



# Penalized for Challenging Traditional Gender Roles: Why Heterosexual Relationships in Which Women Wear the Pants May Be More Precarious

Melissa Vink<sup>1</sup> · Belle Derks<sup>1</sup> · Naomi Ellemers<sup>2</sup> · Tanja van der Lippe<sup>3</sup>

Accepted: 1 November 2022 / Published online: 28 December 2022  
© The Author(s) 2022

## Abstract

There is growing evidence that heterosexual relationships in which traditional gender roles are reversed because women have attained higher societal status than their male partner are more precarious. We argue that this is the case because both partners in role-reversed relationships are evaluated more negatively than partners in more egalitarian or traditional gender role relationships. In two experimental studies conducted in the United States ( $N=223$ ) and the Netherlands ( $N=269$ ), we found that when encountering role-reversed relationships, participants perceive the woman as the more dominant and agentic one and the man as the weaker one in the relationship. They also perceive women in role-reversed relationships as less likeable, have less respect for men in role-reversed relationships, and expect that such relationships are less satisfying. In addition, in a third partner study ( $N=94$  heterosexual couples), we found that both male and female partners in role-reversed relationships considered the man to be the weaker one and the woman to be the more dominant one. Moreover, perceiving the man as the weaker one predicted lower relationship satisfaction in role-reversed couples. Overall, this research indicates that gender stereotypes about heterosexual relationships should be considered in efforts to achieve gender equity.

**Keywords** Heterosexual couples · Social perception · Socioeconomic status · Sex role attitudes · Gender relations · Traditional gender roles · Egalitarian gender roles · Relationship satisfaction

Despite changing gender dynamics in Western countries, gender stereotypes prescribing men to prioritize breadwinning and women to prioritize caregiving persist and are quite resistant to change (Haines et al., 2016; Morgenroth & Heilman, 2017; Park et al., 2010). Some studies show that couples in relationships with these traditional gender roles reversed (e.g., couples in which women are the main provider or in which men are the main caretaker) experience negative relationship outcomes (e.g., decreased marital satisfaction, increased chance of divorce, lower relationship

quality; Bertrand et al., 2015; Vink et al. 2022a; Wilcox & Nock, 2006; Zhang, 2015). Furthermore, role-reversed couples are more likely to experience negative relationship outcomes (e.g., lower relationship satisfaction) in countries that strongly endorse traditional gender role expectations (Vink et al., 2022b). These findings suggest that men and women in role-reversed relationships are sensitive to the gender stereotypes in their environment, potentially affecting their perception of their own partner as well.

The current study aims to test a potential underlying mechanism that explains the negative relationship outcomes of partners in role-reversed relationships. Focusing on relationships in which the woman has surpassed her male partner in social status, we draw on the status incongruity hypothesis (Rudman et al., 2012) to argue that these relationships may be more precarious because of the negative perceptions and expectations that people have about the status divisions that run counter to traditional gender norms in role-reversed relationships. The extent to which women and men are penalized for status violations in role-reversed relationships may explain why role-reversed relationships

---

✉ Melissa Vink  
M.Vink1@uu.nl

<sup>1</sup> Department of Social, Health, and Organizational Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

<sup>2</sup> Organizational Behavior Group, Utrecht University, Utrecht, The Netherlands

<sup>3</sup> Department of Sociology, Utrecht University, Utrecht, The Netherlands

are less socially accepted (Hettinger et al., 2014; MacInnis & Buliga, 2020) and experience more difficulties than traditional role relationships (Bertrand et al., 2015; Vink et al., 2022a; Wilcox & Nock, 2006; Zhang, 2015), and thus may operate to preserve the gender hierarchy.

### Prescriptive Gender Stereotypes and the Status Incongruity Hypothesis

Gender stereotypes follow from frequent observations of men and women in gender-typical social roles, such as men who are the breadwinner of their families and have higher status roles in society and women who are homemakers and have lower status roles (Eagly, 1987; Eagly et al., 2000; Heilman, 2001; Prentice & Carranza, 2002). Men and women who violate these prescriptive stereotypes (e.g., by succeeding in an occupation that is dominated by the other gender) receive social and economic penalties (Heilman & Okimoto, 2007; Heilman & Wallen, 2010; Rudman et al., 2012) or “backlash” (Rudman, 1998; Rudman et al., 2012). Though both men and women risk backlash for violating gender norms, some violations are judged to be more severe than others (Eagly & Karau, 2002; Prentice & Carranza, 2002). For instance, competent and strong women receive favorable responses if they also show modesty and caring qualities (Rudman & Glick, 1999). Further, the status incongruity hypothesis (Rudman et al., 2012) states that stereotype incongruent behavior results in backlash only when the behavior is perceived to be threatening to the current gender hierarchy and men’s higher status (Ridgeway, 2001; Rudman & Kilianski, 2000), as people tend to be motivated to justify and support the gender status quo (Jost et al., 2004). The status incongruity hypothesis implies that behaviors prescribed for men serve to increase or protect their status (e.g., being career-oriented, being dominant; Rudman et al., 2012), whereas behaviors proscribed for men would serve to reduce their status (e.g., being emotional or weak; Rudman et al., 2012). Similarly, behaviors prescribed for women serve to maintain their lower social status (e.g., being warm, caring, modest), whereas behaviors proscribed for women serve to increase their social status (e.g., being assertive or overly confident or placing themselves above men; Rudman et al., 2012).

Though both men and women experience backlash when they show stereotype incongruent behavior, the penalties they face when violating the gender hierarchy are different. Men who violate gender norms by succeeding in a traditionally feminine occupation or showing behaviors associated with low status are perceived to be ineffectual (being a ‘loser’) and respected less, even compared to women who show similar behaviors (i.e., the *weakness penalty*; Heilman & Wallen, 2010; Rudman et al., 2012). Women who violate gender norms by succeeding in a traditional masculine occupation or showing behaviors that might increase status are perceived to be

interpersonally hostile and disliked more (i.e., the *dominance penalty*; Heilman & Okimoto, 2007; Rudman et al., 2012).

### Status Violations and Social Backlash Effects for Role-Reversed Couples

Studies on backlash effects have mainly focused on behaviors and outcomes within the work domain and evaluations of individual women and men. However, a growing literature shows that these mechanisms spill over into the relationship domain. To illustrate, cross-status couples in which either the woman earns more, is more highly educated, or has a higher status occupation than the man were viewed negatively by U.S. adults (MacInnis & Buliga, 2020), which the authors suggest may be because these role-reversed couples challenge the gender hierarchy (MacInnis & Buliga, 2020). In couples where the man has a lower status occupation than his female partner, university students in the U.S. predict the male partner to be less satisfied with the relationship and report less sympathy with the female partner compared to traditional couples, and these negative evaluations are driven by people’s perceptions that a role-reversed man is not ‘masculine’ enough and a role-reversed woman is not ‘feminine’ enough (Hettinger et al., 2014). However, the authors do not make a distinction between positive (agency) and negative masculine traits (dominance), and between positive (communality) and negative feminine traits (weakness) in this study. We argue that especially the negative traits (i.e., dominance and weakness) may explain backlash for role-reversed couples.

In addition, one study of U.S. participants demonstrated that stay-at-home fathers are less respected than fathers who worked outside the home (Brescoll & Uhlmann, 2005). However, in this study, participants judged individual targets and a spouse was not mentioned, making it difficult to assess how these findings generalize to role violations of couples rather than individual men and women. Another study with U.S. adults where both partners were mentioned in their vignettes found that husbands without an income who did most of the domestic chores were perceived to be weaker, less agentic, and less dominant than stay-at-home husbands who work successfully from home or carry out only part of the total domestic chores (Chaney et al., 2019). However, also in this study, the focus was on individual men who violate traditional gender roles rather than on couple dynamics; as the research question was whether men would receive backlash due to a loss of earnings or due to a large share in domestic tasks. When evaluating women and men in role-reversed relationships, we argue that the focus of people’s evaluations shifts from the individual to the dyadic level. Thus, we propose that the negative evaluations of men may be driven by their lower status relative to their female partner rather than these men having lower status in general.

## Do Backlash Mechanisms Also Operate Within Relationships?

In addition to how couples in role-reversed relationships are perceived and evaluated by others, we examine whether men and women *within* role-reversed relationships also evaluate the male partner as the weak one and the female partner as the dominant one in the relationship. Moreover, we test whether relative dominance and weakness perceptions predict negative relationship outcomes for role-reversed couples compared to couples in more egalitarian or traditional relationships. On the one hand, it is not self-evident that others' gendered perceptions of the couples' relationships would be endorsed by couples in role-reversed relationships. Partners have a much more detailed and complete mental representation of one another compared to strangers (Trope & Liberman, 2010). Furthermore, people are motivated to see their relationship in a positive light (Murray, 1999). On the other hand, gender role norms have a strong influence on people, and people often try to avoid gender role violations (Amanatullah & Morris, 2010; Cherry & Deaux, 1978; Rudman et al., 2012; Wallen et al., 2017). Theories on self-categorization and self-stereotyping suggest that men and women can internalize stereotypes about the self and close others under certain conditions and adjust their behavior in line with a negative stereotype when faced with that stereotype (i.e., stereotype threat; Cadinu et al., 2013). Research on the backlash avoidance model illustrates that both men and women are aware of backlash and how it can act to constrain roles within relationships, and how fear of backlash can shape behavior (Rudman et al., 2012).

However, the question remains whether backlash affects the men and women within role-reversed relationships. Based on the processes described above, it might be difficult for couples who break with traditional gender norms to avoid the negative consequences of backlash completely. In the current research, we will examine whether men and women's perceptions of their partner's relative dominance and weakness provide one explanation for the negative relationship outcomes that role-reversed couples experience.

### Overview of Current Research

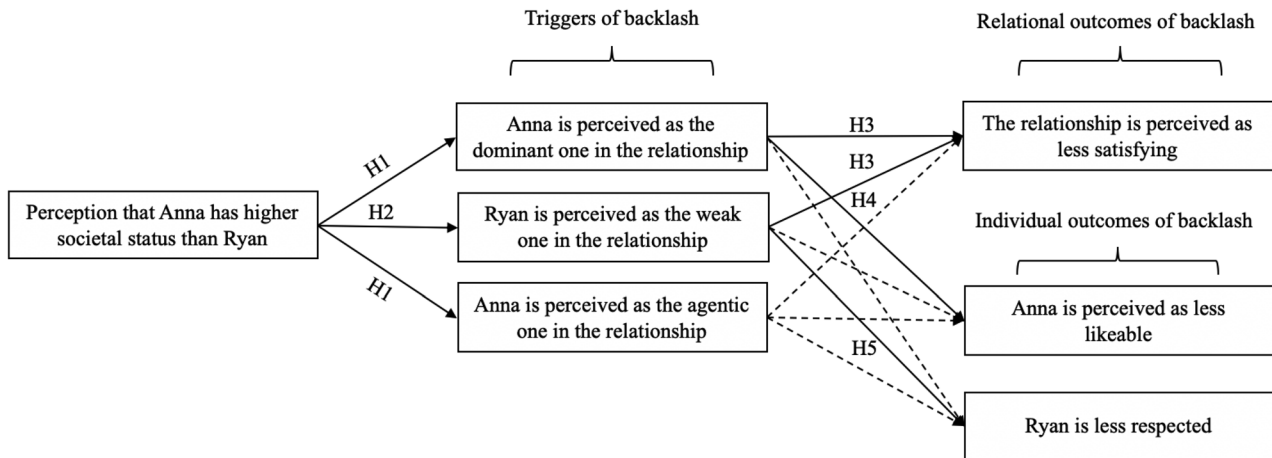
Drawing from the status incongruity hypothesis and applying it at the dyadic level, we examine whether heterosexual couples in a role-reversed relationship are at risk of facing backlash from others because status divisions within their relationship threaten the gender hierarchy (see Fig. 1).

Specifically, across three studies, we examine whether people would perceive role-reversed couples as violating gender roles through rating the women as more dominant than their male partner and rating the men as weaker than their female partner in these relationships, compared to traditional role couples; whether people would negatively evaluate the quality of role-reversed relationships and the individual men and women in role-reversed relationships (i.e., likeability of women and respect for men) as a function of these gender role violations; and whether men and women in role-reversed relationships have similar relative dominance and weakness perceptions of themselves and their partner, and whether these perceptions are related to lower relationship satisfaction.

