

# The Impact of Counterstereotypic Training and Related Correction Processes on the Application of Stereotypes

**Kerry Kawakami**

*York University*

**John F. Dovidio**

*University of Connecticut*

**Simone van Kamp**

*University of Nijmegen*

The present research examined the impact of counterstereotypic training on the application of stereotypes and the moderating effects of correction on these processes. As expected, when receiving no training, participants chose male over female candidates for a supervisory position and rated both male and female candidates as more gender stereotypic. After receiving extensive counterstereotypic association training, however, participants no longer preferred male over female job candidates and no longer attributed stereotypic traits to a greater extent. These latter results, however, were only found after participants had an opportunity to correct for perceived influences on an initial task. These findings provide evidence for the potential moderating effects of correction processes on the success of strategies aimed at decreasing intergroup biases.

**KEYWORDS** correction processes, discrimination, gender bias, stereotyping

STEREOTYPING involves two separate processes, stereotype *activation* and *application* (Gilbert & Hixon, 1991). A stereotype refers to the association of specific traits, roles, and characteristics with a person or a group based on group membership (Dovidio, Brigham, Johnson, & Gaertner, 1996; Stangor & Lange, 1994). While stereotype *activation* relates to the cognitive accessibility of these associations, stereotype *application* represents the use of stereotypic associations in making decisions about a person or a group.

Although stereotype activation can at times be directly related to the application of stereotypes (Kawakami, Dion, & Dovidio, 1998; Lepore & Brown, 1997), the relationship between stereotype activation and application is not

## **Author's note**

Address correspondence to Kerry Kawakami, Department of Psychology, York University, 4700 Keele Street, Toronto, Canada, M3J 1P3 [email: kawakami@yorku.ca]

always so straightforward. The application of stereotypes is frequently moderated by personal standards, general norms, and lay theories about the appropriateness of these beliefs and their expression (Monin & Miller, 2001). Devine (1989), for example, demonstrated that although both high and low prejudiced people activated stereotypes equivalently, low prejudiced people overtly applied stereotypes to a lesser extent than did high prejudiced people. Additional work has demonstrated that people may be able to limit the effect of stereotype activation on their behaviors when they have the motivation and opportunity to control their responses (Dovidio & Fazio, 1992; Kawakami, Dovidio, & Spears, 2005; Kawakami, Spears, & Dovidio, 2002; Monteith, 1993). As a consequence, stereotype activation does not necessarily translate directly to deliberative behaviors and the use of stereotypes (Dovidio, Kawakami, & Gaertner, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995). In particular, when people are motivated and capable of exerting effort, they may correct for the initial activation of stereotypes by suppressing this activation so that they are not applied in subsequent judgments. The present research extends this previous work on stereotype activation and application by further investigating how lay theories related to correction processes can influence the application of stereotypes in a way that *increases* rather than *decreases* their use.

Lay theories are organized knowledge structures that help people understand, predict, and control their environment (Hong, Levy, & Chiu, 2001; Kruglanski, 1990). These naïve theories can direct behaviors, judgments, and evaluations about social groups. One important lay theory, the Flexible Correction Model, is related to how people perceive and correct for the influence of other people and stimuli on their responses (Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994). Specifically, this theory proposes that if people perceive themselves to be unduly influenced, they will attempt to correct for these influences by modifying their assessments in a direction opposite to the perceived bias (Martin, Seta, & Crelia, 1990; Mussweiler & Neumann, 2000).

In accordance with earlier assumptions of reactance theory (Brehm & Brehm, 1981), more recent correction theories suggest that if people think they are being overtly influenced, they may react and reassert what they believe to be their initial true response (Schwarz & Bless, 1992; Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994). Although both reactance and correction theories can be considered lay theories of influence, they differ substantially in their underlying assumptions and implications. Whereas reactance theory posits that people, often without full deliberation, will react *against* threats that limit their personal sense of behavioral freedom, correction theory emphasizes more specific, deliberative, and strategic adjustments. In particular, correction theory proposes that people do not simply resist attempts at control, but their intent is to correct for unwanted influence and they respond in a calibrated fashion. In particular, they form naïve theories about the direction and the extent of potential influence and then adjust their responses to compensate for this impact (Mussweiler & Neumann, 2000; Wegener & Petty, 1997; Wilson & Brekke, 1994). Thus, whereas reactance to the loss of one's personal control is somewhat automatic and diffuse, correction is a more cognitively demanding process that involves a more developed personal theory about the impact of influence attempts and specific ways to counteract them. Furthermore, correction theory suggests that this process will only occur when people are motivated to correct their responses and have the cognitive capacity to engage in correction.

The present research explores the effects of mental correction on intergroup behavior in an attempt to extend theorizing related to correction processes in important ways. Whereas most research on correction processes addresses adjusting for priming or context effects in impression formation (e.g. Martin et al., 1990; Petty & Wegener, 1993) or correction for perceived social influences (e.g. Petty, Wegener, & White, 1998; Wilson, Houston, & Myers, 1998), the present research examines correction for interventions aimed at ameliorating stereotyping. This study, therefore, has the potential

to broaden the scope of lay theories by applying correction theory to an area of intergroup relations, intervention strategies, not often framed in terms of lay theories. Our research hopes to demonstrate the theoretical value of adopting correction theorizing to these processes by directly addressing the intricacies of how and when stereotype activation leads to stereotype use and when it does not.

Whereas previous research on efforts to exert control over stereotyping has focused on how people can make 'corrections' to reduce the impact of stereotype activation on its application (e.g. Devine, 1989), recent research suggests that correction processes can also increase the application of stereotypes. In particular, these findings suggest that people may correct not only for perceived biases against specific social categories but also for interventions that are aimed at reducing these biases. Correcting for the influence of anti-bias procedures on one's attitudes can significantly moderate the effectiveness of these strategies and often produce what appears to be 'backlash'. The primary goal of the present research is to examine an intervention aimed at reducing the activation of stereotypes on the use of stereotypes when evaluating a category group member and the impact of related corrective actions on this process.

Recent research by Kawakami, Dovidio, and van Kamp (2005) demonstrates that, when able, people will correct for the perceived influence of a training session intended to reduce intergroup biases by increasing the subsequent application of stereotypes. That is, if people perceive that strategies aimed at reducing stereotyping may be influencing their responses, they may attempt to reassert their prior stereotypic associations—thereby correcting for the perceived influence of the intervention. Specifically, Kawakami, Dovidio, and van Kamp (2005) found that after engaging in extensive counterstereotypic training, a procedure demonstrated by Kawakami, Dovidio, Moll, Hermsen, and Russin (2000) to eliminate the automatic activation of stereotypes, participants still showed significant bias against women in hiring.

