Supplementary Materials

Chapter 2

Study 2.1. In this section readers can find the correlations between different measures in study 2.1 (see Table A), and a description of additional measures not described in the Chapter itself.

Method.

Inclusion of the self in other scale. This scale requires participants to select one from a series of 7 pictures (Schubert & Otten, 2002). Each picture shows a circle labelled "self" and a larger circle labelled with the group name. Consecutive pictures show decreasing distance between the self and the group. Thus, the first picture shows considerable separation of the self and the group, in the final picture the self is completely within the group circle. Two versions of this scale were used, one asking about the distance between self and women as a group and the other asking about the distance between women and men.

Gender Role Preferences. The 8 items of the gender role preference scale (Becker & Wagner, 2009) examine participants' gender role preferences (α=0.81). This scale was included to compare results for our measure of feminist identification to results for this scale (Becker and Wagner, 2009).

Results.

Inclusion of the self in women as a group. The extent to which participants perceived overlap between themselves and women as a group, was predicted by both women's identification and feminist identification (women's ID: β=0.43, $F(1,85)=20.01$, $p< 0.001$; feminist ID: β=0.21, $F(1,85)=4.76$, $p=0.032$). The item asking about the overlap between women as a group and men as a group was not predicted by either of the identification variables (Fs<2.99, ps>0.87).

Gender role preference. Scores on the Gender Role Preference scale were not predicted by women's identification or feminist identification (Fs<1).

Table A. *Correlation table Study 2.1*

		Women's ID	Feminist ID	Femininity	Self-stereotyping	Solidarity	Satisfaction	Homogeneity	Attitudes to the Feminist movement	Modern Sexism	Disadvantage	Hostile Sexism
Women's ID	Correlation	1.000										
	Significance											
Feminist ID	Correlation	**0.18**	1.000									
	Significance	0.101										
Femininity	Correlation	**0.62**	**0.01**	1.000								
	Significance	0.000	0.961									
Self-stereotyping	Correlation	**0.48**	**0.15**	**0.52**	1.000							
	Significance	0.000	0.167	0.000								
Solidarity	Correlation	**0.60**	**0.39**	**0.40**	**0.52**	1.000						
	Significance	0.000	0.000	0.000	0.000							
Satisfaction	Correlation	**0.55**	**0.12**	**0.50**	**0.46**	**0.40**	1.000					
	Significance	0.000	0.256	0.000	0.000	0.000						
Homogeneity	Correlation	**0.11**	**0.01**	**0.35**	**0.48**	**0.23**	**0.23**	1.000				
	Significance	0.303	0.912	0.001	0.000	0.034	0.032					
Attitudes to the Feminist movement	Correlation	**0.21**	**0.50**	**0.10**	**0.07**	**0.31**	**0.14**	**0.01**	1.000			
	Significance	0.055	0.000	0.353	0.541	0.004	0.197	0.914				
Modern Sexism	Correlation	**0.04**	**0.31**	**-0.21**	**-0.39**	**-0.06**	**-0.12**	**-0.31**	**0.47**	1.000		
	Significance	0.719	0.003	0.059	0.000	0.578	0.262	0.014	0.000			
Disadvantage	Correlation	**-0.05**	**0.42**	**-0.13**	**-0.29**	**0.04**	**-0.21**	**-0.11**	**0.32**	**0.56**	1.000	
	Significance	0.654	0.000	0.218	0.006	0.740	0.056	0.299	0.003	0.00		
Hostile Sexism	Correlation	**-0.17**	**-0.26**	**0.02**	**-0.01**	**-0.07**	**-0.02**	**0.18**	**-0.47**	**-0.44**	**-0.28**	1.000
	Significance	0.123	0.017	0.848	0.945	0.533	0.858	0.097	0.000	0.000	0.010	
Benevolent Sexism	Correlation	**0.06**	**0.07**	**0.19**	**0.24**	0.20	0.16	0.25	-0.14	-0.38	-0.14	0.57
	Significance	0.580	0.523	0.086	0.067	0.067	0.151	0.020	0.155	0.000	0.211	0.000

Study 2.2. In this section readers can find the correlations between different measures in study 2.2 (see Table B), and a description of some additional measures not described in the Chapter itself.

Table B. *Correlation table Study 2.2*

Control Variable			**Women's ID**	**Feminist ID**	**Moderate action**
Efficacy	**Women's ID**	Correlation	1.000		
		Significance			
	Feminist ID	Correlation	**0.12**	1.000	
		Significance	0.193		
	Moderate action	Correlation	**0.13**	**0.26**	1.000
		Significance	0.143	0.004	
	Radical action	Correlation	**-0.10**	**0.34**	**0.22**
		Significance	0.261	0.000	0.013

Method.

As data for this study were collected as part of a larger experiment, there were several dependent variables that are not of central interest to the current study. These measures are described here, in the order in which they were administered.

Inclusion of the self in others scale. Two versions of this scale were used, as in Study 2.1, one asking about the distance between self and women as a group and the other asking about the distance between women and men (Schubert & Otten, 2002).

Perceived efficacy. Participants indicated to what extent they perceived their in-group as efficacious, using three items such as "I think women united can successfully defend their rights", rated on a 7-point scale (α=0.83). Perceived efficacy was measured as it has been shown to be an important predictor of collective actions (Van Zomeren, Spears, Fischer, & Leach, 2004).

Support for Feminist Goals. Attitudes to the feminist movement were measured with the global goals of feminism scale (Morgan, 1996), consisting of 10 items such as "Women should be considered as seriously as men as candidates for the Presidency of Spain" (α=0.56), rated on a scale of 1 to 7.

Affirmative action attitudes. Attitudes towards affirmative action aimed at improving the position of women were measured by 3 items (from Tougas et al, 1999; e.g., "If there are no affirmative action programs helping women in employment, they will continue to be unfairly treated"; α=0.63), rated on a scale of 1 to 7.