Prior research suggests that although women who violate gender norms are perceived to be more agentic, it is their dominance rather than their agency that results in backlash (Rudman et al., 2012). Therefore, if people indeed evaluate women as the agentic partners in the relationship, we did not anticipate that this relative agency would affect their judgment of the women and their relationships. Furthermore, we expect that these backlash mechanisms operate for heterosexual couples in which the women have higher status than the men across absolute status levels. This means that we did not expect men's absolute low status would cause backlash toward heterosexual couples, but rather men having lower status relative to their women partners.

### Study 1

The aim of Study 1 was to examine the judgments of role-reversed couples (i.e., where the woman has higher status than the man) compared to traditional-role couples (i.e., where the man has higher status than the woman) and status-equal couples. Overall, for role-reversed couples, we expected that participants would perceive the woman as the more agentic and dominant one relative to the man (*H1*) and the man to be the weaker one relative to the woman (*H2*). At the relationship level, we predict that when the woman is perceived to be the more dominant one and the man to be the weaker one in the relationship, people will perceive this relationship as less satisfying for the couple (*H3*). On the individual level and consistent with the dominance penalty, we expect that when the woman is perceived as the more dominant one in the relationship, people will evaluate her as less likeable (*H4*). Consistent with the weakness penalty, we expect that people will have less respect for the man when they perceive him as the weaker one (*H5*; see Fig. 1).



*Note.* Dotted lines represent relations of which we have no specific expectations.

**Fig. 1** Theoretical Model of Social Backlash Effects in Role-Reversed Relationships

## Study 1 Method

### Participants and Design

We aimed to recruit 250 participants. A priori calculation of a repeated-measures ANOVA with a within-between interaction, an estimated partial  $\eta^2$  of .02, a power of .80, and a correlation of .03 between repeated measures in G\*Power indicated that we would need a sample size of 228 (Faul et al., 2007). This sample size is also sufficient for our structural equation models, as a priori calculation of linear multiple regression with an  $R^2$  increase of .10 (a small to medium effect), 13 tested predictors, and a total number of 33 predictors indicated a sample size of 174 (Faul et al., 2007). We aimed for 10% higher so that we could exclude those participants that failed the attention check.

Participants ( $N = 266$ ) were individuals living in the United States who were recruited through Amazon Mechanical Turk (MTurk). We included participants who consented to participate and who passed the attention check ( $N = 223$ ; 139, 62.3% men and 84, 37.7% women). Participants ( $M_{age} = 35.19$ ,  $SD_{age} = 9.21$ ) reported being of White ( $n = 115$ , 51.6%), Asian/Pacific Islander ( $n = 94$ , 42.2%), Hispanic or Latino ( $n = 7$ , 3.1%), Black or African American ( $n = 5$ , 2.2%), Native American or American Indian ( $n = 1$ , 0.4%) or Indian ( $n = 1$ , 0.4%) ethnic origin. They were married or in a domestic partnership ( $n = 138$ , 61.9%), single ( $n = 71$ , 31.8%), divorced ( $n = 8$ , 3.6%), widowed ( $n = 1$ , 0.4%) or other (e.g., engaged;  $n = 5$ , 2.2%). Further, participants had a bachelor's degree ( $n = 92$ , 41.3%), were high school graduates (or equivalent;  $n = 41$ , 21.1%), master's degree ( $n = 44$ , 19.7%), associate's degree ( $n = 34$ , 15.2%), professional school degree ( $n = 5$ , 2.2%),

or nursery school to 8<sup>th</sup> grade ( $n = 1$ , 0.4%). Participants were employed for wages ( $n = 162$ , 72.6%), self-employed ( $n = 46$ , 20.6%), students ( $n = 4$ , 1.8%), homemaker ( $n = 3$ , 1.3%), unable to work ( $n = 3$ , 1.3%), out of work and looking for work ( $n = 2$ , 0.9%), out of work but not currently looking for work ( $n = 2$ , 0.9%), or independent contractor ( $n = 1$ , 0.4%). The data collection for this research project was approved post-hoc by the ethical committee of the first author's faculty (FETC21-0502). We applied for post-hoc approval since applying for ethical approval was less commonplace at the time of data collection in 2015.

We conducted an experiment with a 2 (Partner Gender: male [Ryan]/female [Anna], within-participants)  $\times$  2 (Absolute Status Ryan: Low/Medium, between-participants)  $\times$  3 (Relative Status Anna: Lower/Equal/Higher than Ryan, between-participants) mixed design. We manipulated Ryan's status to test whether backlash in the relationship domain is indeed predicted by the relative status of the woman compared to the man compared to the absolute status of the man.

### Procedure

After providing informed consent, participants were told that we were interested in how the careers of dual-earning couples affected their relationships and they would earn \$2.50 for completing the study and passing the attention check. We emphasized that participation was voluntary and anonymous. After filling out demographic background information, participants were randomly assigned to one of the six conditions where they were asked to read one vignette that included the manipulation. After completing questions related to our independent and dependent variables, we debriefed participants on our research goals.

### Manipulation

The manipulation consisted of a description of Ryan and Anna as follows: “Now, we will describe the situation of Ryan and Anna. Ryan and Anna have been in a relationship for five years now. They met each other through mutual friends. Ryan works as [occupation] and Anna works as [occupation].”

We chose to manipulate occupation as a proxy for status. Status can be derived from one’s educational degree, income, and prestige of the job (Adler et al., 2000), which are all signaled by one’s occupation. Using Glick and colleagues’ occupations (1995), we conducted a pilot study ( $N=31$  of which 20 were men,  $M_{age}=34.48$ ,  $SD_{age}=7.22$ ) to find occupations that differed in prestige but were comparable in the perceived gender ratio of the job holders (see Table 1 and Supplement A in the online supplement). Participants responded to two items; “Please indicate what you expect is the ratio of current male and female job holders in the following occupations:” with responses given on a slider from 0 (*only male*) to 100 (*only female*) and “Please indicate the prestige you expect is associated with the following occupations:” with responses ranging from 1 (*very low prestige*) to 7 (*very high prestige*).

For every condition, we selected two occupations for Ryan and Anna that were similar in status. We did this to make sure that effects were driven by status rather than a specific, unforeseen characteristic of one occupation. These occupations were counterbalanced for each participant (see Table 1 for all selected occupations). To illustrate, in the

condition where Ryan had low absolute status and Anna higher status than Ryan, we stated that Ryan worked either as a bookbinder or food store manager and Anna as a professor or dentist.

### Measures

Items were all measured on a 7-point scale with options ranging from 1 (*completely disagree*) to 7 (*completely agree*), unless otherwise indicated. The order of the constructs was based on the hypothesized causal chain in the theorized model. As we had no specific expectations about the perceptions of who is the communal one in the relationship, as communality is not associated with status (Rudman et al., 2012). We included measures of communality to test whether this affected the outcomes. As this was not the case in both Study 1 and 2, we did not report on communality here.

### Manipulation Check of Perceived Societal Status

Ryan and Anna’s perceived status was measured with a subjective socioeconomic status ladder with ten different rungs (Adler et al., 2000). We described that people at the top of the ladder are best off in terms of income, education, and respected jobs, whereas people at the bottom are worst off. Participants were asked to indicate the rung they thought best represented Ryan and Anna’s individual situation.

**Table 1** Means and (Standard Deviations) for Prestige and Gender Ratio Across Status and Occupation in Pilot Study 1

Variable	Status	Occupation – Ryan		Occupation – Anna	
	(Ryan Absolute/Anna Relative)				
Prestige	Ryan Low/Anna Lower	Bookbinder	Food Store Manager	Bus Driver	Dishwasher
		2.19 (1.14)	2.87 (1.09)	1.90 (0.91)	1.39 (0.84)
Gender Ratio	Ryan Low/Anna Lower	44.26 (18.94)	43.90 (14.24)	40.77 (20.74)	36.94 (26.01)
		Prestige	Ryan Low/Anna Equal	Bookbinder	Food Store Manager
2.19 (1.14)	2.87 (1.09)			2.27 (1.20)	2.61 (1.05)
Gender Ratio	Ryan Low/Anna Equal	44.26 (18.94)	43.90 (14.24)	36.26 (15.11)	51.52 (19.49)
		Prestige	Ryan Low/Anna Higher	Bookbinder	Food Store Manager
2.19 (1.14)	2.87 (1.09)			5.84 (0.78)	5.45 (1.15)
Gender Ratio	Ryan Low/Anna Higher	44.26 (18.94)	43.90 (14.24)	43.35 (12.46)	42.29 (11.38)
		Prestige	Ryan Medium/Anna Lower	Radio Announcer	Accountant
4.35 (1.38)	4.48 (1.06)			2.19 (1.14)	2.87 (1.09)
Gender Ratio	Ryan Medium/Anna Lower	35.68 (18.31)	43.90 (13.71)	44.26 (18.94)	43.90 (14.24)
		Prestige	Ryan Medium/Anna Equal	Radio Announcer	Accountant
4.35 (1.38)	4.48 (1.06)			4.58 (1.23)	3.90 (1.26)
Gender Ratio	Ryan Medium/Anna Equal	35.68 (18.31)	43.90 (13.71)	43.61 (17.49)	38.68 (9.90)
		Prestige	Ryan Medium/Anna Higher	Radio Announcer	Accountant
4.35 (1.38)	4.48 (1.06)			6.16 (0.97)	5.90 (1.08)
Gender Ratio	Ryan Medium/Anna Higher	35.68 (18.31)	43.90 (13.71)	45.23 (17.90)	38.65 (11.72)

## Agency

We adapted five items from traits in the Bem Sex Role Inventory (Bem, 1981) to assess how participants rated Ryan's and Anna's level of agency (5 items; e.g., "Ryan/Anna defends his/her own beliefs, is willing to take risks, has a strong personality, is independent, is ambitious";  $\alpha_{Ryan} = .78$ ;  $\alpha_{Anna} = .84$ ).

## Dominance and Weakness

Based on Heilman and Wallen (2010), we asked participants whether they felt that Ryan/Anna was abrasive, manipulative, selfish, and cold (dominance traits,  $\alpha_{Ryan} = .82$ ;  $\alpha_{Anna} = .79$ ), and wishy-washy, wimpy, insecure, spineless, and weak (weakness traits,  $\alpha_{Ryan} = .84$ ;  $\alpha_{Anna} = .89$ ).

## Likeability

Based on Heilman and Wallen (2010), we assessed likeability with two questions: "How much do you think you would like Ryan/Anna?" and "How would you describe Anna/Ryan," on a scale ranging from 1 (*not at all likeable*) to 7 (*very much likeable*) ( $r_{Anna} = .48$ ,  $p < .001$ ;  $r_{Ryan} = .59$ ,  $p < .001$ ).

## Respect

Based on Heilman and Wallen (2010), we included two items to measure perceived respect for Ryan and Anna: "How much do you think Ryan/Anna is someone who commands respect from others?" and "How would you describe Ryan/Anna?" on a scale ranging from 1 (*respectable*) to 7 (*unrespectable*) ( $r_{Anna} = .33$ ,  $p < .001$ ;  $r_{Ryan} = .55$ ,  $p < .001$ ).

Given the correlation of the two respect items for Anna was lower, we also conducted the analyses with only a single item for respect (i.e., "How much do you think Ryan/Anna is someone who commands respect from others?"). The results were similar whether the two-item or single-item measure was used. For this reason, and because we wanted our results to be comparable with those of Heilman and Wallen (2010), we use the two-item measure of perceived respect.

## Relationship Satisfaction

Perceived satisfaction within Anna and Ryan's relationship was measured using five items from Rusbult et al. (1998) measure of relationship satisfaction (e.g., "Ryan and Anna feel satisfied with their relationship," "Ryan and Anna's relationship is close to ideal";  $\alpha = .90$ ).

## Study 1 Results

### Preliminary Analyses

Correlations between the demographic variables and hypothesized variables showed that participants' education level, ethnicity (Asian vs. white ethnic origin), marital status (married vs. single), and employment status (wages vs. self-employed) were associated with several of the dependent variables (see Table 2), and therefore we conducted the analyses with and without these variables as covariates, which showed similar results (see Supplement B in the online supplement for results without the covariates). Moreover, we controlled for the average societal status of the couple (except in the analyses for the manipulation checks) to be sure that differences in the evaluations of Anna and Ryan and perceived relationship satisfaction were not due to the fact that as a couple, Ryan and Anna had higher average status in some conditions than in others. Q-Q plots showed that all hypothesized variables were normally distributed. There were no missing data in the final sample.

### Manipulation Check

We conducted mixed repeated measures ANCOVA's to check that the experimental manipulation evoked the expected differences in dominance, weakness, and agency traits across the conditions. The findings indicated that the manipulation worked as intended, such that participants perceived Ryan to have higher absolute status in the conditions that we assigned him a medium-status occupation compared to the conditions in which he had a low-status occupation. Also, participants perceived Anna's status relative to Ryan as intended across the conditions (see Supplement C in the online supplement for these analyses). Next, we were in line with our hypotheses (i.e., H1 and H2).