Kawakami, Dovidio, and van Kamp (2005) suggested that the reason for the apparent

ineffectiveness of the training on the hiring decisions was due to correction processes. Consistent with the interpretation that participants were strategically correcting for the effect of the training, the results revealed that the level of discrimination was comparable to the effect for the control groups (no training and stereotype-consistent training). Also, the effects of correction disappeared when participants' capacity to control their responses was impaired. That is, training to eliminate stereotype activation was effective at reducing the application of stereotypes, but only when participants were inhibited from engaging in correction processes by decreasing their cognitive capacity with an additional probe reaction task (Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994).

The present research was designed to further explore the effects of mental correction on intergroup relations and extend the work of Kawakami, Dovidio, and van Kamp (2005) by examining the impact of correction processes over multiple tasks. Whereas the Kawakami, Dovidio, and van Kamp (2005) studies investigated the impact of cognitive capacity on how correction processes mitigate the effectiveness of interventions to reduce intergroup bias on one index of stereotype application (i.e. the hiring of a woman for a supervisory position), the current study includes two separate types of stereotype application. By including two indices of stereotyping, the present work has the potential to investigate the impact of opportunities to correct for interventions such as training on an initial task of stereotyping and on subsequent tasks.

Like many other motives, a motive to correct for bias may vary in strength based on a person's opportunity to satisfy the motive. According to Lewin's (1951) classic field theory, needs create a state of tension which are alleviated when the needs are satisfied. Lewin's theory and more recent research suggest that active goals and needs enhance accessibility of need-related concepts and the drive to fulfill these goals. Achievement of a goal, through any of a variety of different reasonable means (Lewin, 1935; Shah, Kruglanski, & Friedman, 2002), reduces the accessibility of the goal and the need to fulfill it (Bargh, Gollwitzer, Lee-Chai,

Barndollan, & Trotschel, 2001; Liberman & Forster, 2000; Wicklund & Gollwitzer, 1982). Following this logic, if training induces a need to correct for the perceived influence of an anti-bias intervention, then letting people correct on an initial task related to the application of stereotypes would satisfy this need and reduce the motivation to correct for the influence of the intervention on a subsequent task related to the application of stereotypes. With respect to the present study, counterstereotypic association training would therefore not initially positively impact stereotyping because of people's attempts to correct for the influence of this training. However, once participants had had the opportunity to correct, the training would prove to be more effective on subsequent tasks.

In summary, a primary aim of the present study was to extend previous theorizing and research on correction processes and the application of stereotypes by examining the impact of correction processes on participants' use of stereotypes when multiple opportunities to correct for anti-bias interventions are present. Following the procedure used by Kawakami, Dovidio, and van Kamp (2005), participants were trained to associate counterstereotypic characteristics with men and women. Specifically, participants in the Training condition were presented with photographs of men and women with stereotypic and counterstereotypic traits under each photograph, and were instructed to choose the trait that was not culturally associated with the category in the photograph. The control condition in the current study was modeled after the No Training condition in the Kawakami, Dovidio, and van Kamp (2005) study in which participants proceeded directly to the job applicant evaluation phase of the research. In the applicant evaluation phase, participants engaged in two tasks involving the application of stereotypes in sequence. In counterbalanced order across groups, participants made hiring decisions about two male and two female job candidates and attributed stereotypic and non-stereotypic traits to male and female candidates.

We hypothesized, in accordance with Kawakami, Dovidio, and van Kamp (2005), that correction

processes would counteract the perceived effect of counterstereotypic association training (Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994). Moreover, we extended theorizing related to the previous research by hypothesizing that the impact of these correction processes and the training would vary depending on whether the participants had an initial opportunity to correct for the training or not. In particular, we proposed that participants would correct for the perceived effects of the training on an initial task of stereotype application. Consistent with this hypothesis, participants in a pilot study indicated, after reading the procedure used in the present study, that they believed the purpose of the training was to change beliefs about men and women and that, in general, people would try to resist this influence. However, we further hypothesized that once participants had an initial opportunity to correct their responses for the perceived influence of training and satisfy their motivation to respond against the influence of others, correction would no longer occur on subsequent tasks. The effectiveness of counterstereotypic association training would therefore become evident and the use of stereotypes would decrease on the second stereotype application task. Thus, our main prediction was that the Counter Stereotypic Association Training and the No Training conditions would show comparable evidence of gender bias on the first task (hiring or trait attributions), but that on the second task (trait attributions or hiring), the Counter Stereotypic Association Training condition would show less gender bias than the No Training condition, and no significant bias against women overall.

## **Method**

### ***Participants and design***

A total of 156 (111 female and 45 male) undergraduate students in the Netherlands took part in the experiment and were paid approximately US\$6. These participants were randomly assigned to one of eight conditions within a 2 (Training: Counter Stereotypic Association Training vs. No Training)  $\times$  2 (Order of Tasks: Candidate Choice Task First vs. Second)  $\times$  2

(Instruction: Attend to Leadership vs. General Impression) between-subjects design.

### **Procedure**

Upon arrival, participants were informed that they would be participating in two studies. The primary aim of the first phase was to train participants to associate counterstereotypic words with specific social categories. Whereas approximately half of the participants received training in making new associations with male and female photographs, the other half of the participants did not receive this training. The goal of the second phase was to examine the effect of the training on the application of stereotypes. All participants in this phase were asked to read four applications for a supervisory position. While half of the participants were instructed to first select one applicant who they thought would be most suitable for the job and then to rate each applicant on a series of job related traits, the other half of the participants were instructed to complete these tasks in the reverse order. First, they were presented with the trait rating task and then they were asked to choose the best candidate.

In the *training phase* of the study, upon entering the laboratory participants were randomly assigned to either the Counter Stereotypic Association Training condition or the No Training control condition. Participants in the Counter Stereotypic Association Training condition were seated behind a Macintosh Performa computer and a button box and informed that they would be presented with photographs of men and women. Furthermore, they were informed that underneath each photograph two traits would appear—one on the left and one on the right. They were instructed to choose the trait that was not associated in our society with the gender category in the photograph and to press the appropriate button on the left or the right side of the button box. For half of the trials, the correct answer was on the right side of the button box; for the other half of the trials, the correct answer was on the left side of the button box.

For example, when participants were presented with a photograph of a woman with the traits 'sensitive' and 'strong', the trait 'strong'

is the correct answer because strong is not culturally associated with women. Alternatively, if participants were presented with a photograph of a man with the traits 'weak' and 'sloppy', the trait 'weak' is the correct answer because that is the trait that is not culturally associated with men.