Study 2.3. In this section readers can find the correlations between different measures in study 2.3 (see Table C), and a description of some additional measures not described in the Chapter itself.

Table C: *Correlation table Study 2.3*

		Women's ID	Feminist ID	Femininity	Modern Sexism	Disadvantage	Hostile Sexism	Benevolent Sexism	Attitude Strength	Problematic stereotypes (pos descr)	Problematic Stereotypes (neg descr)	Problematic stereotypes (pos prescr)
Women's ID	Corr	1.000										
	Sign											
Feminist ID	Corr	**0.25**	1.000									
	Sign	0.001										
Femininity	Corr	**0.51**	**0.11**	1.000								
	Sign	0.000	0.142									
Modern Sexism	Corr	**0.13**	**0.66**	**-0.09**	1.000							
	Sign	0.085	0.000	0.251								
Disadvantage	Corr	**0.20**	**0.41**	**0.05**	**0.55**	1.000						
	Sign	0.006	0.000	0.508	0.000							
Hostile Sexism	Corr	**-0.004**	**-0.49**	**0.14**	**-0.59**	**-0.33**	1.000					
	Sign	0.952	0.000	0.056	0.000	0.000						
Benevolent Sexism	Corr	**0.36**	**-0.12**	**0.31**	**-0.21**	**-0.10**	**0.46**	1.000				
	Sign	0.000	0.098	0.000	0.004	0.182	0.000					
Attitude Strength	Corr	**0.15**	**0.77**	**0.02**	**0.64**	**0.36**	**-0.51**	**-0.15**	1.000			
	Sign	0.031	0.000	0.836	0.000	0.000	0.000	0.042				
Problematic stereotypes (pos descr)	Corr	**-0.13**	**0.24**	**-0.16**	**0.13**	**0.16**	**-0.22**	**-0.29**	**0.27**	1.000		
	Sign	0.074	0.001	0.036	0.085	0.031	0.008	0.000	0.000			
Problematic stereotypes (neg descr)	Corr	**-0.11**	**0.35**	**-0.17**	**0.34**	**0.20**	**-0.41**	**-0.41**	**0.39**	**0.59**	1.000	
	Sign	0.150	0.000	0.023	0.000	0.007	0.000	0.000	0.000	0.000		
Problematic Stereotypes (pos prescr)	Corr	**-0.16**	**0.27**	**-0.18**	**0.30**	**0.33**	**-0.35**	**-0.50**	**0.33**	**0.38**	**0.50**	1.000
	Sign	0.036	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Problematic Stereotypes (neg prescr)	Corr	**-0.06**	**0.21**	**-0.05**	**0.29**	**0.29**	**-0.27**	**-0.38**	**0.27**	**0.23**	**0.45**	**0.72**
	Sign	0.455	0.004	0.521	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000

Method.

Hiring task. In the hiring task, the instructions asked participants to imagine that a new female leader had been hired at a company they worked for, and that this woman would become their new department boss. Participants were then asked to evaluate a series of traits in terms of how important they would be for their new female leader to have. Items focused on warmth, competence and morality. This measure was designed to examine perceptions of stereotypes indirectly; more endorsement of stereotypes would lead to a preference for more stereotypical attributes. However, preliminary analysis of this measure showed that morality traits were valued much more highly than either warmth or competence traits. In fact, morality was valued so highly, that differences between the warmth and competence dimensions were obscured by it. For this reason, the interpretability of the results of this measure was limited and we do not report it in the main text.

Regulatory focus. Research has shown that individuals under promotion focus commit to collective action when it is likely that the objectives will be achieved. Individuals under prevention focus, however, saw collective action for social change as a moral obligation, and were less affected by the likelihood of success (Zaal, Van Laar, Ståhl, Ellemers, & Derks, 2012). Extending this reasoning to the gender context, high/high identifiers may adopt a promotion focus when considering gender issues: they understand the disadvantaged social position of women, but are quite satisfied on a personal level, and therefore would only "risk" engaging in collective action when it is likely that the objectives are reached. Conversely, distinctive feminists may adopt a prevention focus: they see striving for social change on gender issues as a moral obligation. This option was explored as an alternative to our central hypothesis. We measured regulatory focus, both dispositional, and in the context of gender issues. Dispositional regulatory focus was assessed with the Regulatory Focus Proverb Questionnaire (α=0.70, Van Stekelenburg, 2006). Participants rate the extent to which proverbs with promotion and prevention foci reflect the approach they take to life. Examples include "better be safe than sorry" (prevention), and "nothing ventured nothing gained" (promotion).

Alongside this dispositional measure of regulatory focus, we developed some items that reflect situational regulatory focus of statements, specific to the gender context. Six items (α=0.81) examined endorsement of feminist principles when they were framed as having either prevention goals (preventing sexism) or promotion goals (promoting gender equality). Another

five items (α=0.97) represented the objectives of (moderate) collective action as either preventive or promotion-focused (i.e. "drawing attention to the disadvantage faced by women" vs. "showing support for women's rights").

Attitude strength. We wished to exclude the alternative possibility that the interaction between identification with women and feminists affects attitude strength rather than content. Such an explanation would be counter to multiple identities approach, which distinguishes the subgroups based on content, rather than strength of identification. The measure of attitude strength was composed of 8 items (α=0.876, items adapted from Vonofakou, Hewstone, & Voci, 2007), such as "How often do you think about gender and its meaning?".

Results.