### Perceptions of Agency, Dominance, and Weakness in the Role-Reversed Relationships

We conducted mixed repeated measures ANCOVA's to test whether the relative dominance and weakness perceptions of Ryan and Anna varied as predicted.

There was no interaction effect for partner gender and Ryan's absolute status,  $F(1, 145) = 1.52$ ,  $p = .220$ , but consistent with Hypothesis 1, there was an interaction effect for partner gender and Anna's relative status on perceived relative agency (see Table 3). Only in the conditions where Anna had higher status than Ryan did participants perceive her to be more agentic than Ryan. When Ryan had higher status, participants rated him to be more agentic than Anna, and when they had equal status, participants rated their agency similarly.

**Table 2** Zero-Order Correlations Among Background Variables, Predictor Variables (Dummy-Coded) and Outcome Variables in Study 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
1. Participants' Gender:	-																							
Women/Men																								
2. Age	.12	-																						
3. Education	.05	.03	-																					
4. Ethnicity: White/Asian	-.03	-.13	.54**	-																				
5. Ethnicity: White/Other	-.02	-.12	.04	-	-																			
6. Married/Single	-.25**	-.26**	-.20**	-.06	.09	-																		
7. Married/Other	.15	.30**	-.06	-.19*	-.09	-	-																	
8. Wages/Self-employed	.01	.06	-.17*	-.05	-.14	.11	.21*	-																
9. Wages/Other	-.02	-.02	-.24**	-.12	-.13	.28**	.37**	-	-															
10. Status of Couple	.00	-.08	.14*	.20**	-.13	-.01	-.11	.11	.04	-														
11. Status Ryan: High/Low	-.04	.02	.00	-.01	.02	-.01	-.05	.08	.05	.31**	-													
12. Relative Status Anna: Higher/Other	.04	.00	-.09	-.09	-.04	-.01	-.10	.01	.02	.35**	-.02	-												
13. Relative Status Anna: Lower/Other	.06	.05	.12	.06	-.01	-.11	.16	-.01	.01	-.28**	-.06	-.45**	-											
14. Agency Ryan	-.03	-.03	.18**	.17*	-.04	-.16*	-.22**	-.08	-.18*	.19**	.08	-.09	-.02	-										
15. Agency Anna	-.05	.01	.06	.14*	-.01	-.13	-.22**	-.04	-.05	.47**	.10	.33**	-.32**	.27**	-									
16. Dominance Ryan	-.05	-.11	.17*	.28**	.01	-.03	-.01	.04	-.02	.23**	.10	-.02	.04	-.05	-.16*	-								
17. Dominance Anna	-.02	-.06	.12	.27**	-.12	.01	.00	.08	.10	.32**	.08	.13*	-.05	-.24**	.00	.68**	-							
18. Weakness Ryan	-.02	-.06	.09	.21**	-.07	.06	-.03	.03	-.05	.22**	.05	.07	-.05	-.34**	-.10	.71**	.72**	-						
19. Weakness Anna	-.03	-.12	.13	.22**	-.12	.07	.01	.06	-.01	.09	.04	-.09	.14*	-.08	-.39**	.74**	.72**	.69**	-					
20. Likeability Ryan	-.09	.01	.19**	.20**	-.07	-.18*	-.23**	-.10	-.13	.22**	-.08	.13	-.08	.43**	.47**	-.29**	-.15*	-.30**	-.27**	-				

Table 2 (continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
21. Likeability Anna	-.07	.02	.21**	.19**	-.11	-.13	-.16*	-.13	-.16*	.13	-.08	.06	-.12	.32**	.40**	-.20**	-.27**	-.18**	-.29**	.58**	-	-	-
22. Respect Ryan	.02	.03	.14*	.14*	-.03	-.21**	-.09	-.15*	-.19*	.14*	-.04	-.02	.03	.55**	.24**	-.18**	-.20**	-.38**	-.15*	.55**	.38**	-	-
23. Respect Anna	.02	.07	.11	.10	-.04	-.09	-.01	.00	-.20**	.40**	.00	.33**	-.28**	.14*	.64**	-.20**	-.04	-.10	-.35**	.42**	.50**	.39**	-
24. Relationship Satisfaction	-.03	-.02	.20**	.21**	-.01	-.28**	-.15	-.02	-.21**	.25**	.07	-.08	.00	.46**	.32**	.16*	-.24**	-.31**	-.21**	.45**	.44**	.44**	.29**

Note. \*  $p < .05$ ; \*\*  $p < .01$ .

We found no interaction effect for partner gender and Ryan’s absolute status,  $F(1,145) = .04, p = .835$ , but supporting Hypothesis 1, there was an interaction effect for partner gender and Anna’s relative status on perceived relative dominance (see Table 3). Ryan was perceived to be more dominant than Anna when Anna and Ryan had traditional or equal status division. In contrast, when Anna had a higher status than Ryan, participants rated her and Ryan similarly dominant.

There was no interaction effect for partner gender and Ryan’s absolute status,  $F(1,145) = 1.15, p = .286$ , but consistent with Hypothesis 2, there was an interaction effect for partner gender and Anna’s relative status on perceived relative weakness (see Table 3). Whereas Ryan and Anna were perceived to be equally weak when they had a traditional or equal status division, when Anna had higher status than Ryan, participants rated Ryan as significantly weaker than Anna.

### Structural Equation Modelling to Test Our Theorized Model

To test our theorized model shown in Fig. 1, we built structural equation models in Mplus (Muthén & Muthén, 2008). These models provide fit indices and allow us to constrain paths of effects we did not anticipate to zero. This way, we can show that the penalties for men and women are different and that these models show a good fit to our data. We created a dummy variable for condition (i.e., Anna had lower or equal status vs. Anna had higher status than Ryan). We created difference scores for dominance, weakness, and agency between the partners. Covariances among the three mediator variables and the five dependent variables were estimated. The theoretical model showed a poor fit with the data,  $\chi^2(df=27) = 77.67, p < .001, RMSEA = .11, CFI = .91, SRMR = .09$ . Based on the modification indices, we freed the paths of Anna’s relative agency on Anna’s likeability and respect, which significantly improved the fit of the previous model,  $\Delta\chi^2(\Delta df=2) = 34.45, p < .001, RMSEA = .03, CFI = .99, SRMR = .04$ . We tested several alternative models (e.g., a model in which mediator and dependent variables were switched) to confirm that the current model provided the best fit to the data (see Supplement D in the online supplement).

### Perceived Relationship Satisfaction in the Role-Reversed Relationship

In line with Hypothesis 3 and as shown in Fig. 2, participants in the role-reversed condition rated Anna to be relatively more dominant than Ryan compared to the other conditions,  $b = .26, SE = .08, p < .001$ , and rated Ryan weaker than Anna,  $b = .26, SE = .08, p < .001$ . Participants rated the



**Table 3** Estimated Marginal Means, Standard Errors, and Multi- and Univariate Effects for Gender Backlash Variables in Study 1

Variable	Relative Status			Multivariate Interaction Effect Target Gender x Relative Status	Univariate Between-Condition Effect For Each Target
	Anna Lower than Ryan M (SE) [95% C.I.]	Anna Equal as Ryan M (SE) [95% C.I.]	Anna Higher than Ryan M (SE) [95% C.I.]		
Agency				$F(2,145) = 12.92, p < .001, \eta^2 = .11$	
Ryan	5.11 (0.14) <sup>ab</sup> [4.83, 5.39]	5.25 (0.12) <sup>a</sup> [5.00, 5.49]	4.70 (0.14) <sup>bc</sup> [4.42, 4.99]		$F(2,145) = 4.12, p = .018, \eta^2 = .05$
Anna	4.67 (0.15) <sup>c</sup> [4.38, 4.96]	5.16 (0.13) <sup>a</sup> [4.91, 5.42]	5.38 (0.15) <sup>a</sup> [5.08, 5.67]		$F(2,145) = 5.70, p = .004, \eta^2 = .07$
Dominance				$F(2,145) = 3.68, p = .025, \eta^2 = .05$	
Ryan	3.09 (0.17) <sup>a</sup> [2.76, 3.43]	2.82 (0.15) <sup>a</sup> [2.52, 3.12]	2.80 (0.17) <sup>a</sup> [2.45, 3.14]		$F(2,145) = .92, p = .401, \eta^2 = .01$
Anna	2.77 (0.18) <sup>bc</sup> [2.42, 3.13]	2.51 (0.16) <sup>c</sup> [2.20, 2.82]	3.00 (0.18) <sup>ab</sup> [2.64, 3.36]		$F(2,145) = 2.07, p = .129, \eta^2 = .03$
Weakness				$F(2,145) = 5.19, p = .007, \eta^2 = .07$	
Ryan	2.75 (0.18) <sup>ad</sup> [2.40, 3.09]	2.64 (0.15) <sup>a</sup> [2.34, 2.94]	2.95 (0.18) <sup>a</sup> [2.60, 3.30]		$F(2,145) = .85, p = .429, \eta^2 = .01$
Anna	2.91 (0.18) <sup>d</sup> [2.54, 3.27]	2.42 (0.16) <sup>a</sup> [2.10, 2.74]	2.41 (0.19) <sup>d</sup> [2.04, 2.78]		$F(2,145) = 2.37, p = .097, \eta^2 = .03$

Note. Means with different superscripts in rows and/or columns differ significantly from each other,  $p < .05$

role-reversed relationship to be less satisfying when they rated Anna as more dominant than Ryan,  $b = -.17, SE = .07, p = .010$ , and rated Ryan as weaker than Anna,  $b = -.13, SE = .06, p = .036$ , compared to the other conditions. We found a significant indirect effect of condition on perceived relationship satisfaction via Anna's perceived dominance,  $b = -.04, SE = .02, p = .040$ , but not via Ryan's perceived weakness,  $b = -.03, SE = .02, p = .075$ . However, the overall indirect effect was significant,  $b = -.08, SE = .03, p = .006$ , indicating that participants perceived the role-reversed relationship as less satisfying due to their combined evaluation of Anna's relative dominance and Ryan's relative weakness. The direct effect of condition on relationship satisfaction was significant,  $b = -.18, SE = .07, p = .011$ .

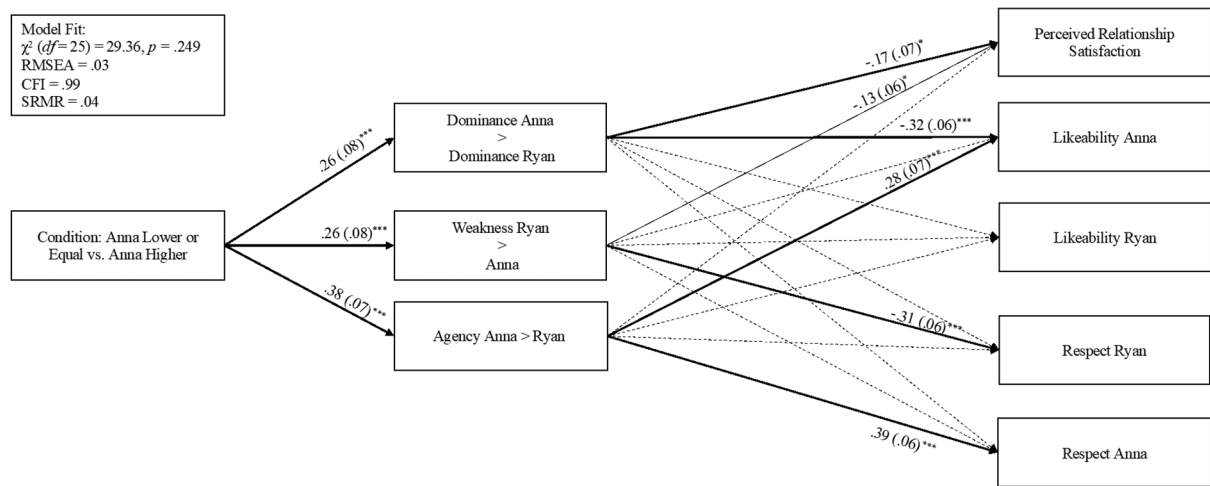
### Perceived Likeability of Anna in the Role-Reversed Relationship

Consistent with Hypothesis 4, when participants rated Anna as more dominant than Ryan in the role-reversed condition (compared to the other conditions), they liked her less,  $b = -.32, SE = .06, p < .001$ , but they did not like Ryan less (see Fig. 2). The indirect effect of condition on how much Anna was liked via Anna's relative dominance was significant,  $b = -.08, SE = .03, p = .005$ , and the direct effect of condition on Anna's likeability was not significant,  $b = .01, SE = .07, p = .972$ . Unexpectedly, when participants rated

Anna as more agentic than Ryan in the relationship, they liked her more,  $b = .28, SE = .07, p < .001$ . The indirect effect of condition on how much Anna was liked via Anna's relative agency was also significant,  $b = .11, SE = .03, p = .001$ . There was no overall indirect effect of condition on how much Anna was liked via Anna's relative dominance and agency,  $b = .02, SE = .04, p = .513$ . Thus, the negative effect via Anna's relative dominance and the positive effect via Anna's relative agency appear to cancel each other out.