Specifically, on each trial the photograph and the traits were presented simultaneously on the computer screen and remained until the participant responded. After each response, a blank screen appeared for 1500 ms before the presentation of the next photograph and traits. In total, participants received 480 trials consisting of six blocks of eighty trials. In each block, 20 black and white photographs of men and women scanned from college yearbooks were presented with stereotypes of men and women under the photograph. On each trial, both stereotypes were either positive or negative. In total, 20 stereotypes of men and women were utilized. For a list of the specific traits used in both the training and job applicant phases, please refer to the appendix.

All trials in each block were presented in a random order. After each block, participants were given a break and were asked to press the mouse when they were ready to continue the experiment. Participants were instructed to complete each trial as quickly and as accurately as possible. Before beginning the actual trials, however, participants were first presented with a practice set of eight trials involving stimuli not used in the experiment.

To examine the effects of the training on the use of stereotypes, all participants were presented with the second job application phase. In this phase, as in the Kawakami, Dovidio, and van Kamp (2005) article, participants were told that a neighboring university was carrying out a national study and had asked their psychology department to help investigate job hiring procedures. Specifically, participants were presented with a real advertisement for a chairperson of a District Doctor's Association selected from a national newspaper. They were asked to read the advertisement along with four ostensibly real résumés and cover letters for the job.

The advertisement described a leadership job in which the successful candidate would supervise doctors in his or her district. The candidate would advise policy on financial, organizational, and juridical affairs, as well as supervise and set into motion action plans for health care in the district. The candidate was required to have strong management skills and to be able to network well with people in his or her field and to supervise the office staff. The function required an academic level of thinking and working, management qualities and experience, and inspirational leadership skills. The candidate was expected to be a convincing negotiator, and be flexible, creative, and have vision.

The résumés and cover letters of four candidates briefly described their education and experience. Materials were developed such that all applicants were suitable for the advertised job. Two of the applicants, however, were given male names (i.e. Sander van Cleef and Anton Fleuren), and two of the applicants were given female names (i.e. Marian Koenders and Helma Driessen). It is important to note that the specific application and the sex of the applicant were counterbalanced. Specifically, for half of the participants, Applications A and B were women and C and D were men. For the other half of the participants, Application A and B were men and C and D were women. Participants were presented with one of four different orders of the applications.

After reading all the applications, participants were presented with one of two possible orders of the tasks. Specifically, some of the participants were first asked to choose the best applicant for the job. Next, these participants were asked to judge each applicant on 16 ostensibly job-related traits on a scale that ranged from 1 (totally not applicable) to 9 (totally applicable). Eight of the traits were female stereotypes (e.g. helpful and irrational) and eight of the traits were male stereotypes (e.g. risk-taker and intimidating). None of the traits utilized in the job application phase were included in the training phase. The rest of the participants were given the same two tasks but in the reverse order. Specifically, these participants were first asked to rate

each candidate on the 16 traits and then asked to choose the best applicant.

Because previous research has revealed that gender bias is greater when the characteristics associated with a position are more gender linked (Nelson, 2001), in one condition we simply asked participants to choose the best candidate, whereas in another condition, to further increase the salience of the task as masculine, we instructed participants to choose the best candidate *on the basis of leadership qualities*.

The stereotypes used in both the training and job application phase were chosen on the basis of a pilot study ( $N = 50$ ) that evaluated the valence and attribution of traits to men and women on 9-point scales. The male and female stereotypes were comparable in both word length (both  $M_s = 7.7$ ) and valence ( $M_s = 5.17$  and  $5.64$ , respectively). Furthermore, male stereotypes were attributed more to men ( $M = 7.28$ ) than women ( $M = 4.11$ ). Female stereotypes, alternatively, were attributed more to women ( $M = 7.38$ ) than men ( $M = 4.10$ ).

## Results

The initial analyses examined the speed of participants' responding over time in the Counter Stereotypic Association Training condition to assess their learning of the new associations (see Kawakami, Dovidio, & van Kamp, 2005; Kawakami et al., 2000). Subsequent sets of analyses considered the application of stereotypes in both the choice of candidate and the trait attributions. While the choice of the best candidate was a dichotomous measure (i.e. recommended hiring of either a man or a woman), the attribution of masculine and feminine traits to male and female candidates was a continuous measure.

### *Speed of response on the Counter Stereotypic Association Task*

Before the response data on the Counter Stereotypic Association Task were analyzed, response latencies in which participants gave incorrect answers (6.38%) and outlier latencies that were more than three standard deviations from the mean (1.87%) were excluded (Ratcliff, 1993).

For each participant, the mean values associated with the response latencies for the 40 trials for male and female photographs in each of the six blocks of trials were computed. Because initial analyses demonstrated no systematic effects for sex of participant on the training data, this variable was not included in the subsequent analyses.

To examine the effect of the amount of training on speed of responding, a 2 (Photograph: Male vs. Female)  $\times$  6 (Blocks)  $\times$  2 (Order of Task: Candidate Choice First vs. Second)  $\times$  2 (Instructions: Attend to Leadership vs. General Impression) analysis of variance was performed on the response latencies with the Photograph and Block variables as repeated measures. The Block variable was analyzed in terms of linear trends.

A significant effect for type of Photograph was found ( $F(1, 66) = 9.79, p < .01$ ). Participants responded faster in general to female ( $M = 1716$ ) in comparison to male ( $M = 1778$ ) photographs. As expected, a significant linear effect for Block was also found ( $F(1, 69) = 81.13, p < .001$ ). The response latencies reflect a strong and consistent learning effect from block 1 to block 6 ( $M_s = 2021, 1869, 1769, 1694, 1607, 1525$ ). With more training, participants became faster in associating

counterstereotypes with male and female categories. These findings replicate earlier results related to counterstereotypic training (Kawakami, Dovidio, & van Kamp, 2005; Kawakami et al., 2000) and underline the importance of practice for learning new associations to existing social categories.

### Choice of the candidate

Preliminary analyses revealed no systematic effects of sex of participant on candidate selection decisions, and thus this variable was not included in subsequent analyses. The independent variables in the main analysis were 2 (Training: Training vs. No Training), 2 (Order of Tasks: Candidate Choice Task First vs. Second), and 2 (Instructions: Attend to Leadership vs. General Impression). Because choice is a dichotomous dependent measure, the effects of these variables and their interactions were analyzed using logistic regression and chi-square analyses.