Hiring task. In the hiring task, participants were asked to rate the importance of several traits representing warmth, competence and morality, in terms of how important they would be for a new female leader in their company to have. There was a significant 3-way interaction between feminist identification, women's identification and dimension, $F(1, 185)=4.314$, $p=0.039$, such that the differences between these conditions are amplified as scores on the identification variables go up. Non-identifiers do not show significant differences between any of the dimensions ($F(1, 185)=1.71$, $p=0.193$), while high/high identifiers rate all dimensions differently ($F(1, 185)=25.31$, $p<0.001$, $\eta^2_{p=}$ 0.30).

These findings suggest that as identification goes up, opinions on the traits of female leaders become more pronounced, such that morality is considered most important, followed by warmth, followed by competence. The third measure of perceptions of stereotypes was the hiring task. We hypothesised that more endorsement of stereotypes would lead participants would give more importance to stereotypical traits in a woman. It was found that participants rated morality as the most important aspect of a female leader, followed by warmth, followed by competence. These effects were more pronounced as identification (with feminists and women) went up. The finding that warmth is considered more important than competence may reflect the fact that the measure stated that the female leader would become the participants' boss at work. As warmth traits serve communal functions focused on other-interest, while competence traits could be seen as more focused on self-interest (Abele &Wojciszke, 2007), participants' preference for warmth over competence may reflect a concern for one's own outcomes, rather than a desire for the leader to embody female stereotypes (warmth). That is, the

preference for warmth may not be related to the stereotypicality of this dimension. However, this explanation could not be tested directly in this study.

Regulatory focus. The first component of the regulatory focus measure, assessed dispositional regulatory focus through the RFPQ (Van Stekelenburg, 2006), to examine general preference for prevention and promotion focus. Results showed that, in general, participants reported more agreement with promotion focused items ($F(1, 185)=6.36$, $p=0.013$, $\eta^2_{p=}$ 0.03). Additionally, there was a main effect of women's identification ($F(1, 185)=4.46$, $p=0.036$, $\eta^2_{p=}$ 0.02), such that higher women's identifiers endorsed the statements more, regardless of their focus.

Endorsement of feminism was affected by a main effect of feminist identification, $F(1, 185)=57.31$, $p<0.001$, $\eta 2p= 0.24$, such that higher feminist identification lead to more endorsement of the aims of feminism, regardless of their prevention or promotion focus ($Fs<1.41$, $ps>0.24$). Additionally, there was a main effect of focus, such that participants in general perceived feminism as having a preventive focus (i.e. preventing women's disadvantage, rather than supporting women's rights), $F(1, 185)=18.21$, $p<0.001$, $\eta^2_{p=}$ 0.09.

Endorsement of gender-related collective action also showed a main effect of feminist identification, $F(1, 185)=79.97$, $p<0.001$, $\eta^2_{p=}$ 0.31, such that higher feminist identification increased endorsement of collective action. Again, there was a main effect of focus, $F(1, 189)=5.13$, $p=0.025$, $\eta^2_{p=}$ 0.03, such that participants endorsed more collective action when it focused on preventing women's disadvantage. Arguably, this effect of focus could be explained by the perceived seriousness of the issue at stake: the negatively framed collective action items may garner more support because preventing disadvantage is perceived as more important than providing support.

In sum, the hypothesis that the interaction between women's identification and feminist identification affects the regulatory foci adopted when thinking about gender issues was not supported.

Attitude Strength. Attitude strength was associated with feminist identification, $\beta=0.50$, $F(1, 185)=237.52$, $p<0.001$, such that higher feminist identification predicted stronger attitudes on gender issues. Crucially, the interaction between feminist identification and women's identification did not reach significance ($F<1$). Thus, there is no evidence that the interaction between women's identification and feminist identification produces differences in the strength of attitudes.

Study 2.4. In this section readers can find the correlations between different measures in study 2.4 (see Table D), and a description of some additional measures not described in the Chapter itself.

Method.

Hiring task. In the hiring task, the instructions asked participants to imagine that a new female leader had been hired at a company they worked for, and that this woman would become their new department boss. Participants were then asked to evaluate a series of traits in terms of how important they would be for their new female leader to have. Items focused on warmth, competence and morality. This measure was designed to examine perceptions of stereotypes indirectly; more endorsement of stereotypes would lead to a preference for more stereotypical attributes. However, as in Study 3, preliminary analysis of this measure showed that morality was valued so highly, that differences between the warmth and competence dimensions were obscured by the large effect it produced. For this reason, the interpretability of the results of this measure were limited and we do not report it in the main text.

Regulatory focus. The previous study measured regulatory focus generally, as well as regulatory focus when considering gender issues. The Regulatory Focus Proverbs Questionnaire used in the previous study (van Stekelenburg, 2006) was not affected by the manipulation and omitted in this study. The items focusing on feminism and collective action showed that feminists tended to agree with all statements, regardless of the regulatory focus it reflected. Therefore, in the current study these items were presented as forced choice. Participants were asked to select the option that reflected their opinions most closely, with on one end of the scale a preventively framed option (e.g. the aim of feminism is to prevent sexism) and at the other end of the scale a promotion-focused option (e.g. the aim of feminism is to promote gender equality). Using the statements as opposite scale anchors halved the number of items from 12 to 6, with 3 items focussing on the aims of feminism, and 3 items focusing on the objectives of collective action.

Attitude strength. This measure examined the alternative possibility that identification with women and feminists affects attitude strength rather than content. Such an explanation would be counter to the TGIF model, which distinguishes the subgroups based on content, rather than strength of identification. The measure of attitude strength was composed of 8 items (α=0.876,

items adapted from Vonofakou et al., 2007), such as "How often do you think about gender and its meaning?"

Results.