### Perceived Respect of Ryan in the Role-Reversed Relationship

Consistent with Hypothesis 5, when participants rated Ryan as weaker than Anna, they reported less respect for him,  $b = -.31, SE = .06, p < .001$ , but not for Anna (see Fig. 2). There was an indirect effect of condition on how much Ryan was respected via Ryan's relative weakness,  $b = -.08, SE = .03, p = .004$ , whereas there was no direct effect of condition on Ryan's respect,  $b = -.09, SE = .07, p = .201$ . Unexpectedly, when participants rated Anna as more agentic than Ryan, they had more respect for her,  $b = .39, SE = .06, p < .001$ . There was also an indirect effect of condition on how much Anna was respected via Anna's relative agency,  $b = .15, SE = .04, p < .001$ . Also, the direct effect of perceived relative status on respect for Anna was significant,  $b = .13, SE = .06, p = .002$ .



Note: Thick lines represent significant indirect paths. Reported statistics are standardized coefficients and errors. Dotted lines are paths that were constrained to zero.

**Fig. 2** Results of Theoretical Model in Study 1 Controlling for Average Status of the Couple, Education, Ethnicity: Asian vs. White, Marital Status: Married vs. Single and Employment Status: Employed for Wage vs. Self-Employed

## Study 1 Discussion

In Study 1, we provide initial evidence that backlash effects occur in heterosexual role-reversed relationships. Regardless of their gender, people rated the partner with the higher occupational status to be more agentic than the other. However, they rated Anna as similarly dominant to Ryan in the role-reversed relationship, whereas they rated the Ryan as more dominant than Anna in the traditional and status-equal relationships. Also, people rated Ryan as weaker than Anna in the role-reversed relationship, whereas there was no difference between Anna and Ryan in perceived weakness in the traditional and status-equal relationships. Moreover, participants expected role-reversed relationships to be less satisfying compared to more traditional relationships. Women's relative dominance in role-reversed relationships was also associated with less liking of them, and men's relative weakness in these relationships was associated with less respect attributed to them. There were also positive individual outcomes for women in role-reversed relationships, such that women's relative agency perceptions were associated with more respect and increased liking of them. As we did not expect these two effects, we aimed to replicate our results in Study 2.

## Study 2

In order to replicate the findings of Study 1 in a different national context, we conducted a similar experiment ( $N=269$ ) in the Netherlands. Although both the Netherlands and the

United States are Western countries, they differ in meaningful ways. In the Netherlands, more women work part-time (58.7%) compared to women in the U.S. (17.1%; Organisation for Economic Cooperation and Development, 2018). Dutch men also work more part-time (18.9%) compared to U.S. men (8.2%; Organisation for Economic Cooperation and Development, 2018), but this difference is smaller. Furthermore, in dual-earning couples, 12% of women in the Netherlands have a higher income than their partner (Portegijs & Van den Brakel, 2018) compared to 40% of women in the United States (Pew Research Center, 2013). By conducting a second study in a Western country that differs in how common it is for couples to reverse gender roles, we could examine the generalizability of these findings to other contexts (Swanborn, 2010). We expect that backlash for role-reversed couples will be comparable in both countries because research has shown that backlash effects are quite persistent across different contexts (Rudman & Fairchild, 2004). We will therefore test the same hypotheses in Study 2 as we did in Study 1.

## Study 2 Method

### Participants and Design

Participants ( $N=269$ ; 88 men, 179 women, 2 unknown gender) were living in the Netherlands and were recruited by convenience sampling and Prolific Academic. We again aimed for around 250 participants based on a similar power analysis as Study 1. Participants ( $M_{age} = 32.96, SD_{age} = 13.33$ ) were single

( $n=94$ , 34.6%), married ( $n=71$ , 26.4%), in a non-cohabiting relationship ( $n=52$ , 19.3%), in a cohabiting relationship ( $n=42$ , 15.6%), separated ( $n=6$ , 2.2%), divorced ( $n=4$ , 1.5%), or indicated that it was complicated ( $n=1$ , 0.4%). Participants had a bachelor's degree (or equivalent;  $n=84$ , 31.2%), a higher educational degree (or equivalent;  $n=78$ , 29%), a master's degree (or equivalent;  $n=64$ , 23.8%), a vocational degree (or equivalent;  $n=25$ , 9.3%), lower educational degree (or equivalent;  $n=10$ , 3.7%), doctoral degree ( $n=6$ , 2.2%), or they finished elementary education ( $n=2$ , 0.7%). Participants were employed for wages ( $n=118$ , 43.9%), students ( $n=93$ , 34.6%), entrepreneurs ( $n=20$ , 7.4%), unemployed and looking for a job ( $n=10$ , 3.7%), unable to work ( $n=8$ , 3.0%), unemployed and not looking for a job ( $n=5$ , 1.9%), retired ( $n=3$ , 1.1%), homemaker ( $n=2$ , 0.7%), or other (e.g., students and employed for wages;  $n=10$ , 3.7%). The data collection for this research project was approved post-hoc by the ethical committee of the first author's faculty (FETC21-0502). We applied for ethics approval after data collection as applying for ethical approval was less commonplace at the time of data collection for these types of studies. The design was identical to Study 1.

## Procedure

The procedure of Study 2 was identical to Study 1 except for the form of participant compensation. Depending on where participants were recruited, they were either entered into a

lottery in which we raffled a 50-Euro gift voucher for every 50 participants, received 1.50 GBP, or 0.25 credit toward partial course requirement.

## Manipulation

We used similar vignettes as in Study 1, except that we changed Ryan and Anna's occupations based on a pilot test that we ran in the Netherlands (see Table 4 and Supplement A in the online supplement). To illustrate, in the condition where Ryan had low absolute status and Anna higher status than Ryan, we stated that Ryan worked either as a bookbinder or tailor and Anna as a radiologist or architect.

## Measures

We used identical measures as in Study 1 except for the measure of respect. Instead of two items in Study 1, we administered one item in Study 2 to assess respect ("How much do you think Ryan/Anna is someone who commands respect from others?") as the other item did not translate well into Dutch. Reliability analyses showed similar and satisfactory alphas for all other included measures: agency ( $\alpha_{Ryan} = .82$ ;  $\alpha_{Anna} = .87$ ), dominance ( $\alpha_{Ryan} = .80$ ;  $\alpha_{Anna} = .79$ ), weakness ( $\alpha_{Ryan} = .90$ ,  $\alpha_{Anna} = .90$ ), likeability ( $r_{Ryan} = .47$ ,  $p < .001$ ;  $r_{Anna} = .48$ ,  $p < .001$ ), relationship satisfaction ( $\alpha = .82$ ).

**Table 4** Means and (Standard Deviations) for Prestige and Gender Ratio Across Status and Occupation in Pilot Study 2

Variable	Status				
	(Ryan Absolute/Anna Relative)	Occupation – Ryan		Occupation – Anna	
	Ryan Low/Anna Lower	Bookbinder	Tailor	Dishwasher	Bellhop
Prestige		2.54 (1.20)	2.67 (1.13)	1.61 (0.94)	2.06 (1.06)
Gender Ratio		48.92 (10.95)	47.31 (17.49)	41.67 (13.49)	40.60 (18.81)
	Ryan Low/Anna Equal	Bookbinder	Tailor	Upholsterer	Food Preparation Worker
Prestige		2.54 (1.20)	2.67 (1.13)	2.46 (1.13)	2.51 (1.17)
Gender Ratio		48.92 (10.95)	47.31 (17.49)	46.54 (20.72)	56.54 (11.21)
	Ryan Low/Anna Higher	Bookbinder	Tailor	Radiologist	Architect
Prestige		2.54 (1.20)	2.67 (1.13)	5.04 (1.47)	5.29 (1.13)
Gender Ratio		48.92 (10.95)	47.31 (17.49)	47.67 (12.59)	42.62 (12.60)
	Ryan Medium/Anna Lower	Geologist	Police Officer	Factory Worker	Bartender
Prestige		4.27 (1.30)	4.08 (1.22)	2.10 (1.02)	2.25 (1.14)
Gender Ratio		42.62 (11.67)	42.94 (8.97)	45.23 (17.23)	49.37 (8.61)
	Ryan Medium/Anna Equal	Geologist	Police Officer	Management Analyst	Business Operations Specialist
Prestige		4.27 (1.30)	4.08 (1.22)	3.92 (1.13)	4.18 (1.23)
Gender Ratio		42.62 (11.67)	42.94 (8.97)	45.10 (13.70)	42.50 (13.86)
	Ryan Medium/Anna Higher	Geologist	Police Officer	Lawyer	Professor
Prestige		4.27 (1.30)	4.08 (1.22)	5.71 (1.29)	5.96 (1.33)
Gender Ratio		42.62 (11.67)	42.94 (8.97)	52.01 (10.89)	39.17 (13.07)

## Study 2 Results

### Preliminary Analyses

Correlations showed that participants' educational level, employment status (wages vs. students), and whether they filled out the survey via Prolific Academic were all correlated with several dependent variables (see Table 5). Therefore, we controlled for these variables. Again, we also conducted the analyses without these covariates, which did not show major differences (see Supplement B in the online supplement for the results without the covariates). Similar to Study 1, we controlled for the couple's average societal status in all analyses. Again, Ryan's absolute status did not affect how stereotypically Ryan and Anna were perceived, nor how their relationship was perceived. Therefore, we decided not to report these effects again in this study. The manipulation worked as intended, such that participants perceived Anna's status relative to Ryan also as intended across the conditions (see Supplement E in the online supplement). Q-Q plots showed that all hypothesized variables were normally distributed. There were no missing data in the final sample.

### Perceptions of Agency, Dominance, and Weakness in the Role-Reversed Relationships

Consistent with Hypothesis 1, we found an interaction effect for partner gender and Anna's relative status on perceived relative agency (see Table 6). Participants rated the partner with the higher status in the relationship to be more agentic than the other partner.

Consistent with Hypothesis 1, we found an interaction effect for partner gender and Anna's relative status on perceived relative dominance (see Table 6). Participants rated Anna and Ryan to be equally dominant when they had a traditional or equal status division. When Anna had higher social status than Ryan, participants rated her to be more dominant than her partner.

Contrary to Study 1 and Hypothesis 2, Ryan was not rated as weaker than Anna in conditions where Anna had higher status than Ryan. There were no significant main or interaction effects (see Table 6).

### Structural Equation Modelling to Test Our Theorized Model

We started our analysis by building a structural equation model in Mplus identical to the final model in Study 1 (see Fig. 3). This model provided a good fit with the data,  $\chi^2(df=22)=53.84$ ,  $p < .001$ , RMSEA = .08, CFI = .94, SRMR = .06. Similar to Study 1, this model showed better fit than our initial

theoretical model (see Fig. 1),  $\Delta\chi^2(\Delta df=2)=60.29$ ,  $p < .001$ , RMSEA = .13, CFI = .82, SRMR = .16 (see Supplement F in the online supplement for alternative models). Furthermore, we conducted a multiple group comparison in SEM to compare whether there were significant differences between the American sample (Study 1) and the Dutch sample (Study 2) for the final model. The constrained model provided no worse fit than the unconstrained model,  $\chi^2_{diff}(df=17)=23.45$ ,  $p = .135$ , indicating that there were no significant differences between the two samples for the final models.