There was no main effect for Training ( $B = 0.11, df = 1, \text{Wald statistic} = 0.42, p = .52$ ), or any other variables. However, a Training  $\times$  Order interaction, was obtained ( $B = 0.33, df = 1, \text{Wald statistic} = 3.62, p = .05$ ). As indicated in Figure 1,

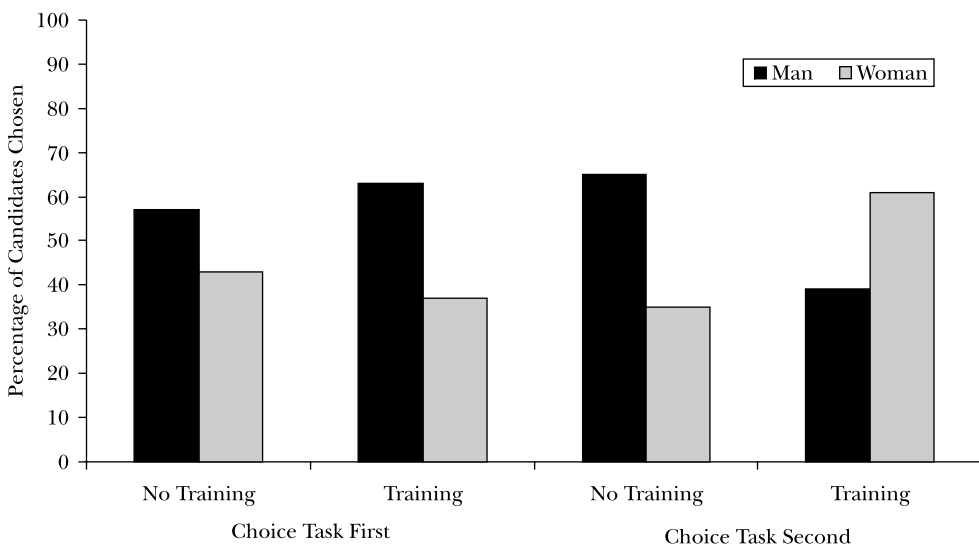


Figure 1. Effects of counterstereotypic training and order of tasks on choice of male and female candidates.

the effect of training was moderated by whether participants performed the candidate Choice Task first or second.

In particular, when the Choice Task was first, *before* the Trait Task, there was no difference between the Counter Stereotypic Association Training and No Training Control conditions in the extent to which male and female candidates were selected for the position ( $\chi^2(1, N = 102) = .37, p = .54$ ). In the No Training condition, 43% of the participants selected a woman and 57% selected a man. In the Counter Stereotypic Association Training condition, 37% of the participants selected a woman and 63% selected a man. Thus overall, across both training conditions, women were selected significantly less than men, 40% vs. 60% ( $\chi^2(1, N = 102) = 3.92, p < .05$ ).

In contrast, when the Choice Task was performed second, *after* the Trait Task, participants in the Counter Stereotypic Association Training and the No Training conditions differed in their selection of male and female candidates for the leadership position ( $\chi^2(1, N = 54) = 3.71, p = .05$ ). In the No Training condition, only 35% of the participants selected a woman and 65% selected a man. These percentages are similar to the selection pattern in the Counter Stereotypic Association Training and the No Training conditions when the Choice Task was first (i.e. 40% vs. 60%). However, when the Choice Task was second and after the Counter Stereotypic Association Training condition, 61% of the participants selected a woman and only 39% selected a man. The Counter Stereotypic Association training was therefore successful at reducing participants' preference for men over women when the choice of the candidate did not directly follow the training task. Although, as expected, the difference in the extent to which men and women were chosen for the job was no longer significant ( $\chi^2(1, N = 31) = 1.58, p = .21$ ), it is important to note that the percentages related to this gender bias are similar to when the Choice Task was first (40% vs. 60%). However, when the Choice Task is second, this bias is reversed and favors women over men. As predicted, the order in which participants performed the tasks critically moderated the impact

of counterstereotypic association training on hiring decisions.

### **Trait judgments**

Preliminary analyses of the trait judgments included a 2 (Participant Sex: Male vs. Female)  $\times$  2 (Job Candidate Sex: Male vs. Female)  $\times$  2 (Training: Training vs. No Training)  $\times$  2 (Order: Candidate Choice First vs. Second)  $\times$  2 (Instructions: Attend to Leadership vs. General Impression)  $\times$  2 (Trait: Stereotypic vs. Nonstereotypic) analysis of variance with repeated measures on the Job Candidate Sex and Trait variables. Because no systematic effects were found associated with sex of participant, this variable was not included in subsequent analyses. Also because no significant effects were found for job candidate sex that were of theoretical importance, and because parallel patterns for male and female applicants were obtained for comparisons between responses to stereotypic associations (i.e. feminine traits for women and masculine traits for men) and non-stereotypic associations (i.e. feminine traits for men and masculine traits for women), this variable was also excluded. Consequently, the results reported below are related to a 2 (Training)  $\times$  2 (Order)  $\times$  2 (Instructions)  $\times$  2 (Trait) analysis of variance with repeated measures on the last independent variable.

This analysis revealed a main effect for Trait ( $F(1, 148) = 10.42, p < .01$ ). In general, participants attributed stereotypic traits ( $M = 5.41$ ) in comparison to nonstereotypic traits ( $M = 5.14$ ) to a greater extent to the candidates. This effect, however, was qualified by a significant Training  $\times$  Trait  $\times$  Order of Task interaction ( $F(1, 148) = 5.82, p < .05$ ). In accordance with the results for the Choice Task, this three-way interaction was pursued by examining the effects of the Training on each order of tasks separately.

An examination of the results related to the participants who completed the Choice Task second, and therefore rated each candidate on the 16 traits first, demonstrated a marginally significant effect for Trait ( $F(1, 52) = 3.58, p = .06$ ). As indicated in Figure 2, these participants tended to attribute stereotypic traits ( $M = 5.51$ ) in comparison to nonstereotypic traits