Effect of the manipulation. The manipulation exposed participants to the views of two different women on issues of gender equality, with one speaker being critical of gender stereotypes, and the other speaker endorsing stereotypes. Results showed that the pro-stereotype speaker was perceived more positively than the anti-stereotype speaker on all dimensions (warmth, intelligence, liking, agreement), $F(1,189)=94.36$, $p<0.001$, $\eta^2_{p=}$ 0.34. Additionally, there were main effects of feminist identification, $F(1,189)=15.79$, $p<0.001$, $\eta 2p=$ 0.08, and women's identification, $F(1,189)=5.16$, $p=0.024$, $\eta^2_{p=}$ 0.03, such that higher feminist identification and higher women's identification lead to more positive ratings being given, regardless of who the speaker is, and regardless of the dimension on which speakers are rated. Finally, an interaction between the speaker, the dimension on which the speaker was rated, and feminist identification ($F(1,189)=9.21$, $p=0.003$), showed that low feminists' ratings of agreement were less strongly affected by the content of the speakers' arguments than high feminist identifiers, $F(1,189)=4.50$, $p=0.035$, $\eta^2_{p=}$ 0.02.

In sum, participants gave more positive ratings to the pro-stereotype speaker than the anti-stereotype speaker; this preference for the pro-stereotype speaker was reflected particularly strongly in high feminists' agreement ratings. This is contrary to our hypothesis that distinctive feminists would agree more with the anti-stereotype speaker. One reason why feminists disliked the anti-stereotype speaker in this sample, may be that her arguments were framed quite prescriptively ("women should not behave according to stereotypes"). Findings from Study 3.3 have shown that women dislike prescriptions for women's behaviour, a finding which is confirmed in the current study (see below). Thus, feminists may have preferred the pro-stereotype speaker over the anti-stereotype speaker, because the arguments of the pro-stereotype speaker were more accepting of women's choices.

Hiring task. In the hiring task, participants were asked to rate the importance of several competence and warmth-related traits for a hypothetical female leader. Results showed that there was an interaction between feminist and women's identification, $F(1,189)=6.35$, $p=0.013$, $\eta^2_{p=}$ 0.03, such that non-identifiers placed less importance on a female leader having positive traits than other women did.

Positive traits also represented different dimensions, of warmth, com-

petence and morality. There was a main effect of dimension, $F(1,189)=17.27$, $p<0.001$, $\eta^2_{p=}$ 0.09, such that participants overall placed greatest importance on morality, followed by warmth, followed by competence.

Regulatory Focus. There were no significant effects of feminist or women's identification on regulatory focus (Fs<1).

Attitude Strength. As in Study 2.3, attitude strength was positively predicted by feminist identification, $F(1,189)=72.88$, $p<0.001$, such that feminist identifiers reported stronger gender attitudes than non-feminist identifiers. The interaction between feminist identification and women's identification did not reach significance (F<1).

Table D. *Correlation table Study 2.4*

		Women's ID	Feminist ID	Femininity	Modern Sexism	Disadvantage	Hostile Sexism	Benevolent Sexism	Attitude Strength	Gender differentiation	Problematic stereotypes (pos descr)	Problematic stereotypes (neg descr)	Problematic stereotypes (pos prescr)
Women's ID	Corr	1.000											
	Sign												
Feminist ID	Corr	**0.26**	1.000										
	Sign	0.000											
Femininity	Corr	**0.56**	**0.20**	1.000									
	Sign	0.000	0.007										
Modern Sexism	Corr	**0.15**	**0.50**	**0.10**	1.000								
	Sign	0.038	0.000	0.169									
Disadvantage	Corr	**0.27**	**0.43**	**0.24**	**0.44**	1.000							
	Sign	0.000	0.000	0.001	0.000								
Hostile Sexism	Corr	**-0.11**	**-0.50**	**-0.02**	**-0.39**	**-0.26**	1.000						
	Sign	0.152	0.000	0.772	0.000	0.000							
Benevolent Sexism	Corr	**-0.05**	**-0.31**	**0.08**	**-0.26**	**-0.11**	**0.82**	1.000					
	Sign	0.519	0.000	0.269	0.000	0.155	0.000						
Attitude Strength	Corr	**0.10**	**0.55**	**-0.05**	**0.57**	**0.37**	**-0.39**	**-0.28**	1.000				
	Sign	0.170	0.000	0.546	0.000	0.000	0.000	0.000					
Gender differentiation	Corr	**0.16**	**-0.07**	**0.12**	**-0.10**	**-0.15**	**0.13**	**0.13**	**-0.15**	1.000			
	Sign	0.031	0.085	0.091	0.197	0.042	0.074	0.079	0.041				
Problematic stereotypes (pos descr)	Corr	**-0.12**	**0.15**	**-0.14**	**0.19**	**0.13**	**-0.15**	**-0.12**	**0.18**	**-0.06**	1.000		
	Sign	0.116	0.037	0.052	0.008	0.077	0.048	0.120	0.014	0.432			
Problematic stereotypes (neg descr)	Corr	**-0.10**	**0.08**	**-0.04**	**0.09**	**0.04**	**-0.14**	**-0.23**	**0.03**	**-0.04**	**0.36**	1.000	
	Sign	0.172	0.289	0.590	0.224	0.559	0.060	0.002	0.692	0.604	0.000		
Problematic stereotypes (pos prescr)	Corr	**-0.02**	**0.15**	**-0.10**	**0.37**	**0.22**	**-0.29**	**-0.33**	**0.25**	**-0.18**	**0.26**	**0.32**	1.000
	Sign	0.808	0.046	0.175	0.000	0.000	0.000	0.026	0.001	0.014	0.000	0.000	
Problematic stereotypes (neg prescr)	Corr	**0.04**	**0.20**	**0.06**	**0.30**	**0.32**	**-0.20**	**-0.17**	**0.21**	**-0.09**	**0.14**	**0.30**	**0.70**
	Sign	0.604	0.007	0.448	0.000	0.007	0.007	0.021	0.005	0.201	0.051	0.000	0.000

Chapter 3

In this section, readers may find details of the studies described in Chapter 3. The section is divided into 4 sections, each providing details on a separate issue noted in the text. The first section provides an overview of the procedure of each study. The second section describes three outcome measures for which preliminary analyses showed some methodological problems, which hampered interpretation of results. Additionally, some measures that are described in the text were not affected by the manipulation, but did show some other significant effects. Those effects are described in section 3. The last section discusses findings specific to Study 3.1, which included 2 conditions that were not present in the other studies.