### Perceived Relationship Satisfaction in the Role-Reversed Relationship

Similar to Study 1 and in line with Hypothesis 3, participants in the role-reversed condition rated Anna as more dominant than Ryan as compared to participants in the other conditions,  $b = .20$ ,  $SE = .07$ ,  $p = .003$ . Contrary to Study 1 and Hypothesis 3, Ryan was not rated as weaker than Anna in the role-reversed relationship,  $b = .12$ ,  $SE = .07$ ,  $p = .075$  (see Fig. 3). Anna's perceived relative dominance predicted less perceived relationship satisfaction,  $b = -.21$ ,  $SE = .06$ ,  $p = .001$ , but this was not the case for perceived relative weakness,  $b = -.01$ ,  $SE = .06$ ,  $p = .829$ . Thus, contrary to Study 1 and Hypothesis 3, participants' perceptions of Ryan's relative weakness were not associated with perceived relationship satisfaction. We found a significant indirect effect of condition on perceived relationship satisfaction via Anna's perceived dominance,  $b = -.04$ ,  $SE = .02$ ,  $p = .028$ , but not via Ryan's perceived weakness,  $b = -.00$ ,  $SE = .01$ ,  $p = .830$ . The overall indirect effect was significant,  $b = -.04$ ,  $SE = .02$ ,  $p = .027$ , indicating that participants perceived the role-reversed relationship as less satisfying due to the combined perception that Anna is more dominant than Ryan and Ryan is weaker than Anna. There was no direct effect of condition on perceived relationship satisfaction,  $b = .07$ ,  $SE = .07$ ,  $p = .325$ .

### Perceived Likeability of Anna in the Role-Reversed Relationship

Consistent with Hypothesis 4, when participants rated Anna as more dominant than Ryan in the relationship, they liked her less,  $b = -.49$ ,  $SE = .06$ ,  $p < .001$ , but not Ryan (see Fig. 3). We found an indirect effect of condition on how much Anna was liked via Anna's perceived relative dominance,  $b = -.10$ ,  $SE = .04$ ,  $p = .006$ . When participants rated Anna as more agentic than Ryan, they liked her more,  $b = .17$ ,  $SE = .07$ ,  $p = .016$ . There was an indirect effect of condition on how much Anna was liked via Anna's perceived relative agency,  $b = .07$ ,  $SE = .03$ ,  $p = .023$ . Furthermore, we found no overall indirect effect,  $b = -.03$ ,  $SE = .04$ ,  $p = .526$ ,

**Table 5** Zero-Order Correlations Among Background Variables, Predictor Variables (Dummy-Coded) and Outcome Variables in Study 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1. Participants' Gender: Women/ Men	-																									
2. Age	.06	-																								
3. Education	.12*	.41**	-																							
4. Ethnicity: Dutch/Other	.00	-.03	.04	-																						
5. Married/Single	-.09	-.68**	-.48**	.05	-																					
6. Married/Other	.03	-.49**	-.21**	-.02	-																					
7. Wages/Students	-.01	-.73**	-.58**	.06	.72**	.48**	-																			
8. Wages/Other	.00	-.08	-.25**	.04	.14	.06	-																			
9. Prolific/Other	-.31**	-.16**	-.08	.04	.34**	.06	.06	.13	-																	
10. Status of Couple	.12*	.07	.04	.01	-.11	.01	-.03	.13	-.06	-																
11. Status Ryan: High/Low	-.01	.06	.02	.04	.02	-.02	-.02	.08	.00	.40**	-															
12. Relative Status Anna: Higher/Other	.05	.02	.07	.04	.13*	-.01	-.01	.01	.03	.37**	.01	-														
13. Relative Status Anna: Lower/Other	-.01	.02	.02	-.11*	-.01	-.06	.01	-.04	-.05	-.41**	-.01	-.51**	-													
14. Agency Ryan	.11	.06	-.07	-.09	-.03	-.05	-.02	.11	-.02	.27**	.31**	-.01	.06	-												
15. Agency Anna	.09	-.04	.01	-.07	.03	.05	.04	.05	-.02	.47**	.17**	.42**	-.38**	.30**	-											
16. Dominance Ryan	-.06	-.01	-.03	.09	-.01	-.01	.10	.22**	.05	.14*	.24**	-.06	.04	.32**	-.02	-										
17. Dominance Anna	-.07	-.11	.00	.00	.06	.13	.19**	.21*	.02	.19**	.20**	.12	-.14*	-.02	.27**	.47**	-									
18. Weakness Ryan	-.06	-.06	.07	-.01	.09	.11	.12	.07	.10	-.12*	-.10	-.04	.03	-.35**	-.16**	.36**	.44**	-								
19. Weakness Anna	-.11	-.01	.11	.10	-.03	.06	.01	.05	.06	-.09	-.01	-.14*	.08	-.05	-.45**	.49**	.35**	.49**	-							
20. Communnality Ryan	.08	-.04	-.04	-.16**	.01	-.01	-.07	-.05	.04	.01	-.19**	.05	-.05	.26**	.37**	-.34**	-.09	-.16**	-.26**	-						
21. Communnality Anna	.07	-.11	-.11	-.06	-.04	-.03	-.03	.02	.06	.07	-.03	-.02	.03	.46**	.30**	.00	-.28**	-.35**	-.18**	.40**	-					
22. Likeability Ryan	.03	-.06	-.06	-.09	.06	.02	.01	-.08	.01	.09	-.04	.06	-.05	.16**	.33**	-.29**	-.10	-.16**	-.32**	.60**	.28**	-				
23. Likeability Anna	.03	-.07	-.07	.04	.00	-.02	-.04	-.01	.00	.08	-.05	.07	-.04	.24**	.19**	.07	-.42**	-.30**	-.22**	.17**	.56**	.42**	-			
24. Respect Ryan	.04	.07	.07	-.01	-.13	-.08	-.27**	.00	-.02	.18**	.29**	.05	.06	.42**	.05	.19**	-.02	-.26**	.08	-.02	.12	.04	.16**	-		
25. Respect Anna	-.04	.07	.07	.01	.02	.16*	-.10	.02	.08	.35**	.13*	.44**	-.36**	-.02	.56**	-.06	.22**	.03	-.27**	.05	-.04	.16**	-.02	.19**	-	
26. Relationship Satisfaction	.05	.00	.00	-.04	-.07	-.09	-.11	-.06	.14*	.15*	-.06	.09	-.04	.24**	.24**	-.03	-.21**	-.18**	-.19**	.29**	.38**	.31**	.37**	.11	.05	

Note. \*  $p < .05$ ; \*\*  $p < .01$ .

**Table 6** Estimated Marginal Means, Standard Errors, and Multi- and Univariate Effects for Gender Backlash Variables in Study 2

Variable	Relative Status			Multivariate Interaction Effect Target Gender x Relative Status	Univariate Between- Condition Effect For Each Target
	Anna Lower than Ryan M (SE) [95% C.I.]	Anna Equal as Ryan M (SE) [95% C.I.]	Anna Higher than Ryan M (SE) [95% C.I.]		
Agency				$F(2,201)=20.84$ , $p<.001$ , $\eta^2=.17$	
Ryan	4.97 (0.13) <sup>a</sup> [4.71, 5.23]	4.55 (0.12) <sup>b</sup> [4.32, 4.78]	4.47 (0.12) <sup>b</sup> [4.24, 4.70]		$F(2,201)=4.07$ , $p=.018$ , $\eta^2=.04$
Anna	4.26 (0.13) <sup>b</sup> [4.00, 4.52]	4.59 (0.12) <sup>b</sup> [4.35, 4.83]	5.17 (0.12) <sup>a</sup> [4.93, 5.40]		$F(2,201)=11.99$ , $p<.001$ , $\eta^2=.11$
Dominance				$F(2,201)=4.91$ , $p=.008$ , $\eta^2=.05$	
Ryan	3.05 (0.14) <sup>ab</sup> [2.78, 3.32]	3.00 (0.12) <sup>ab</sup> [2.75, 3.24]	2.77 (0.12) <sup>a</sup> [2.53, 3.01]		$F(2,201)=1.22$ , $p=.296$ , $\eta^2=.01$
Anna	2.79 (0.13) <sup>ab</sup> [2.53, 3.06]	3.00 (0.12) <sup>ab</sup> [2.76, 3.24]	3.13 (0.20) <sup>b</sup> [2.90, 3.37]		$F(2,201)=1.55$ , $p=.214$ , $\eta^2=.02$
Weakness				$F(2,201)=2.01$ , $p=.137$ , $\eta^2=.02$	
Ryan	2.81 (0.14) <sup>a</sup> [2.54, 3.08]	2.85 (0.13) <sup>a</sup> [2.60, 3.10]	2.86 (0.12) <sup>a</sup> [2.61, 3.10]		$F(2,201)=.03$ , $p=.970$ , $\eta^2=.00$
Anna	2.79 (0.13) <sup>ab</sup> [2.53, 3.06]	2.85 (0.12) <sup>ab</sup> [2.61, 3.09]	2.53 (0.12) <sup>b</sup> [2.30, 2.77]		$F(2,201)=1.83$ , $p=.163$ , $\eta^2=.02$

Note. For every variable, means with different superscripts in rows and/or columns differ significantly from each other,  $p < .05$ .

indicating that the negative indirect effect of Anna's perceived relative dominance and the positive indirect effect of Anna's perceived relative agency on how much she was liked cancelled each other out. There was also no direct effect of condition on Anna's perceived likeability,  $b = .05$ ,  $SE = .06$ ,  $p = .478$ .

### Perceived Respect of Ryan in the Role-Reversed Relationship

Consistent with Hypothesis 5, when participants rated Ryan's as weaker than Anna in the role reversed condition, they had less respect for him,  $b = -.28$ ,  $SE = .06$ ,  $p < .001$ , but not for Anna (see Fig. 3). Unexpectedly and contrary to Study 1, there was no indirect effect of condition on how much Ryan was respected via Ryan's perceived relative weakness,  $b = -.03$ ,  $SE = .02$ ,  $p = .099$ . There was also no direct effect of condition on Ryan's perceived respect,  $b = .00$ ,  $SE = .07$ ,  $p = .989$ .

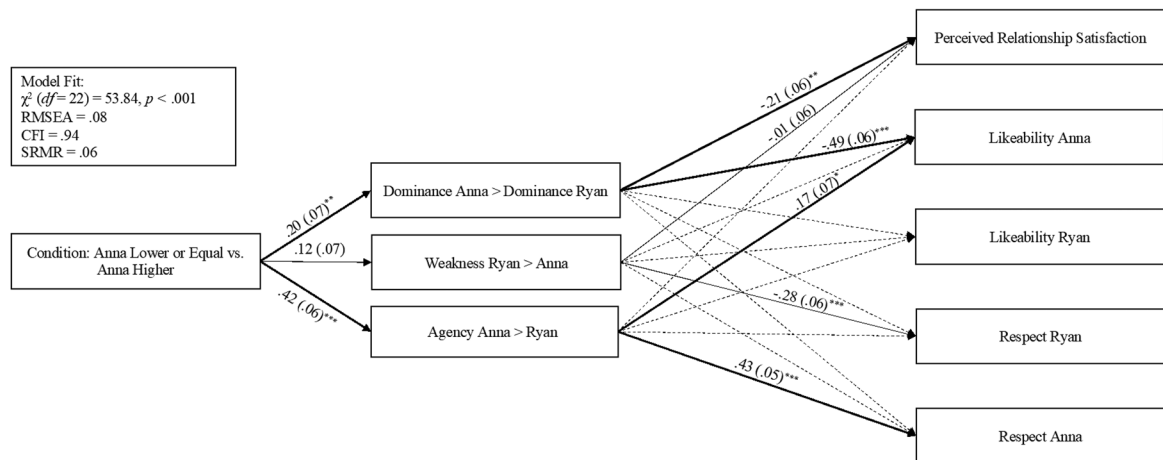
When participants rated Anna as more agentic than Ryan, they had more respect for her,  $b = .43$ ,  $SE = .05$ ,  $p < .001$ . There was also an indirect effect of condition on how much Anna was respected via Anna's perceived relative agency,  $b = .18$ ,  $SE = .03$ ,  $p < .001$ . Moreover, the direct effect of perceived relative status on how much Anna was respected was significant,  $b = .21$ ,  $SE = .06$ ,  $p < .001$ .

## Study 2 Discussion

We again provide evidence that backlash effects occur at the level of the individual as well as the relationship in role-reversed relationships. Regardless of the partner's gender, people rated the partner with the higher occupational status to be more agentic than the other partner. Participants only rated Anna as more dominant in the relationship when she had a higher status occupation than her partner. Though we replicated the direction of this effect, in Study 1, participants rated Anna to be similarly dominant as Ryan when violating traditional gender norms in the relationship. Contrary to Study 1, we did not replicate the weakness penalty for men. We found that role-reversed couples face repercussions both on the relationship level (lower perceived relationship satisfaction) and the individual level (women's likeability). We replicated the finding that women were more respected and perceived as more likeable because of their relative agency in role-reversed relationships.

## Study 3

Our next step was to examine how the backlash mechanisms affect the relationship quality *within* role-reversed relationships. Specifically, we investigated whether women



Note: Thick lines represent significant indirect paths. Reported statistics are standardized coefficients and errors. Dotted lines are paths that were constrained to zero.