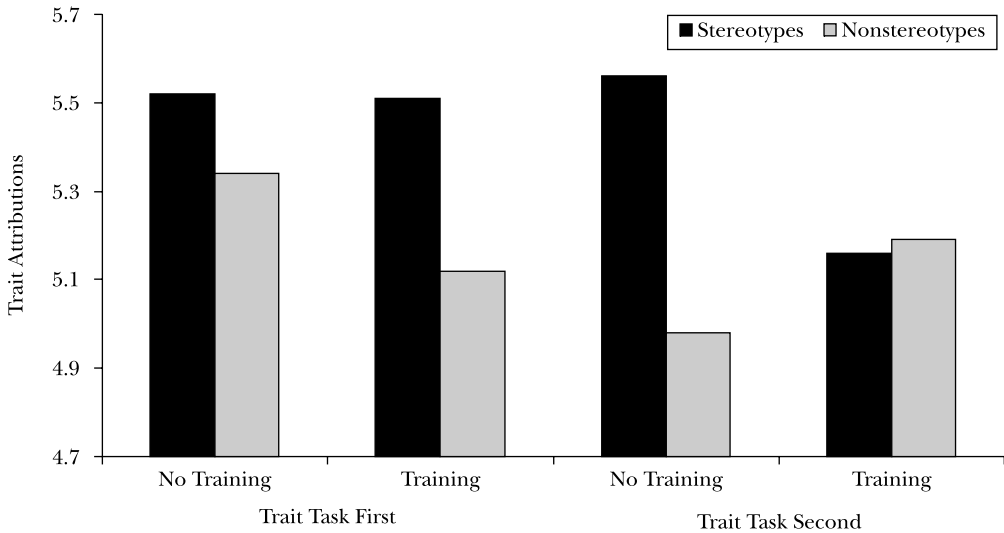


Figure 2. Effects of counterstereotypic training and order of tasks on attributions of stereotypic and nonstereotypic traits to male and female candidates.

( $M = 5.25$ ) to a greater extent to the candidates. Consistent with the results of the Choice Task, training had no effect on trait attributions when the Trait Task was first. Specifically, the Training  $\times$  Trait interaction was not significant ( $F(1, 52) = .56, p = .46$ ).

An examination of results for participants who completed the Choice Task first, before rating the applicants on the 16 traits, however, demonstrated a significant Training  $\times$  Trait interaction ( $F(1, 100) = 7.69, p < .01$ ). As predicted, participants who did not receive training attributed stereotypic traits ( $M = 5.56$ ) in comparison to nonstereotypic traits ( $M = 4.98$ ) to a greater extent to the applicants ( $t(50) = 4.98, p < .001$ ). In contrast, participants who received extensive counterstereotypic association training did not differentiate in the extent to which they attributed stereotypic traits ( $M = 5.16$ ) and nonstereotypic traits ( $M = 5.19$ ), ( $t(50) = .19, p = .85$ ).

Examining the Trait  $\times$  Training interaction from a different perspective further shows that for participants who completed the Choice Task first before the Trait Task, no difference was found between participants who received training

( $M = 5.19$ ) and those who did not receive training ( $M = 4.98$ ) on the attribution of nonstereotypic traits ( $t(100) = 1.52, p = .13$ ). However, these participants did differ in stereotypic traits. Specifically, participants who received training ( $M = 5.16$ ) attributed stereotypic traits to male and female candidates to a lesser extent than did participants who did not receive training ( $M = 5.56$ ), ( $t(100) = 2.49, p < .05$ ). Thus, analogous to the results for the Choice Task, training had no effect on trait attributions when the Trait Task immediately followed the training, but training did have the anticipated effect when the Trait Task was completed after the Choice Task.

## Discussion

In general, the results of the present research support the hypothesis that correction is a deliberate and calibrated process that people use strategically to compensate for undesired external influence. Whereas most previous research on correction processes has focused on adjusting for priming or context effects in impression formation (e.g. Martin et al., 1990;

Petty & Wegener, 1993) or correction for perceived social influences (e.g. Petty et al., 1998; Wilson et al., 1998), the present research examined correction for interventions aimed at ameliorating stereotyping. Furthermore, while Kawakami et al.'s (2000) original studies on the training procedure demonstrated that extensive practice associating counterstereotypic characteristics with social categories can reduce the *activation* of stereotypes, more recent research (Kawakami, Dovidio, & van Kamp, 2005) has revealed that training can also, at least under some circumstances, reduce the *application* of stereotypes. Specifically, the latter research demonstrated a decrease in gender bias in hiring decisions after extensive training, but only when the possibility for correction processes was reduced. For example, when cognitive demand was high and participants were required to perform a secondary probe-reaction task.

The present study extends this previous work on the impact of correction processes and counterstereotypic association training on stereotyping in significant ways. Specifically, the present results provide evidence for the generalizability of the effects related to both training and correction processes by replicating the findings for personnel selection and showing parallel effects for trait attributions. Training and correction processes not only influenced general decisions to hire or not hire a male or a female candidate for a leadership position but also influenced the extent to which particular gender traits were ascribed to men and women. For example, in the present study, training actually changed the extent to which a female candidate was perceived to be helpful or submissive and a male candidate was perceived to be competitive or determined. Even though the information related to each candidate (i.e. the résumés and cover letters) remained the same across training conditions, participants evaluated and interpreted the information differently after training.

We also found that the effects of training on reducing gender biases were not evident on all stereotyping tasks that followed the training. Participants did not reduce their stereotyping

on the first task regardless of whether this task involved candidate selection or trait attributions. As we proposed, when an intervention aimed at reducing intergroup bias, such as counterstereotypic association training, is directly followed by the stereotyping target task, participants may assume that the intervention unduly influenced their judgments of group members. They, therefore, may modify their assessments of the group members (e.g. job applicants) in a more biased direction on an initial task to adjust for this influence. By correcting for the perceived influence of training, participants may think that they are judging others more objectively and according to their own perceptions, even though it involves the use of stereotypes. In contrast to studies of stereotype activation that utilize automatic processing conditions (e.g. Kawakami et al., 2000), the more controlled processing conditions related to the application of stereotypes in the present study allowed participants to successfully modify their answers to decrease the perceived influence of the training.

However, the effects of the training had a differential impact depending on whether the stereotyping task immediately followed the training or not. Importantly, the present findings extend theorizing on correction processes by demonstrating that once participants have the opportunity to adjust their responses according to the perceived influence of the training, subsequent motivation to correct and related correction processes may decrease (Lewin, 1951; Liberman & Forster, 2000; Wicklund & Gollwitzer, 1982). In particular, in the present study once participants had initially corrected for the training on the first stereotyping task, regardless of whether it involved candidate selection or trait attributions, they were subsequently less inclined to correct their responses on the second stereotyping task, even though the first and second tasks were conceptually similar and both were related to the application of stereotypes.

In accordance with theorizing related to the motivational properties of goals, these results suggest that fulfillment of a goal may signify to people that subsequent activation of the goal

construct is no longer necessary or useful and therefore this goal will be inhibited (Lieberman & Forster, 2000; Zacks & Hasher, 1994). After the need to correct for the influence of the anti-bias intervention is satisfied by one task related to the application of stereotypes, the need to correct on subsequent tasks related to stereotyping will dissipate because of means substitution (Lewin, 1935; Shah et al., 2002). Any activities that are perceived to be related to the same goals can substitute for one another and therefore can release tension and related motivation to fulfill the same goal system.