Additional measures. Aside from the measures described in the main text, each study included some additional measures that are described here, because preliminary analysis of these measures showed some methodological problems, which hampered interpretation of results. Study 3.1 included an AMP (Affect Misattribution Procedure) as a measure of collective self-esteem, and Study 3.3 included a story-writing task as a behavioural indicator of resistance. Additionally, Study 3.3 included a measure that was analysed together with other studies in a pooled analysis reported in Chapter 4. We briefly report the nature and results of these measures here.

Methods.

Collective Self-esteem. In Study 3.1, an AMP (Payne, Cheng, Govorun, & Stewart, 2005) assessed participants' implicit collective self-esteem. Participants rated 10 Chinese characters on how visually pleasing they found them. Each character was presented twice, once preceded by a female name prime (42 ms) and once preceded by a male name prime. Higher ratings of the characters preceded by female primes indicate higher collective self-esteem. The stimuli were pilot tested and selected based on similar ratings of pleasantness overall. That is, we aimed for all stimuli to have a comparable level of baseline pleasantness. However, when these stimuli were used in Study 3.1, preliminary analysis showed that not all characters were rated as equally pleasant. Some characters were rated as very pleasant and some as quite unpleasant; indicating that pilot testing had not been successful in identifying stimuli that had comparable baselines. This variance produced a large main effect on ratings of pleasantness, overriding possible effects of the gender of the prime on ratings of pleasantness. Therefore, results of this measure provided little scope for interpretation.

Stories task. In the stories task (adapted from Branscombe, Spears, Ellemers, & Doosje, 2002) included in Study 3, participants were asked to write two short stories about (1) a characteristic of themselves that they were proud of and (2) a characteristic of women as a group that they were proud of. This task was included to measure individual, rather than group-based, resistance responses. That is, after exposure to stereotypes, participants might wish to distance themselves from the in-group and write more words about their pride in their personal characteristics. Conversely, resistance on this task is indicated by a greater number of words written about pride in the group following implicit stereotypes. Participants were able to choose which story they wanted to write first. Results of this task were difficult to interpret due to the fact that there was a time limit of 10 minutes on this task. This was done to make sure the experiment took roughly the same amount of time for all participants (in view of possible decay over time for implicit priming effects). However, this meant that participants who spent a long time on the first story, had less time remaining for the second story. That is, some participants were cut off while writing the second story, and the number of words they had written could therefore no longer be used as an indicator of motivational processes. While the number of participants who were affected by this was small (N=11), we consider this problematic because precisely those participants who had written a lot during the first story (and presumably were motivated to complete the task), were the ones who were cut off.

Moral Choice Dilemmas task. Study 3.3 included a moral choice dilemma task to measure in-group bias in a more behavioural way. It consisted of 8 items, adapted from Thomson (1986), in which one person has to be sacrificed to save the lives of a larger group. Four of these were adapted so that the target person to be sacrificed was a man, in the other four it was a woman who had to be sacrificed. Results of this task are described in detail in Chapter 4.

Results.

Collective self-esteem. The measure of collective self-esteem included was affected by women's identification ($F(1,73)=8.50$, $p=0.005$), so that higher women's identifiers gave lower ratings of pleasantness for the Chinese characters. There was no evidence for an interaction between exposure condition, and feminist identification or women's identification ($F<1.65$, $p>0.20$). That is, there was no evidence for resistance.

Stories task. The number of words written about pride in the self was predicted by the 3-way interaction between feminist identification, wom-

en's identification and exposure condition (F(1,249)=3.71, p=0.055), so that non-identifiers wrote more words about their pride in themselves after exposure to counter-stereotypes compared to stereotypes (M_{diff}=13.12, F(1,249)=3.88, p=0.050). There were no significant effects of any of the predictors on words written about pride in the group (Fs<2.09, ps>0.14).

Additional Results. Here we report some additional findings for some of the measures described in the main text. Firstly, we describe the step-by-step breakdown of higher order interactions observed on the implicit measures. When breaking down the interactions, we describe only those lower-order effects that reached significance, effects that did not reach significance are not described here. Secondly, some measures did not show effects of the manipulation (which is reported in the main text), but other terms did reach significance. These findings are also described here.

Implicit threat experience. The implicit threat experience task (Studies 2 & 3) showed a 4-way interaction between exposure condition, feminist identification, type of target and direction of the response, which is described in the main text. Additionally, there was a marginal 5-way interaction between exposure condition, feminist identification, women's identification, direction of response and type of target (F(1,37551)=3.26, p=0.071), showing that the main effect of direction (approach vs avoidance) was less strong for non-identifiers and dual identifiers' responses to neutral targets in the stereotype condition (F(1,37551)=3.02, p=0.082). As this effect regarded neutral targets, we considered that scope for interpretation is limited.

Self-esteem. All Studies measured self-esteem. Results from the pooled analysis showed no effect (F<1) of exposure condition on self-esteem. Nevertheless, there was a main effect of feminist identification (F(1,387)=8.69, p=0.003, d=0.15), so that increased feminist identification was associated with higher self-esteem. Increased women's identification, on the other hand, was associated with reduced self-esteem (F(1,387)=91.24, p<0.001, d=0.49). In sum, there was no evidence that implicit stereotypes negatively affected self-esteem in these studies.

Explicit in-group bias.