**Fig. 3** Results of Theoretical Model in Study 2 Controlling for Average Status of the Couple, Education, Source: Prolific vs. Other, Employment Status: Employed for Wage vs. Student

in role-reversed relationships perceive their male partner to be the weaker one in the relationship, whether men in role-reversed relationships perceive their female partner to be the more dominant one, and whether these perceptions affect their relationship outcomes. In this way, we can examine how backlash mechanisms may operate among couples in role-reversed relationships. Furthermore, as we conducted a dyadic study, we can investigate the extent to which men and women agree about the status division in their relationship. Investigating couples' perceptions allows us to analyze the effects of relative status perceptions on men and women's own outcomes (actor-effects) as well as on their partner's outcomes (partner-effects; Cook & Kenny, 2005). Finally, we will explore whether men and women in role-reversed relationships have similar relative dominance and weakness perceptions of themselves and their partner, and whether these perceptions are related to lower relationship satisfaction.

## Study 3 Method

### Participants and Design

Participants were heterosexual couples in the Netherlands who had been in a relationship for at least one year and of which both partners were over 18 years of age and worked at least 12 h a week. Following Kenny and Cook (1999), we calculated power based on mediation models that were not treated as dyadic models and then translated the sample size for individuals to couples. A priori calculation of multiple linear regression, with an estimated partial  $R^2$  of

.10, a power of .80, and six predictors (i.e., two independent variables and four mediators) in G\*Power showed that we would need a sample size of 130 (Faul et al., 2007). Therefore, we aimed to recruit at least 65 couples. In total, 94 heterosexual couples ( $N = 188$ ) met the requirements. Participants' demographics were similar for both men and women (see Table 7). This study had a dyadic design as we recruited heterosexual partners. The ethics committee of the first author's faculty approved the study (FETC17-043).

### Procedure

Participants were recruited through convenience sampling. We advertised our study on social media, in supermarkets, and we approached eligible couples in our personal network. Potential participants left their contact information, which we coupled to a unique number for each dyad. Upon receiving the link to the survey, each couple received their own code that they filled out at the start of the survey. We asked each participant to share the survey with their partner if they had not already completed it. After collecting the data, we deleted the file with the personal information of participants to assure their anonymity, whereas we could still identify couples through their unique codes.

At the start of the survey, participants gave their informed consent and then completed a survey, including questions regarding their background information (i.e., reporting sex, age in years, highest degree of education, hours working per week, their relationship duration in years, whether they cohabited with their partner, whether they were married to their partner, number of children, age of their youngest child in years, organizational tenure in

**Table 7** Characteristics of Participants from Study 3

Measure	N	Men		Women				
		%	M	SD	N	%	M	SD
Age	105		43.63	13.49	118		41.60	13.47
Highest Education	105				118			
Lower educational degree	3	2.9			4	3.4		
High school degree	6	5.7			4	3.4		
Vocational degree	13	12.4			11	9.3		
College degree/bachelor's degree	45	42.9			45	38.1		
University master's degree	36	34.4			53	44.9		
PhD	2	1.9			1	0.8		
Duration of relationship in years	103		17.06	12.91	117		16.88	12.84
Cohabiting with partner	102				117			
Yes	94	92.2			106	90.6		
No	8	7.8			11	9.4		
Married with partner	103				117			
Yes	49	46.7			55	47		
No	54	52.4			62	53		
Number of children	103		2.36	1.20	117		2.31	1.21
Age youngest child	55		17.22	7.32	10		17.73	7.89
Age child (for parents with one child)	9		11.89	9.47	60		10.10	10.47
Organizational tenure in years	105		10.73	10.39	117		10.88	16.13
Hours working per week	92		39.74	9.24	100		34.45	9.06
Hours working per week when self-employed	13		41.38	6.38	17		35.24	12.40
Net income in euros	103		2952.28	1559.34	111		2204.73	1128.63
Self status in relationship	104		7.44	.90	117		7.20	1.15
Partner status in relationship	104		7.27	1.04	117		7.50	1.32

years, and their monthly net income in euros), social status, and relationship outcomes. We distributed three €100 dinner vouchers through a lottery to couples. The survey took, on average, 15 min to complete.

## Measures

All items were measured on 7-point Likert scales with response options ranging from 1 (*completely disagree*) to 7 (*completely agree*), unless otherwise indicated.

### Perceived Relative Status

We used the same subjective socioeconomic status ladder measure as the previous studies (Adler et al., 2000). In this study, we asked women and men to think about their own situation and to indicate the rung where they would place themselves ( $M_{women} = 7.20$ ,  $SD_{women} = 1.15$ ;  $M_{men} = 7.44$ ,  $SD_{men} = .90$ ) and their partners ( $M_{women} = 7.50$ ,  $SD_{women} = 1.32$ ;  $M_{men} = 7.27$ ,  $SD_{men} = 1.04$ ). We then

measured participants' own perceived relative status by subtracting the man's perceived status from the woman's perceived status.

### Perceived Relative Dominance

To measure relative dominance, we used an adjusted version of Heilman and Wallen (2010) dominance traits by attenuating the tone of the traits. We asked participants the extent to which they felt themselves to be ruthless, dominant and whether they hold the reins in the relationship ( $\alpha_{women} = .69$  and  $\alpha_{men} = .70$ ). We asked participants to indicate the extent to which they felt their partners possess these traits in the relationship ( $\alpha_{women} = .67$  and  $\alpha_{men} = .76$ ). In the original scale, we also included the extent to which participants felt themselves and their partner to be 'firm' in the relationship. Reliability analyses showed that the alphas for the 4-item scale were lower than the alphas for the 3-item scale ( $\alpha_{womenself} = .64$ ;  $\alpha_{menself} = .73$ ;  $\alpha_{womenpartner} = .65$ ;  $\alpha_{menpartner} = .81$ ). We calculated women's perceived relative dominance by subtracting women's own



perception of their dominance from women's perception of their partner's dominance in the relationship. We calculated men's perceived relative dominance by subtracting men's perception of their partner's dominance from men's own perception of dominance in the relationship.

### Perceived Relative Weakness

To measure relative weakness, we used an adjusted version of Heilman and Wallen (2010) weakness traits by again attenuating the tone of the traits. In the original scale, we asked participants to indicate the extent to which they felt themselves to be passive, insecure, compliant, and a push-over in the relationship ( $\alpha_{women} = .38$  and  $\alpha_{men} = .34$ ). Next, we asked participants the extent to which they felt their partners possessed these traits in the relationship ( $\alpha_{women} = .41$  and  $\alpha_{men} = .47$ ). As the reliability of this scale was insufficient, we decided to include two items 'passive' and 'being a push-over' in the final scale ( $r_{womenself} = .22$ ,  $p = .016$ ;  $r_{womenpartner} = .36$ ,  $p < .001$ ;  $r_{menself} = .28$ ,  $p = .004$ ;  $r_{menpartner} = .28$ ,  $p = .005$ ). We measured women's perceived relative weakness by subtracting women's own perception of weakness in the relationship from women's perception of their partners' weakness. We calculated men's perceived relative weakness by subtracting men's perception of their partner's weakness from men's own perception of weakness in the relationship.

Because of the four weakness items' reduced reliability, we further examined whether dominance and weakness are two different constructs. We conducted four principal component analyses (PCA) with orthogonal rotation (varimax): 1) on the eight items for women's own evaluation of their dominance and weakness in the relationship, 2) on the eight items for women's evaluation of their partner's dominance and weakness, 3) on the eight items for men's own evaluation of their dominance and weakness in the relationship and 4) on the eight items for men's evaluation of their partner's dominance and weakness. The Kaiser–Meyer–Olkin (KMO) measures revealed that each PCA's sampling adequacy was mediocre to good for the analysis,  $KMO > .63$ . Bartlett's test of sphericity indicated that correlations between items were sufficiently large for all four PCA's,  $\chi^2(28) > 124.02$ , all  $p$ 's  $< .001$ . After rotation, all factor loadings showed that the items we included in our final analyses represented two different components (see Supplement G in the online supplement for all rotated factor loadings).

### Perceived Relative Agency

We asked participants to rate both themselves and their partner on perceived agency in the relationship. We adapted four items from traits used in the Bem Sex Role Inventory (Bem, 1981) to assess to what extent participants felt

themselves to be competitive and independent in their relationship, whether they think they defend their own beliefs and whether it is easy for them to make decisions in their relationship ( $\alpha_{women} = .52$  and  $\alpha_{men} = .51$ ). Next, we used the same four items to assess the extent to which participants felt their partner is agentic in the relationship ( $\alpha_{women} = .53$  and  $\alpha_{men} = .39$ ). Scale reliability remained insufficient, and we, therefore, decided to exclude relative agency from our analyses.

### Relationship Quality

We measured participants' relationship quality using one item of the time competition survey (Van der Lippe & Glebbeek, 2003). The item ("In general, how satisfied are you with your relationship?") was rated on a 10-point scale ranging from 1 (*very unsatisfied*) to 10 (*very satisfied*). Previous studies established that relationship quality is a construct that can be reliably measured with a single item (see, e.g., Blom & Hewitt, 2019; Hardie et al., 2014).

### Data Analysis Strategy

To test our theoretical model, we conducted a series of Actor-Partner Interdependence Mediation Model analyses (APIMeM; Ledermann et al., 2011) using structural equation modeling in Mplus (Muthén & Muthén, 2008). We treated dyad members as distinguishable as they were in heterosexual relationships. Following Ledermann et al. (2011), we first estimated saturated models and tested all effects, including control variables. Then, we used the step-wise modeling procedure to find the parsimonious model. We excluded control variables that did not significantly predict any variable. We also investigated gender effects by comparing the saturated model with a model in which the actor and partner effects were constrained across gender. This model was not a worse in fit compared to the saturated model,  $\chi^2_{diff}(df = 10) = 8.81$ ,  $p = .551$ ,  $RMSEA = .00$ ,  $CFI = 1.00$ ,  $SRMR = .02$ .

## Study 3 Results

### Preliminary Analyses

Correlations between background, predictor, and outcome variables were analyzed to identify potential covariates (see Table 8). Participants' age, relationship duration, whether they had children, were in a cohabiting relationship, or were married were associated with our most important predictor and outcome variables. For instance, participants' age was associated with women's perception that the men were more dominant in the relationship compared to the women.

Also, being married was associated with higher relationship quality among women, whereas having children was associated with lower levels of relationship quality among women (see Table 8). Interestingly, men and women strongly agreed on the status division within their relationship ( $r = .75, p < .001$ ). There was also convergence in their perception of who is the more dominant one in the relationship ( $r = .56, p < .001$ ), their perception of who is the weaker one in the relationship ( $r = .45, p < .001$ ) and their relationship quality ( $r = .34, p = .001$ ). Not surprisingly, correlations of relevant background variables between men and women (e.g., women reporting about whether they were cohabiting with their partner and their partner reporting about whether they were cohabiting with their partner) were almost completely overlapping ( $r > .94, p < .001$ ).

Q-Q plots showed that all variables were normally distributed. Missing data ranged from 11 cases (8.6%, for relative weakness and dominance according to female partners) to 26 cases (20.3%, for relative weakness and dominance according to male partners and men's perceived relationship quality). Missing values were handled based on full information maximum likelihood (Muthén & Muthén, 2008).

### Actor-Partner Interdependence Mediation Models

First, we tested a saturated APIMeM model including control variables based on the correlational analysis. These were women's age, relationship duration, having children (yes/no), having a cohabiting relationship (yes/no), and being married (yes/no). Furthermore, in this model, we also included the absolute status of the couple as a covariate. We calculated the couple's absolute status by calculating the mean of women's and men's own perceptions of their status ( $M = 7.25, SD = .90$ ). This saturated APIMeM showed that only women's age and whether couples cohabited affected predictor and outcome variables. Therefore, we dropped the other covariates from further analyses to make the models as parsimonious as possible. A model in which we constrained gender effects showed no worse fit compared to the saturated model,  $\chi^2_{diff} (df = 10) = 8.81, p = .551, RMSEA = .00, CFI = 1.00, SRMR = .02$ . This model showed significant actor effects, but no partner effects. For this reason, we tested model fit for a model in which all partner effects were constrained. This model showed no worse fit compared to the model in which we only constrained gender effects,  $\chi^2_{diff} (df = 5) = 2.65, p = .754, RMSEA = .00, CFI = 1.00, SRMR = .03$ .

