrom, T. M., & Sedikides, C. (1992). The outgroup homogeneity effect in natural and minimal groups. *Psychological Bulletin*, 112, 536-552.

Pilz, G. (2000). *Fußball und Gewalt – Auswertung der Verwaltungsentscheide und Sportgerichtsurteile im Bereich des Niedersächsischen Fußballverbandes Saison 1998-1999* [Soccer and violence – Archival analysis of sports court decisions

in Lower Saxony for the season 1998 to 1999]. Unpublished manuscript, University of Hannover.

Rojahn, K. & Pettigrew, T. F. (1992). Memory for schema-relevant information: A meta-analytic resolution. *British Journal of Social Psychology*, 31, 81-109.

Scarabis, M., Florack, A., & Gosejohann, S. (2003). *Implicit associations between Turks and unfair behavior in soccer*. Unpublished raw data, University of Münster.

Shaffer, D. R. & Collier, C. A. (2002). Activation and use of stereotyped beliefs in personnel decisions: A mock (football) draft. In S. P. Shohov (Ed.), *Advances in psychology research* (Vol. 14, 186-195). Nova Science Publishers: Hauppauge.

Sherif, M. & Sherif, C.W. (1953). *Groups in harmony and tension: An integration of studies on intergroup relations*. New York: Octagon.

Sherman, J. W. & Frost, L. A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin*, 26, 26-34.

Sherman, J. W., Stroessner, S. J., Loftus, S. T., & DeGuzman, G. (1997). Stereotype suppression and recognition memory for stereotypical and non-stereotypical information. *Social Cognition*, 15, 205-215.

Simon, B. (1992). The perception of ingroup and outgroup homogeneity: Reintroducing the intergroup context. *European Review of Social Psychology*, 3, 1-30.

Sporer, S. L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law*, 7, 36-97.

Stangor, C. & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social development

literatures, *Psychological Bulletin*, 111, 42-61.

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.

Weber, R., & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology*, 45, 961-977.

Wegener, D. T., & Petty, R. E. (1997). The Flexible Correction Model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141-208). Mahwah: Erlbaum.

Wegener, D. T., Dunn, M., & Tokusato, D. (2001). The Flexible Correction Model: Phenomenology and the use of naive theories in avoiding or removing bias. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 277-290). Mahwah: Erlbaum.

Wegner, D.M. (1994) Ironic processes of mental control. *Psychological Review*, 101, 34-52.

Wilder, D. A. (1984). Intergroup contact: The typical member and the exception to the rule. *Journal of Experimental Social Psychology*, 20, 177-194.

Appendix

Example for vignettes depicting fair behavior

The game is over after 90 exhausting minutes. The match winner is determined. Both teams didn't give up and fought for the win until the very end. ... [*Turkish or German name*] congratulates every player of the winning rival team.

... [*Turkish or German name*] accidentally hits his opposite number's leg. Despite not being punished by the referee, he instantly leaves the ball to the rival player.

Example for a vignettes depicting unfair behavior

... [*Turkish or German name*] commits an assault behind the referees back. Without a cause, he hits his opposite number directly in the face.

Directly after a corner ball, [*Turkish or German name*] bounces his opposite number. So he receives the ball and manages to score a goal.

Footnotes

$$^1 A' = 0.5 + ((\text{hits} - \text{false alarms}) * (1 + \text{hits} - \text{false alarms}) / 4 * \text{hits} * (1 - \text{false alarms}))$$

Author Note

Martin Scarabis, Psychology Department, University of Muenster, Germany.

Arnd Florack, Psychology Department, University of Basel, Switzerland.

The research reported in this article was supported by a research grant from the Deutsche Forschungsgemeinschaft to the first and second author (SC 27/1-2).

Correspondence concerning this article should be addressed to Martin Scarabis, Psychology Department, University of Muenster, Fliednerstrasse 21, 48149 Muenster, Germany. Electronic mail may be sent to scarabis@psy.uni-muenster.de.

Table 1

Recognition Sensitivity as a Function of Fairness Instruction and Group Membership of the Targets

		Recognition sensitivity (A')	
		Turkish targets	German targets
Experimental condition	No fairness instruction	.56	.61
	Fairness instruction	.61	.55