Study 3.2. In Study 3.2, explicit in-group bias was measured through a hiring paradigm, to examine explicit associations of warmth and competence with the genders. There was no evidence that the exposure conditions affected the ratings of the candidates who appeared in the vignette, or men and women in general (Fs<2.17, p>0.13). Nevertheless, there were a number of

other effects that reached significance. Firstly, lower women's identification lead to higher ratings of the candidates (both female and male) who appeared in the vignettes ($F(1,96)=5.31$, $p=0.023$, d=0.14) an effect which was marginally stronger amongst distinctive feminists than non-identifiers, $F(1,96)=3.11$, $p=0.081$, $d=0.15$.

For ratings of men and women in general, there was again a main effect of women's identification so that lower women's identifiers gave higher ratings on all dimensions ($F(1,96)=9.98$, $p=0.002$, $d=0.31$). Moreover, an interaction between women's identification, dimension of the evaluation and gender being rated ($F(1,96)=4.47$, $p=0.037$) showed that compared to lower women's identifiers, higher women's identifiers rate women as more competent ($F(1,96)=4.53$, $p=0.036$, $d=0.22$). Similarly, a marginal interaction between feminist identification, and dimension of the evaluation affected ratings of men, such that compared to high feminist identifiers, low feminist identifiers rated men as more competent ($F(1,96)=3.74$, $p=0.056$, $d=0.18$). In sum, these results suggest that increased women's identification leads to more favourable evaluations of women, while increased feminist identification leads to less favourable ratings of men. However, as these results are independent of exposure condition, scope for further interpretation is limited.

Ambivalent Sexism. In Study 3.2, we examined whether exposure condition affected endorsed sexism, and found that exposure to implicit counter-stereotypes reduces endorsement of sexism, particularly amongst high feminist identifiers. These effects are described in the main text. Additionally, women's identification was related to more endorsement of hostile and benevolent sexism (Hostile: $F(1,101)=4.49$, $p=0.037$, d=0.24; Benevolent: $F(1,101)=3.47$, $p=0.066$, d=0.19). Feminist identification was marginally related to less endorsement of benevolent and hostile sexism (Benevolent: $F(1,101)=3.86$, $p=0.052$, d=0.14; Hostile: $F(1,101)=4.002$, $p=0.048$, d=0.29).

Study 3.1: The evaluative conditions. In Study 3.1, participants were divided over 4 conditions, giving 46 participants in the (counter-) stereotype conditions, and 46 participants in two additional conditions. In this section we describe the reasoning behind, and results of, the inclusion of these 2 additional conditions. Results for these conditions were less clear-cut than those for the stereotype conditions, and were not included in Studies 2 and 3. This meant that there was no possibility to include these conditions in the pooled analysis, and therefore they are described here.

Introduction. A stereotype can be said to have a stereotypical and an

evaluative dimension. The evaluative dimension refers to a positive or negative evaluation of the group, while the stereotypical dimension refers to the implied characteristics of the group. Often, both dimensions act together, resulting in a representation of the group which is both negative and stereotypical. However, there are cases in which the representation of the group is stereotypical, but positive. For instance, benevolent sexism expresses positive attitudes about women ("women are warm and caring") while at the same time reinforcing stereotypical views that legitimise inequalities between the sexes (Glick & Fiske, 1996). Thus, social identity threat may come from the stereotypical content of the threat, the implied negative evaluation of the group, or the combined effect of both dimensions. Amodio and Devine (2006) have shown that whether social identity threat is operationalized with a focus on stereotypical associations, or evaluative associations leads to different outcomes. Specifically, in Study 3.1, we expect that exposure to implicit stereotypes will elicit a stronger reaction than exposure to implicit identity threat on the evaluative dimension. While stereotypes reflect social structures, evaluative identity threat simply associates women as a group with negative valence. That is, stereotype exposure has an ecological validity than the evaluative dimension does not. In this way, we hope to show that stereotypes may lead to identity threat independent of their valence.

Method. The methods of Study 3.1 are described in the main section of the paper. However, the study included 2 conditions that were not present in the other studies, those conditions are described here. This study used a 2 (Threat: present vs. absent) x 2 (Dimension: stereotype vs. evaluative) between participants design to manipulate implicit social identity threat, creating 4 conditions, notably stereotype exposure, counter-stereotype exposure, which were also present in the others studies, as well as evaluative threat exposure, and counter-evaluative threat exposure. The evaluative threat and evaluative counter-threat conditions were not present in the other studies. Social identity threat was created by pairing female primes with negative pictures in 95% of trials, while the counter-threat condition associated female primes with positive pictures in 95% of trials. Stimuli consisted of 10 moderately positive and10 moderately negatively valenced pictures. The pictures were selected to be independent of gender stereotypes. Where human actors appeared in the picture (N=4), their gender was not visible (e.g., they were wearing a chemical jumpsuit) or several actors of both genders appeared (a crowd). Participants' task was to answer a question unrelated to the gen-

der-valence association that was primed, and report whether a person (N=4) or animal (N=6) appeared in the picture.

Results.

Implicit stereotypes. Examining the results of the evaluative conditions on implicit stereotypes showed an interaction between exposure condition, gender of the prime, type of target and feminist identification ($F(1,90)=3.88$, $p=0.049$) . Decomposition of the interaction showed that, after seeing positive associations with women, high feminist identifiers responded faster to warmth targets than competence targets when preceded by a female prime ($F(1,9932)=7.74$, $p=0.005$). Conversely, after seeing negative associations with women, high feminist identifiers responded marginally faster to competence targets than warmth targets when preceded by a female prime, $F(1,9932)=2.92$, $p=0.088$. There are two possible explanations of these findings. Firstly, it may be the case that while the latter effect is indicative of resistance to negative representations of women, the first effect instead is an effect of identity safety: only when women are presented positively do high feminist identifiers associate them with warmth. In spite of this, when considering these issues it must be kept in mind that power in this sample was modest, and therefore results must be interpreted with caution.