### Predicting Relationship Quality via Perceived Relative Dominance

Both men and women perceived the woman as the more dominant one in the relationship when they perceived the woman to have the higher status (see Table 9 and Fig. 4). However,

perceived relative dominance was not related to men's and women's relationship satisfaction. The 95% bias-corrected bootstrapping confidence interval of the indirect effect also contained a zero, indicating no indirect effect. Thus, though men and women perceived the woman to be the more dominant one in the relationship when they were in a relationship in which the woman had the higher status, this was not related to their relationship quality (see Table 9 and Fig. 4).

### Predicting Relationship Quality via Perceived Relative Weakness

As shown in Table 9 and Fig. 4, there were no direct effects of men's and women's perceived relative status on their perceived relationship quality. However, men and women who perceived the woman to have higher status in the relationship also rated the man as the weaker one. This perceived relative weakness negatively predicted relationship quality. The 95% bias-corrected bootstrapping confidence intervals (C.I.) showed that the indirect effect of men and women's perceived relative status on their relationship quality via their relative weakness perceptions was significant. Thus, for men and women in role-reversed relationships, perceiving the man to be the weak one in the relationship was related to lower relationship quality (see Table 9 and Fig. 4).

## Study 3 Discussion

In Studies 1 and 2, we showed how backlash operates when observers evaluate a couple in a role-reversed relationship. In Study 3, we show that men and women in role-reversed relationships have similar perceived relative dominance and weakness. They perceived the woman to be the dominant one and the man to be the weak one in the relationship when they believed the woman had a higher social status than her male partner. Furthermore, perceived relative weakness predicted lower relationship quality for couples in role-reversed relationships. In Study 3, we demonstrated that when couples in role-reversed relationships perceived the man to be weaker than the woman the relationship, they reported lower relationship quality compared to couples in traditional relationships.

## General Discussion

Previous research has examined when and why women and men experience backlash when they show behaviors that are incongruent with their gender roles (Eagly & Karau, 2002; Heilman & Okimoto, 2007; Heilman & Wallen, 2010; Rudman et al., 2012). These penalties have mainly been examined in the work domain and for men and women individually. With three studies, we add to this work by showing that backlash also occurs in the relationship domain.

In this case, backlash in the form of perceived dominance of women and perceived weakness of men in role-reversed heterosexual relationships was associated with more negative evaluations of the couple and the individual members of the couple. Specifically, in two experimental studies, we found that a woman with higher status than her male partner risks a dominance penalty; she is perceived to be the more dominant one in the relationship and is therefore disliked more compared to a woman in a traditional or equal status relationship. Moreover, in one of the two experiments, we found that a man with lower status than his partner risks a weakness penalty; he is perceived as the weaker one in the relationship and consequently disrespected more compared to a man in a traditional or equal status relationship. Importantly, beyond these effects at the individual level, we add to the backlash literature by showing that, at the level of relationships, the dominance and weakness penalty also result in the perception that a role-reversed relationship must be less satisfying than a more traditional relationship.

We also demonstrated that this form of social backlash may affect the partners themselves within these relationships when they perceive status divisions that are incongruent with traditional gender roles. Previous research suggests that couples have more intimate knowledge of each other, making them less susceptible to stereotypical judgments than strangers (Trope & Liberman, 2010). However, our results suggest that couples are sensitive to gender norms in their environment and experience negative relationship consequences (e.g., lower relationship quality) when breaking with these norms.

Specifically, both men and women in role-reversed relationships perceive the man as the weaker one and the woman as the more dominant one in the relationship. The perception that the man is the weaker one in the relationship may explain, in part, why couples in role-reversed relationships experience lower relationship satisfaction compared to traditional couples. These findings are a first indication that at least some backlash may spill over to the couples themselves and that couples in role-reversed relationships experience the negative consequences of deviating from the gender hierarchy when male partners have higher status than female partners.

### Buffering Effect of Relative Agency for Women

The perception that the woman is the agentic one in a role-reversed relationship also led to more positive impressions of the woman, such that people liked her more and found her worthy of more respect compared to women in traditional or equal status relationships. In the past half-century, women entered male-dominated roles in larger numbers, making women's agency more common and accepted (Croft et al., 2015; Diekmann & Eagly, 2000; Eagly et al., 2020). Also,

studies show that over the years, women have increasingly described themselves as agentic (Twenge, 1997, 2009). The changing nature of agency's valence might have even led to situations where agency is desired for women (such as in relationships, which we find here). Although agency is usually seen as benefiting the self, in interpersonal relationships agency might also benefit the close other (Abele & Wojciszke, 2007). In a similar vein, it has been shown that the caregiving penalty for working women can be reduced by providing people with information that the woman is the breadwinner of the family in order to provide for her family rather than for pursuing a career for her own benefit (Bear & Glick, 2017). These findings suggest that having higher status because of family-oriented, communal goals can reduce backlash in the relationship domain for women.

The changing nature of prescriptive stereotypes for women may also explain why the perception that the woman is the more dominant one in the relationship did not cause role-reversed couples to report lower relationship satisfaction. Our results suggest that within role-reversed relationships, only men's lower status in the relationship has repercussions for couples' relationship satisfaction, which is consistent with a growing body of research showing that the violation of gendered norms is more penalizing for men than women (Croft et al., 2015; Vandello et al., 2008). An explanation that is often given for this discrepancy is that lower-status groups (i.e., women) aspire to move into higher status groups (i.e., men), but higher status groups may not be so willing to give up their status (Schmader et al., 2001). Also, according to precarious manhood theory, manhood is unstable and more easily lost than femininity, with harsher social penalties as a consequence (Vandello et al., 2008). Together with the literature showing greater social acceptance of agentic women, the current findings suggest that role-reversed couples may experience less relationship satisfaction due to violating the gender stereotypes prescribing men to have high status within the relationship.

### Limitations and Future Research Directions

A limitation across the studies is that we created difference scores of relative status, dominance, weakness, and agency to test the perceived trait divisions within the relationship. It is argued that using difference scores is suboptimal because it causes absolute scores to be aggregated (Cronbach & Furby, 1970). However, in our case, we think this is less of an issue because Anna and Ryan's individual statuses were experimentally manipulated and, therefore, relatively fixed. Furthermore, although we used difference scores as predictors, we also controlled for the absolute average status of the couple to make sure that we investigated the relative effects of the woman in relation to her partner only, rather than effects caused by the absolute status of the man. In the

**Table 8** Correlation Analyses of Relevant Background Variables, Independent Variables, Mediators and Dependent Variables of Study 3

	<i>M</i>	<i>SD</i>	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	
1. Age M	43.63	13.49	-																					
2. Age W	41.60	13.47	.97**	-																				
3. Education M	.17	.13		.38**	-																			
4. Education W	-.08	-.21			.38**	-																		
5. Cohabiting M	-.15	-.15		-.01	-.11	-.01	-																	
6. Cohabiting W	-.22*	-.27**		-.13	.05	.94**		-																
7. Married M	-.55**	-.62**		.01	.25*	.27**	.31**		-															
8. Married W	-.61**	-.64**		.05	.28**	.29**	.30**	1.00**		-														
9. Tenure M	10.73	10.39	.62**	.64**	.09	.01	-.09	-.13	-.31**	-.40**	-													
10. Tenure W	10.88	16.13	.55**	.40**	.03	-.14	-.13	-.12	-.41**	-.38**	.47**	-												
11. Relationship duration M	17.06	12.91	.77**	.81**	.06	-.21*	-.35**	-.35**	-.68**	-.70**	.47**	.50**	-											
12. Relationship duration W	16.88	12.84	.78**	.82**	.04	-.29**	-.32**	-.35**	-.69**	-.70**	.48**	.38**	.99**	-										
13. Children M	.78**	.79**	.79**	.79**	.04	-.11	-.18	-.19	-.61**	-.65**	.44**	.35**	.67**	.66**	-									
14. Children W	.80**	.76**	.76**	.76**	.06	-.16	-.17	-.27**	-.69**	-.67**	.48**	.32**	.70**	.68**	.96**	-								
15. Relative status M	-.17	1.07	-.06	-.11	-.16	.42**	.07	.12	.05	.06	-.09	-.04	-.14	-.14	-.10	-.14	-							
16. Relative status W	-.31	1.39	.00	-.07	.00	.36**	.06	.11	.03	.03	-.07	.09	-.10	-.08	-.06	-.04	.75**	-						
17. Relative dominance M	.08	1.41	-.03	-.01	.05	.03	.00	-.05	-.14	-.14	-.04	.13	.02	.05	.02	.04	.19	.19	-					
18. Relative dominance W	.38	1.22	-.22*	-.22*	.00	.01	-.17	-.06	.04	.02	-.11	-.05	-.06	-.05	-.12	-.17	.15	.22*	.56**	-				
19. Relative weakness M	-.04	1.11	-.12	-.08	-.09	.11	-.08	-.09	-.14	-.15	-.08	-.03	.03	.07	-.09	-.01	.19	.12	.52**	.37**	-			
20. Relative weakness W	.06	.99	-.12	-.10	-.04	.10	-.05	-.15	.03	.02	-.28**	-.02	-.11	-.05	-.08	-.03	.19	.23*	.50**	.45**	.45**	-		
21. Relationship quality M	8.3	0.98	-.12	-.11	-.04	.01	.12	.12	.15	.09	-.01	-.07	-.17	-.16	-.14	-.13	-.08	.05	-.17	-.01	-.11	-.12	-	
22. Relationship quality W	8.47	1.04	-.29**	-.31**	.06	.12	-.08	.00	.25*	.22*	-.13	-.18*	-.28**	-.31**	-.28**	-.31**	-.01	-.08	-.08	.10	-.06	-.19*	.34**	

Note. Cohabiting and Married are dummy variables with 1 = yes, 0 = no

*M* = men, *W* = women

\*  $p < .05$ ; \*\*  $p < .01$

**Table 9** Direct and Indirect Effects for the APIMem on Relationship Quality

Effects	Estimate	SE	<i>p</i>	C.I. 95% bias-corrected
Direct effects				
Actor effects				
RS -> RQ	-.09	.08	.266	[-.25, .08]
<b>RS -&gt; DOM</b>	<b>.20</b>	<b>.08</b>	<b>.017</b>	<b> [.03, .35]</b>
<b>RS -&gt; WEAK</b>	<b>.21</b>	<b>.07</b>	<b>.001</b>	<b> [.09, .35]</b>
DOM -> RQ	.03	.09	.716	[-.14, .22]
<b>WEAK -&gt; RQ</b>	<b>-.17</b>	<b>.08</b>	<b>.031</b>	<b> [-.34, -.02]</b>
Indirect effects				
Total effect	-.12	.08	.129	[-.29, .03]
Total indirect effect	-.03	.02	.159	[-.08, .00]
RS -> DOM -> RQ	.01	.02	.741	[-.03, .06]
<b>RS -&gt; WEAK -&gt; RQ</b>	<b>-.04</b>	<b>.02</b>	<b>.070</b>	<b> [-.09, -.01]</b>

Note. Following Ledermann et al. (2011), all variables are standardized. The final results reveal no gender differences and partner effects. Therefore, only actor effects are reported. Significant coefficients are in bold. We controlled for women's age and whether couples cohabited.

APIMem actor-partner interdependence mediation model, RS perceived relative status of the woman compared to the man, DOM perceived dominance of the woman compared to the man, WEAK perceived weakness of the man compared to the woman, RQ relationship quality.

dyadic study, we find that the outcomes were not affected by couples' average absolute status, only by their relative status differences.

Another limitation is that in Study 3, we measured dominance and weakness traits of couples *within* the relationship, whereas we measured the perceived dominance and weakness of the fictitious couples in general (e.g., Anna is dominant) in Studies 1 and 2. We decided not to assess perceived dominance and weakness within the relationship in Studies 1 and 2 to keep the measures closely related to the original scale by Heilman and Wallen (2010). As couples have a more detailed construal of their partner than strangers (Trope & Liberman, 2010), we decided to provide more context for these items in Study 3 by asking participants to rate themselves and their partner on traits in the relationship rather than in general. Future research should investigate whether similar findings are found when asking strangers to rate couples' dominance and weakness within their relationship as we found here for ratings of dominance and weakness more generally.