Recent studies related to research on affirmation of specific aspects of the self-concept provide further evidence in how fulfilling a temporary motivation can reduce efforts on subsequent tasks. In particular, this research has revealed that when aspects of the self are threatened, people are motivated to reaffirm their self-concept (Steele, 1999; Steele & Lui, 1983). However, once this self-affirmation has occurred, participants are less likely to be influenced by motivations related to this threat. For example, whereas people will cling to their original views when they feel threatened and reassert these beliefs despite considerable evidence that disconfirms them, immediately after an important aspect of their self has been affirmed they are more open and susceptible to being persuaded (Cohen, Aronson, & Steele, 2000; Correll, Spencer, & Zanna, 2004).

Although the present pattern of findings is clearly compatible with theorizing related to the impact of goal fulfillment on subsequent goal-related efforts and consistent with current theorizing related to correction processes (Kawakami, Dovidio, & van Kamp, 2005; Lieberman & Forster, 2000; Steele, 1999; Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994), future research might productively focus on identifying more fully the dynamics of anti-bias interventions and correction processes on stereotyping. In particular, there are several alternative explanations related to the present findings that warrant further investigation.

One alternative explanation is related to the passage of time. It is possible that correction processes dissipate not directly because of the

opportunity to correct, as hypothesized, but simply with the passage of time. With more time, the biasing influence of the training task may be less apparent, and so people will be less inclined to correct for this influence. Although this explanation is quite plausible when long delays occur between training and the first opportunity to correct for its perceived influence, it does not readily account for the results of the present study. If passage of time was the key element in the current research, then gender bias on the second task would be expected to vary as a function of whether participants completed the relatively brief, single-item Choice Task or the longer, 64-item (rating all four candidates on 16 traits) Trait Task first. Specifically, the time passage explanation suggests that after counterstereotypic association training less bias would be observed when the longer Trait Task was first in comparison to the shorter Choice Task because of the weakening of correction motivation over time. Although it is difficult to assess this possibility given that the two tasks and measures are so different, as illustrated in Figures 1 and 2, the pattern of effects is essentially symmetric. In addition, even when participants have to make only one intervening judgment, to hire a woman or man, this single decision task is enough to eradicate correction processes. Furthermore, if participants' correction on the first item was sufficient to reduce correction processes on subsequent items, we would not have produced the predicted and obtained pattern of effects on the attribution task. If correcting on only one item is enough to alleviate correction motivation, then the lack of correction on all 63 of the remaining traits would make the results look similar across task orders (whether first or second). This was not the case. Participants clearly stereotyped less when the Trait Task was preceded by the Choice Task than when it immediately followed the training. These results suggest that, at least within the context of our relatively brief (one-hour) experimental session, the amount of the time or the number of judgments between the training and the second stereotyping task was less important than the perception that these two tasks involved separate judgments. Participants appeared to shift

judgment strategies across tasks due to something other than time or number of judgments.

What may be more important than the passage of time per se is what occurs during that period of time. According to correction theory, activities that distract people from their goal of correcting for the influence of training involve high cognitive demand that interferes with developing a calibrated response to perceived influence, or enable people to satisfy their goal of correcting for influence should all be effective at eliminating the effect of correction processes and therefore should produce a successful reduction of stereotyping following counterstereotypic training. The Kawakami, Dovidio, and van Kamp (2005) re-search demonstrated that a distraction task (an intervening arithmetic task) and a demanding activity while making an initial response (a probe reaction task) eliminated correction effects. In the present study, we found that whether participants responded in a stereotypic way on the shorter candidate Choice Task or on the longer Trait Task, correction effects were significantly reduced on subsequent tasks. Because both tasks were related, it was unlikely that the first task distracted participants from the goal of correction on the second task. Furthermore, because the tasks were sequential and not simultaneous (as with the probe reaction task), cognitive demand also does not seem to readily account for the results. The most plausible explanation for the present finding is therefore that correction processes dissipated following the first stereotyping task because participants were given the opportunity to correct on this initial task which reduced their motivation to correct on subsequent tasks.

Nevertheless, future research might disentangle these alternative explanations by varying within the same study whether participants who receive extensive counterstereotypic training perform an initial task that is (a) brief and easy, or intensive and demanding, and (b) that is unrelated or related to stereotyping before performing a subsequent stereotyping task. To the extent that either distracting or goal attaining activities can alleviate the effects of correction, participants who perform a brief task that is

unrelated to stereotyping would be expected to show higher levels of bias (due to correction) in comparison to participants in the other three conditions.

It is also possible that the effects of the training on stereotyping in the present study are not due to correction processes related to the first task as hypothesized but to an ineffective training procedure. Specifically, if the training did not have an effect and stereotyping remained the same, responses by participants in the training and the no training condition would not differ on the first stereotyping task and correction processes would not have occurred. Although this possibility would predict the effects related to the first task in the present study, previous research has produced consistent evidence for the ability of the training procedure to reduce the activation and the application of stereotypes (Kawakami, Dovidio, and van Kamp, 2005; Kawakami et al., 2000). Furthermore, research by Kawakami, Dovidio, and van Kamp (2005) provides convincing evidence for the existence of correction processes. Specifically, when cognitive capacity was limited, participants were unable to act according to their correction motivations and the training was shown to be effective in reducing biases on tasks similar to the present stereotyping task.

The alternative explanation that counterstereotypic training was ineffective also does not account for why training was found to be effective on secondary stereotyping tasks. It is possible that the initial stereotyping tasks following the training somehow created an implicit demand to reduce stereotyping on subsequent tasks. Although this explanation is somewhat plausible when the Trait Task is first, it is less plausible when the Choice Task is first. Furthermore, this account is less convincing once one realizes that implicit demand to reduce stereotyping would be much stronger following the training than following a simple trait judgment task. If demand characteristics were a valid explanation for these results, one would expect the greatest reduction in stereotyping to be on the first task immediately following the counterstereotypic association training. This was not the case in the present findings.

Future research should also include different control conditions related to the training and different types of anti-bias interventions. Because the present study included only two training conditions, a Counter Stereotypic Association Training condition and a No Training control condition, it is possible that the results related to the present study may be related to something idiosyncratic about receiving 480 trials of training rather than about learning to associate counterstereotypic traits with specific categories. For example, simply being asked to associate traits to social categories for an extended period of time may activate reactance in the participant so that they act contrary to perceived expectations on subsequent tasks.