Implicit in-group bias. Examining the results of the evaluative conditions on implicit in-group bias showed an interaction between exposure condition, gender of the prime, type of target and women's identification ($F(1,9637)=4.57$, $p=0.033$) . Decomposition of the interaction showed that, after seeing positive associations with women, low women's identifiers are faster to classify positive targets when they are preceded by a female rather than a male prime ($F(1,9637)=11.27$, $p=0.001$). That is, low women's identifiers learn the associations they are exposed to. High women's identifiers on the other hand, show the opposite response: after seeing women represented positively (and men negatively) the classification of positive words is facilitated following male rather than female primes ($F(1,9637)=6.25$, $p=0.012$). One explanation for this finding may be that high women's identifiers object to the negative representation of men in this condition.

Explicit in-group bias. The evaluative threat conditions, feminist identification, and their interactions, did not affect explicit in-group bias (Fs<1.57, ps>0.21).

Mood. The evaluative threat conditions, feminist identification, and their interactions, did not affect mood (Fs<2.41, ps>0.12).

Self-esteem. The evaluative threat conditions did not affect personal self-esteem, either as a main effect ($F<1$) or in interaction with feminist and women's identification (Fs<1.54, ps>0.21). However, there was an interaction between women's and feminist identification ($F(1,90)=6.38$, $p=0.014$), showing that traditional women report lower self-esteem than other groups of women.

Collective Self-esteem. The evaluative threat conditions, feminist identification, and their interactions, did not affect collective self-esteem (Fs<2.59, p>0.11).

Discussion. A methodological aim of Study 3.1 was to distinguish between stereotypical and evaluative components of social identity threat, to examine which dimension drives possible effects. As expected, disentangling the two dimensions shows that the stereotype dimension and the evaluative dimension have different effects on responses (Amodio & Devine, 2006; Fazio et al., 1995). The effects produced by the evaluative conditions were less clear-cut than findings produced by the stereotype conditions. Additionally, the sample size in these conditions was low (N=22 per condition), and since the conditions were not included in other studies, power could not be boosted through pooled analysis.

Chapter 4

Alternative analysis. As an alternative to the repeated measures model we consider in the results section, we also considered a multilevel model, in which Sacrifice is a binary outcome per vignette (1= sacrificed, 0=not sacrificed). Vignette is included as a random factor to examine whether the vignettes produce differences in sacrificing behavior, but this term did not reach significance ($F<2.013$, $p=0.126$). Table E shows a direct comparison between the results of the repeated measures ANOVA, reported in the main text, and the results of the multilevel model. The simple effects reflect the effect of the stereotype versus counter-stereotype exposure on the evaluations of men.

Order of tasks. As the data presented here were collected as part of 2 larger studies, this section outlines the measures used in each of the two studies (see Table F for an overview). The measures of interest in this study are underlined in Table F, and described in the main text. Other measures are described below

Table E. Comparison of results of the repeated measures ANOVA to the multilevel model

	Results			
	RM ANOVA		**Multilevel model**	
	F-value	*p*-value	*F*-value	*p*-value
4-way interaction (Condition*GenderVictim*FemID*WomID)	8.10	$p = 0.005$	6.98	$p = 0.008$
Simple Effect for distinctive feminists	4.42	$p = 0.036$	3.80	$p = 0.051$
Simple Effect for dual identifiers	2.95	$p = 0.087$	2.25	$p = 0.113$
Simple Effect for non-identifiers	6.69	$p = 0.010$	8.40	$p = 0.004$
Simple Effect for traditional women	$F<1$	n.s.	$F<1$	n.s.

NB: The simple effects reflect the effect of the stereotype versus counter-stereotype exposure on the evaluations of men.

Table F. *Procedure of Studies 4.1 and 4.2*

Study 4.1	**Study 4.2**
Demographics	Demographics
Manipulation	Manipulation
Implicit measures Evaluative decision task Approach-Avoidance task	
Moral Choice Dilemma task	Moral Choice Dilemma task
Persistence	Persistence
Collective Action intentions	Explicit in-group bias
Mood	Mood
Identification with women	Identification with women
Identification with feminists	Identification with feminists
Debriefing	Debriefing

Other measures. The measures in Table F that are not underlined are described in this section.

Study 4.1.

Evaluative decision task. Study 4.1 included an evaluative decision task (Fazio, Jackson, Dunton, & Williams, 1995), which examined responses to positive and negative targets associated with the genders. The task consisted of 120 trials. Each trial presented a subliminal gender prime (i.e., male or female name) with forward and backward masks (100 ms), followed by a supraliminal target. Targets were positive (N=20) or negative (N=20) words without stereotypical connotations, such as 'corpse' or 'vacation'(adapted from Roefs et al., 2005). Participants were asked to classify targets as positive or negative, and analyses focused on the speed with which this decision was made. In this task, the facilitation of female-positive pairs, relative to male-positive pairs, following identity threat would be indicative of implicit in-group bias (de Lemus, Spears, Lupiañez, Moya, & Bukowski, 2016).

Approach-Avoidance task. In Study 4.1, approach and avoidance tendencies were assessed using an approach-avoidance task (De Houwer, Crombez, Baeyens, & Hermans, 2001). Participants direct a little stick-person to approach or avoid neutral and threatening word stimuli, and analyses focused on the speed with which this was done. The task consisted of 120 trials: 30 threat-approach trials, 30 threat-avoidance trials, 30 neutral-approach trials and 30 neutral-avoidance trials. The target words were selected to be unrelated to stereotypes. If avoidance of threatening stimuli is facilitated compared to approach of threatening stimuli, this indicates an implicit threat experience.

Collective action intentions. Study 4.1 included a measure of collective action intentions by asking participants to sign a petition to urge the relevant ministry to create policies to improve the position of women. As no explicit reference to gender had been made in the study up to this point, it was necessary, in introducing the petition, to make gender issues salient to participants.

Study 4.2.

Explicit in-group bias. Study 4.2 examined explicit in-group bias in the moral domain through a "bail-task". Participants have to determine the bail amount for 12 individuals who have been arrested for a variety of crimes. Gender of the individual in question was counterbalanced, and gender-stereotypicality of the crime was controlled for (e.g. violent crime). Lower bail amounts for women as opposed to men would signify in-group bias.

Measures present in both studies.

Persistence tasks. Both studies included measures of persistence. We distinguish competencies that are considered stereotypically feminine, such as verbal skills, and those that are stereotypically masculine, such as mathematics and spatial abilities (de Lemus et al., 2016; Steele, Spencer, & Aronson, 2002). Thus, a math task was used to reflect performance and persistence in a male-typical domain, while an anagram task taps a more stereotypical performance domain for women. The two tasks consisted of 10 questions each. If the participant did not know the answer to the question, they could skip the item. The difficulty of the items increased throughout the task, and the final item (unbeknownst to participants) was unsolvable. These measures yielded 1) a performance measure: number of items answered correctly, and 2) a persistence measure: time spent on the unsolvable item.

Mood. Both studies included a mood scale, which was created from a combination of the dejection/agitation scale (Higgins, 2001), and the PANAS (Watson, Clark, & Tellegen, 1988), resulting in a 28-item scale asking about mood (α=0.84). Participants indicated their response on 7-point Likert scale.

Chapter 5

Study 5.1. In this section readers can find a description of the pre-test of the stimuli used in the manipulation and implicit measures, and details of main effects of in-group identification on the explicit measures.

Method.

Pre-test of stimuli. A list of 40 stereotyped traits was created based on previous studies of stereotypes in Spain (Morales & Rodriguez-Bailón, 2004; Willis & Rodriguez-Bailón, 2008). A group of 67 students at the University of Granada indicated to what extent they thought the stereotypical targets to be a) positive versus negative, b) reflecting high or low competence and c) typical of the in-group versus the out-group. A list of 40 positive and negative targets related to the economic crisis was derived from media content. The pilot participants indicated to what extent they thought these targets a) reflected advantage or disadvantage, b) were associated with the economic crisis, and c) were associated with the in-group versus the out-group.

Aside from the selection of stimuli, we used the pre-test data to examine which targets participants perceived as typical of the in-group versus the out-group. Participants considered low competence (t=6.99, p<0.001) and crisis targets (t=29.14, p<0.001) to be more typical of the in-group than to the out-

group. This provides some preliminary evidence that participants are familiar with the supposed inter-group differences which form the basis of our manipulation.

Results.

Math task. Our hypotheses regarding the math task were not supported: there was no evidence for resistance on the math task. However, there were main effects of identification on both persistence (F(1,145)=3.69, p=0.057) and performance (F(1,145)=4.53, p=0.035), such that low identifiers perform better and persist longer than high identifiers.

Hiring task. The hiring task was designed to measure in-group bias indirectly, and asked participants to rate in-group and out-group job candidates in terms of competence, warmth and suitability for a job vacancy. Although the threat condition did not affect results (Fs<1.95, ps>0.21) There was a main effect of identification on ratings of warmth for the in-group candidate (F(1,145)=4.60, p=0.034): the in-group candidate was rated as more warm as identification with the in-group increases.

Mood. The main effect of threat condition did not reach significance for either positive mood or negative mood, and neither did the interaction between threat condition and identification (Fs<1). There was a main effect of identification on positive mood (F(1,145)=8.78, p=0.004), such that higher identifiers experienced more positive mood.

Collective Action. The hypothesised effect of threat condition did not reach significance on either moderate or radical collective action, and neither did the interaction with identification (Fs<1.22, ps>0.306). Nevertheless, there were main effects of identification on moderate (F(1,145)=3.81, p=0.053) and radical collective action (F(1,145)=7.14, p=0.008), such that high identifiers indicated more support for collective action that combats in-group disadvantage.

Study 5.2. In this section readers can find details of influence of in-group identification on the explicit measures.

Results.

Math task. When taking identification into consideration, the performance measure showed a marginal interaction between threat condition and identification (F(3,138)=2.55, p=0.059), such that low identifiers performed worse after exposure to the disadvantage condition than after exposure to the stereotype condition (M_{diff}= -0.98, F(1,138)=4.85, p=0.030) and also performed worse than high identifiers in the disadvantage condition (M_{diff}=0.997,

F(1,138)=7.74, *p*=0.007). This result suggests that low identifiers experience detrimental effects on their performance when they are exposed to the disadvantage condition. However, there was no evidence for resistance against implicit stereotypes.

Mood. As in Study 5.1, positive mood was affected by identification, such that high identifiers reported more positive mood (*F*(1,138)=6.32, *p*=0.013). Moreover, positive mood was affected by a marginal interaction between threat condition and identification (*F*(3,138)=2.62, *p*=0.053), such that overall positive mood was higher amongst high identifiers than low identifiers, but this difference was smaller in the disadvantage condition. The negative mood factor was not affected by identification, threat condition, or the interaction between them (*F*s<1.07, p*s*>0.361).

Explicit responsibility and legitimacy. There was a main effect of identification on both perceived legitimacy of the situation and perceived responsibility of the in-group. High identifiers found the crisis situation and its consequences for Spain less legitimate (*F*(1,138)=4.81, *p*=0.032), and the in-group less responsible (*F*(1,138)=10.84, *p*=0.002).

Collective Action. Moderate and radical collective action were not affected by either threat condition, identification, or the interaction between them (*F*s<1.79, *p*s >0.150).