This reactance explanation, however, is not entirely consistent with the present findings or previous findings. In contrast to reactance, which is a broad response to threats to personal freedom, correction is a more strategic attempt to compensate for influences that people believe distort their true response. Our findings that the use of stereotypes in the training and no training conditions on the first task showed no difference indicates that when people do correct for the perceived effect of the training, they do so in the opposite direction and use stereotypes to approximately the same extent as those participants who did not receive training. These results suggest the use of fairly sophisticated and accurate naïve theories of influence and they implicate the role of the types of volitional assessment and controlled processing outlined in correction theory.

Furthermore, results related to past research that has included various training conditions would also cast doubt on the possibility that idiosyncratic features of the training per se are a plausible explanation for the current findings. Specifically, these findings demonstrate that while extensive training in associating counterstereotypic traits can reduce the activation of stereotypes related to specific social categories, it does not influence the activation of counterstereotypic new associations (Kawakami et al., 2000). Furthermore, participants who were extensively training in associating stereotypic associations do not differ in the extent to which

they activated or applied stereotypes when compared to participants who received no training (Kawakami, Dovidio, & van Kamp, 2005; Kawakami et al., 2000). Together these findings suggest that the judgments in the present study are not a function of simple reactance or mood effects related to the training per se, or the activation of stereotype inconsistent traits, but are related to the reduced activation of stereotype consistent traits following a particular type of training: counterstereotypic association training.

It is also important to note that whereas different traits were utilized in the training phase and the trait attribution phases of this study, these traits are to some extent related. We assume, however, that the training influences not only traits related to those used in the training but that it induces a broader stereotyping process within the person that influences the interpretation of the information and its association to the target category group member. In accordance with this assumption, the training was shown to influence not only the trait attributions of participants but also the more general global hiring decisions.

Additional research related to the training procedure, however, is clearly warranted. Although it is important to note that to some extent a replication of the order effects in this study are already built-in, in that the same pattern of results related to the first and second task following the training was found regardless of whether the task was the Trait Task or the Choice Task, future research could provide a useful replication of the present findings and could also target specific mechanisms which may underlie the effectiveness of the training.

Further research related not only to the specific training procedure we used but also other types of anti-bias interventions would also be informative. Importantly, our findings have direct implications for the effectiveness of certain types of anti-bias programs. Strong interventions to reduce bias which appear 'heavy-handed' may arouse correction motivations, at least initially, to control for these influences. As Plant and Devine (2001) demonstrated, pro-Black pressure, such as that attributed to 'political

correctness', can initially produce backlash in affective, attitudinal, and behavioral responses. Thus, in terms of the immediate consequences, these interventions may actually be ineffective in reducing bias. However, as illustrated in the present research, the intended effects of anti-bias interventions, such as counterstereotypical training, may subsequently be observed after initial opportunities to correct for this influence. Thus, it is important to consider the short- and long-term effects of anti-bias interventions and the type of intervention to understand the underlying processes that can moderate their effectiveness.

Another potentially productive direction for future research would be to examine anti-bias interventions, stereotype activation, motivations to correct, and stereotype application within the same study. Whereas the present analysis suggests a disassociation between stereotype activation and application, at least on initial tasks following the training, it also predicts that they will be more interrelated on subsequent tasks. In particular, whereas counterstereotypic association training would be expected to reduce stereotype activation in the first task, participants may still use cultural stereotypes in their decisions at this time because of correction processes. In subsequent tasks or when correction processes are not otherwise operating (e.g. because of cognitive demand), processes related to stereotype activation and application should be more congruent and demonstrate reduced stereotyping with counterstereotypic training.

Notably, the Kawakami et al. (2000) experiments examined only stereotype activation and the Kawakami, Dovidio, and van Kamp (2005) and present research assessed only stereotype application. To understand how training interacts with correction actions to influence the full stereotyping process, however, future research should examine each of the following processes in sequence to fully investigate: (a) how training influences the activation of stereotypes, (b) how training influences motivations to correct for the influence of the training, (c) how motivations to correct influence the initial application of stereotypes, (d) how initial correction processes

influence further motivations to correct, and finally (e) how reductions in motivations to correct influence the subsequent application of stereotypes.

Future research might also productively examine individual differences in correction for the perceived influence of others (Wegner & Petty, 1995, 1997; Wilson & Brekke, 1994) and characteristics of the person doing the influencing. For example, with respect to individual differences, Plant and Devine (2001) found that backlash to pro-Black pressure was experienced primarily by people who had high external and low internal motivation to respond without prejudice (measured on separate dimensions). From the perspective of the present research, people who are oriented toward the influence of others may be particularly sensitive to unwanted influence and thus be more likely to engage in correction processes. With respect to characteristics of the perceived influencer, it is also possible that the type of person perceived to be the source of the influence or intervention can also influence correction processes. Interventions from a highly respected legitimate source, for example an elementary teacher instructing her fourth grade class on multiculturalism, may invoke fewer correction processes than would those from someone who is not respected or legitimate. It is possible that undergraduate students partaking in a study by a psychologist on stereotyping and prejudice in the laboratory may be especially sensitive to being influenced.

In summary, while the present findings suggest that correction processes can indeed impact the effectiveness of strategies aimed at reducing bias, it is clear that further research is warranted. These results, however, even in this preliminary stage, have important practical implications related to stereotyping and discrimination and raise significant questions related to anti-bias interventions. For example, when a lawyer uses the 'race card' in a criminal trial (e.g. in the O.J. Simpson trial), will the jury attempt to correct for the perceived influence of this strategy and become more racist or will they become less racist and more vigilant to prejudice as the lawyer presumably intended? While the present study

suggests that participants may actually react to this intervention and respond with more bias at least initially, future research may provide more precise information on who corrects, when they correct, and why they correct.

In conclusion, the findings from the present study and their implications demonstrate the importance of understanding both implicit and explicit forms of intergroup bias (Bargh, 1996, 1997; Kawakami & Dovidio, 2001; Kawakami, Young, & Dovidio, 2002) and how correction processes can moderate their disassociation or concordance. Although a major focus in recent research has been on the reduction of stereotype activation (see Blair, 2002, for a review), and a variety of methods focusing on the characteristics of the individual category members, stimulus cues, focus of attention, self-enhancement, and social motives have shown that this process can be modified (Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001; Galinsky & Moskowitz, 2000; Kunda, 1999), it is unclear what the effects of these strategies will be on other manifestations of discrimination and on the use of stereotypes when encountering an actual group member. The present findings underline the importance of examining the potential of these strategies to influence not only the activation but also the deliberate use of stereotypes in future research. Because people's personal perceptions about strategies aimed at reducing bias influence how they will respond, it is critical to include investigations of lay theories such as the Flexible Correction Model (Wegener & Petty, 1995, 1997; Wilson & Brekke, 1994). Whereas the present research is a first step in this direction, future research needs to continue to examine both the nature of the intervention and of the response that it is targeted to change, as well as the potential role of motivations and perceptions related to these processes.

## Acknowledgments

The research reported in this paper was supported by a Social Science and Humanities Council of Canada (SSHRC) grant and a Netherlands Science Foundation (NWO) grant to the first author.

## References

- Bargh, J. A. (1996). Automaticity in social psychology. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169–183). New York: Guilford.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Erlbaum.
- Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*, 1014–1027.
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–261.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841.
- Brehm, S., & Brehm, J. (1981). *Psychological reactance: A theory of freedom and control*. San Diego, CA: Academic Press.
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*, 1151–1164.
- Correll, J., Spencer, S., & Zanna, M. P. (2004). An affirmed self and an open mind: Self-affirmation and sensitivity to argument strength. *Journal of Experimental Social Psychology*, *40*, 350–356.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 828–841.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276–319). New York: Guilford.
- Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about survey questions: Meaning, memory,*

- attitudes, and social interaction (pp. 204–237). New York: Russell Sage Foundation.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62–68.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). The nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, 33, 510–540.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.
- Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, 78, 708–724.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60, 509–517.
- Hong, Y., Levy, S.R., & Chiu, C. (2001). The contribution of the lay theories approach to the study of groups. *Personality and Social Psychology Review*, 5, 98–106.
- Kawakami, K., Dion, K. L., & Dovidio, J. F. (1998). Racial prejudice and stereotype activation. *Personality and Social Psychology Bulletin*, 24, 407–416.
- Kawakami, K., & Dovidio, J. F. (2001). Implicit stereotyping: How reliable is it? *Personality and Social Psychology Bulletin*, 27, 212–225.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, 78, 871–888.
- Kawakami, K., Dovidio, J. F., & Spears, R. (2005). *A change of habit: The consequences of inducing bias for motivation and stereotyping*. Manuscript submitted for publication.
- Kawakami, K., Dovidio, J. F., & van Kamp, S. (2005). Kicking the habit: Effects of nonstereotypic association training and correction processes on hiring Decisions. *Journal of Experimental Social Psychology*, 41, 68–75.
- Kawakami, K., Spears, R., & Dovidio, J. F. (2002). Disinhibition of stereotyping: Context, prejudice, and target characteristics. *European Journal of Social Psychology*, 32, 517–530.
- Kawakami, K., Young, H., & Dovidio, J. F. (2002). Automatic stereotyping: Category, trait, and behavioral activations. *Personality and Social Psychology Bulletin*, 28, 3–15.
- Kruglanski, A.W. (1990). Lay epistemic theory in social cognitive psychology. *Psychological Inquiry*, 1, 181–197.
- Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.
- Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, 72, 275–287.
- Lewin, K. (1935). *A dynamic theory of personality: Selected papers* (D. E. Adams & K. E. Zener, Trans.). New York: McGraw Hill.
- Lewin, K. (1951). *Field theory in social science*. New York: Harper.
- Lieberman, N., & Forster, J. (2000). Expression after suppression: A motivational explanation of postsuppressional rebound. *Journal of Personality and Social Psychology*, 79, 190–203.
- Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology*, 59, 27–37.
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81, 33–43.
- Monteith, M.J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, 65, 469–485.
- Mussweiler, T., & Neumann, R. (2000). Sources of mental contamination: Comparing the effects of self-generated versus externally provided primes. *Journal of Experimental Social Psychology*, 36, 194–206.
- Nelson, T. D. (2001). *The psychology of prejudice*. Boston: Allyn and Bacon.
- Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context induced contrast. *Journal of Experimental Social Psychology*, 29, 137–165.
- Petty, R. E., Wegener, D. T., & White, P. (1998). Flexible correction processes in social



judgment: Implications for persuasion. *Social Cognition*, 16, 93–113.

Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology*, 37, 486–501.

Ratcliff, R. (1993). Methods for dealing with response time outliers. *Psychological Bulletin*, 114, 510–532.

Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 217–245). Hillsdale, NJ: Erlbaum.

Shah, J. Y., Kruglanski, A. W., & Friedman, R. (2002). Goal systems theory: Integrating the cognitive and motivational aspects of self-regulation. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.), *Motivated social perception: The Ontario Symposium* (Vol. 9, pp. 247–276). Mahwah, NJ: Erlbaum.

Stangor, C., & Lange, J. E. (1994). Mental representations of social groups: Advances in understanding stereotypes and stereotyping. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 357–416). New York: Academic Press.

Steele, C. M. (1999). The psychology of self-affirmation: Sustaining the integrity of the self. In R. F. Baumeister (Ed.), *The self in social psychology* (pp. 372–390). New York: Psychology Press.

Steele, C. M., & Lui, T. J., (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology*, 45, 5–19.

Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naïve theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68, 36–51.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). New York: Academic Press.

Wicklund, R. A., & Gollwitzer, P. M. (1982). *Symbolic self completion*. Hillsdale, NJ: Erlbaum.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117–142.

Wilson, T. D., Houston, C. E., & Meyers, J. M. (1998). Choose your poison: Lay beliefs and unintended attitude change. *Social Cognition*, 16, 114–132.

Zacks, R. T., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241–264). San Diego, CA: Academic Press.

*Paper received 27 October 2004; revised version accepted 9 July 2005.*

## Appendix

### Stereotypic stimulus words

Female traits in training		Male traits in training	
modest	gentle	strong	brave
kind	tender	adventurous	courageous
tidy	caring	athletic	powerful
attentive	understanding	practical	active
sensitive	impulsive	realistic	handy
naïve	jealous	macho	closed
talkative	vain	aggressive	stubborn
anxious	weak	coarse	abrupt
vulnerable	fickle	untidy	loud
careful	complaining	headstrong	messy
Female traits in trait judgment task		Male traits in trait judgment task	
helpful	social	authoritative	determined
precise	open	risk-taker	pioneering
insecure	submissive	dominant	hard
irrational	dependent	competitive	intimidating

## **Biographical notes**

KERRY KAWAKAMI is an associate professor at York University. Her research interests, in general, are related to social cognition. Specifically, she studies automatic and controlled processes involving stereotyping and prejudice and strategies aimed at reducing these biases.

JOHN F. DOVIDIO is currently professor of psychology at the University of Connecticut. Before coming to the university, he was

Charles A. Dana Professor of Psychology at Colgate University, where he also served as Provost and Dean of the Faculty. His research interests are in stereotyping, prejudice, and discrimination; social power and nonverbal communication; and altruism and helping.

SIMONE VAN KAMP is an undergraduate honor's thesis student at the University of Nijmegen. Her research interests include gender biases, stereotyping, prejudice, and discrimination.