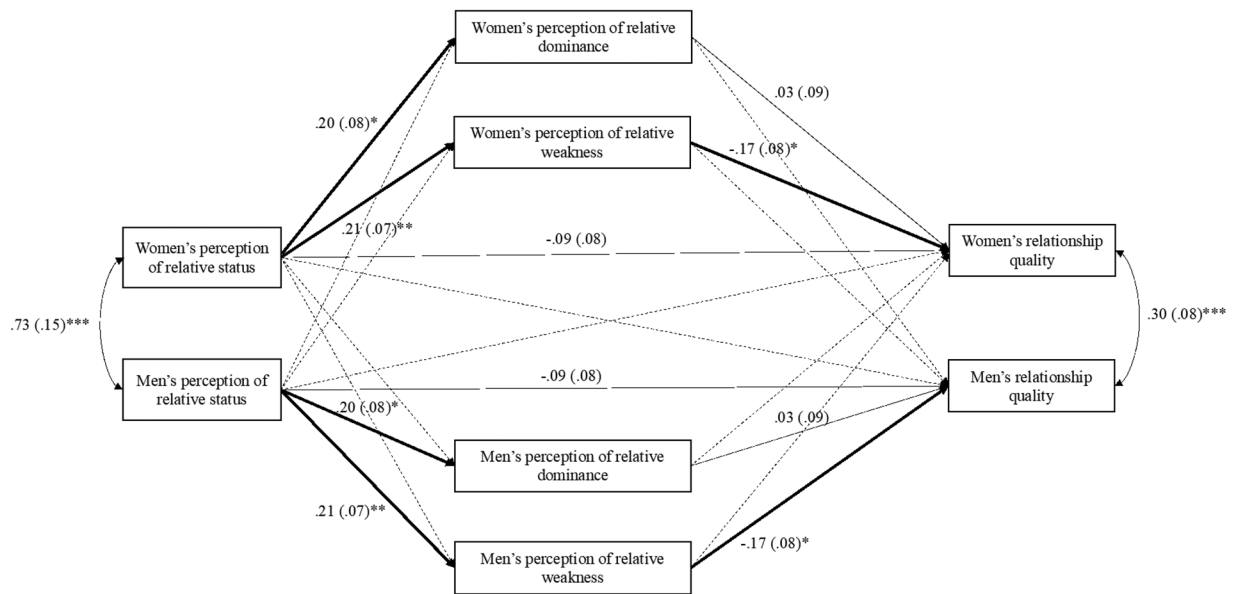
Additionally, as we only investigated explicit reports of relationship satisfaction, more research is needed to investigate the effects of self and partner evaluations on a nonconscious or implicit level. A growing body of research shows that explicit partner evaluations are not associated with implicit partner evaluations and that these implicit partner evaluations are predictive of marital satisfaction over time (McNulty et al., 2013). An often-given explanation for the

discrepancy between explicit and implicit relational evaluations is that individuals are motivated to see their relationship in a positive light (McNulty et al., 2014; Murray, 1999; Olson et al., 2007). Although these implicit evaluations of self and relationship mainly include valence (i.e., having a positive versus negative implicit association of one's partner), the findings might indicate a potential drawback of our study. It could be that the explicit evaluations of relative dominance and weakness and experienced relationship satisfaction might not have accurately captured individuals' implicit attitudes of relative dominance, weakness, and relationships satisfaction (e.g., Joel et al., 2017). Future research could examine whether backlash effects also operate at the implicit level.

Furthermore, we have investigated backlash effects in heterosexual relationships only, which raises the question to what extent the findings can be translated to same-gender relationships. It may be the case that men and women in same-gender relationships are less susceptible for gender stereotypes within our society as these relationships have less clear gender role prescriptions. On the other hand, there are also some indications that same-gender couples do not differ that much from heterosexual couples in their extent to which they value status in their partners (Ha et al., 2012; Lippa, 2007). Furthermore, same-gender couples are also susceptible to the influences of gender stereotypes in how others view them (i.e., gay men and lesbians are stereotyped as 'feminine' and 'masculine' by virtue of their respective orientations towards male and female partners; Hegarty & Pratto, 2001; Kite & Deaux, 1987). Future research may for instance investigate whether women who have surpassed their female partner in status have fewer negative evaluations of their partner compared to women who have surpassed a male partner. By including gay men and women in future research on the influence of gender stereotypes for couples who break with traditional gender stereotypes, we will develop a better understanding of when and how gender stereotypes affect the relationship outcomes of non-traditional couples more broadly.

## Practice Implications

The findings of these studies are in line with other work showing the social penalties that men and women may face when their relationship is not meeting gendered expectations for the division of status (e.g., Bertrand et al., 2015; Pierce et al., 2013; Wilcox & Nock, 2006; Zhang, 2015). We would like to stress that these findings do not imply that traditional relationships are the most desirable or optimal relationships that couples should strive for. There is strong evidence suggesting that couples who adhere to traditional gender role divisions also experience negative relationship outcomes and may be perceived negatively (Hammond & Overall, 2013;



Note: Bold paths are significant paths. These effects are controlled for women's age and whether participants cohabited.

**Fig. 4** Results of Final APIMeM model

Helms et al., 2006; Marshall, 2010; Sanchez et al., 2005). For example, while both men and women desire warmth, affection, and understanding when they devote themselves to intimate relationships (Reis & Gable, 2003), individuals low in feminine traits are less likely to experience these outcomes (Miller et al., 2003). In addition, women who perceived their partner to be a feminist report better relationship outcomes (Rudman & Phelan, 2007). Scholars have argued that the most adjusted and happiest individuals in life possess both agentic and communal qualities (Stake & Eisele, 2010). Applied to romantic relationships, individuals who possess both these traits are likely to be seen as desirable spouses and to have satisfied partners (Marshall, 2010). Past research findings on interpersonal outcomes within couples highlight how processes within couples affect their outcomes, whereas our findings suggest that group-related processes (i.e., gender-stereotypical expectations) may affect relationships as well, such that role-reversed couples partially internalize backlash effects.

Future research should continue to explore how the negative evaluations of others outside the relationship directly affect experiences and behaviors of role-reversed couples. Couples are not passive bystanders and may also be influenced by the extent to which they themselves internalize and endorse traditional gender role divisions. For example, egalitarian individuals have more favorable perceptions of role-reversed couples (Gaunt, 2013). Within couples, young women with high ambitions seek a communal and family-oriented male partner

(Meeussen et al., 2019). Men who endorse hostile sexist attitudes (i.e., perceiving women who challenge men's status as manipulative and subversive) behave more negatively toward their romantic partners and experience lower relationship satisfaction compared to men who do not endorse these attitudes (Hammond & Overall, 2013). The current findings suggest that the gender and sexist stereotypes in couples' environments may have a negative impact on role-reversed couples regardless of couples' own egalitarian views. This is supported by other past work showing that role-reversed couples experience more difficulties when they live in countries that still endorse traditional gender norms (Vink et al., 2022b). In order to help couples' circumvent the negative consequences of performing non-traditional gender roles, it may not be enough to solely focus on couples' own experiences, but rather the systemic context should be considered. This perspective applies not only to researchers investigating role-reversed couples' experiences, but also to relationship therapists, and employers or coaches who aim to achieve gender equality.

## Conclusion

Not only do men and women risk penalties from others when they violate traditional gender norms of close relationships, men and women who perceive the woman to have higher status than the man in the relationship also internalize and may exact, at least in part, this backlash in their own relationship. Backlash occurs for couples who

challenge the gender status quo by occupying a social status that runs counter to traditional gender norms. The gender norms that proscribe dominance for women, especially relative to their male partner, and proscribe weakness for men, especially relative to their female partner, leads to less liking of women and less respect for men in those relationships, as well as evaluating role-reversed relationships as less satisfying for the partners. Moreover, men's and women's own subjective experience of breaking from traditional status divisions was also related to less relationship satisfaction, which was explained by perceiving the male partner to be the weaker one in the relationship. Overall, these findings suggest that backlash effects for role-reversed, heterosexual relationships are another way in which the gender hierarchy is protected and why traditional gender roles are persistent and difficult to change.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11199-022-01339-5>.

**Acknowledgements** This research is part of the research program Sustainable Cooperation – Roadmaps to Resilient Societies (SCOOP). The authors are grateful to the Netherlands Organization for Scientific Research (NWO) and the Dutch Ministry of Education, Culture and Science (OCW) for generously funding this research in the context of its 2017 Gravitation Program (grant number 024.003.025).

**Funding** This study was supported by the Netherlands Organization for Scientific Research (NWO) and the Dutch Ministry of Education, Culture and Science (OCW; 2017 Gravitation Program, grant number 024.003.025).

**Availability of Data and Materials** (Anonymized) datasets and materials for all three studies can be made available upon request to the first author.

**Code Availability** Coding reports/syntaxes for the data analysis of all three studies can be made available upon request to the first author.

## Compliance with Ethical Standards

**Ethical Approval** The studies involved human subjects, and all procedures performed were in accordance with the ethical standards of the institutional and/or national research committee. For the dyadic study, we requested and received ethical approval by the Faculty's Ethical Board, as this study required participants to answer questions about their own romantic relationship and required participants' partner to fill out the survey as well. For the experiments, we requested and received post-hoc ethical approval as requesting ethical approval was less commonplace at the moment of data collection.

**Informed Consent** Each participant in all three studies completed the informed consent process.

**Consent for Publication** We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal.

**Conflicts of Interest** The authors declare that there are no potential conflicts of interest with regards to ethical standards.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*, 751–763. <https://doi.org/10.1037/0022-3514.93.5.751>
- Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology*, *19*, 586–592. <https://doi.org/10.1037/0278-6133.19.6.586>
- Amanatullah, E. T., & Morris, M. W. (2010). Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology*, *98*, 256–267. <https://doi.org/10.1037/a0017094>
- Bear, J. B., & Glick, P. (2017). Breadwinner bonus and caregiver penalty in workplace rewards for men and women. *Social Psychological and Personality Science*, *8*, 780–788. <https://doi.org/10.1177/1948550616683016>
- Bem, S. L. (1981). The BSRI and gender schema theory: A reply to Spence and Helmreich. *Psychological Review*, *88*, 369–371. <https://doi.org/10.1037/0033-295X.88.4.369>
- Bertrand, M., Kamenica, E., & Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, *130*, 571–614. <https://doi.org/10.1093/qje/qjv001>
- Blom, N., & Hewitt, B. (2019). Becoming a female-breadwinner household in Australia: Changes in relationship satisfaction. *Journal of Marriage and Family*, 1–18. <https://doi.org/10.1111/jomf.12653>
- Brescoll, V. L., & Uhlmann, E. L. (2005). Attitudes toward traditional and role-reversed parents. *Psychology of Women Quarterly*, *29*, 436–445. <https://doi.org/10.1111/j.1471-6402.2005.00244.x>
- Cadinu, M., Latrofa, M., & Carnaghi, A. (2013). Comparing self-stereotyping with in-group-stereotyping and out-group-stereotyping in unequal-status groups: the case of gender. *Self and Identity*, *12*, 582–596. <https://doi.org/10.1080/15298868.2012.712753>
- Chaney, K. E., Rudman, L. A., Fetterolf, J. C., & Young, D. M. (2019). Paying a price for domestic equality: Risk factors for backlash against role-reversed husbands. *Gender Issues*, 1–20. <https://doi.org/10.1007/s12147-017-9207-8>
- Cherry, F., & Deaux, K. (1978). Fear of success versus fear of gender-inappropriate behavior. *Sex Roles*, *4*, 97–101. <https://doi.org/10.1007/BF00288380>
- Cook, W. L., & Kenny, D. A. (2005). The actor-partner interdependence model: A model of bidirectional effects in developmental studies. *International Journal of Behavioral Development*, *29*, 101–109. <https://doi.org/10.1080/2F01650250444000405>
- Croft, A., Schmader, T., & Block, K. (2015). An underexamined inequality: Cultural and psychological barriers to men's engagement

- with communal roles. *Personality and Social Psychology Review*, 19, 343–370. <https://doi.org/10.1177/1088868314564789>
- Cronbach, L. J., & Furby, L. (1970). How we should measure “change” – or should we? *Psychological Bulletin*, 74, 68–80. <https://doi.org/10.1037/h0029382>
- Diekmann, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, 26, 1171–1188. <https://doi.org/10.1177/0146167200262001>
- Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Erlbaum.
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598. <https://doi.org/10.1037/0033-295X.109.3.573>
- Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, 75, 301–315. <https://doi.org/10.1037/amp0000494>
- Eagly, A. H., Wood, W., & Diekmann, A. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). Erlbaum.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Gaunt, R. (2013). Breadwinning moms, caregiving dads. *Journal of Family Issues*, 34, 3–24. <https://doi.org/10.1177/0192513X12438686>
- Glick, P., Wilk, K., & Perreault, M. (1995). Images of occupations: Components of gender and status in occupational stereotypes. *Sex Roles*, 32, 565–582. <https://doi.org/10.1007/BF01544212>
- Ha, T., Van Den Berg, J. E., Engels, R. C., & Lichtwarck-Aschoff, A. (2012). Effects of attractiveness and status in dating desire in homosexual and heterosexual men and women. *Archives of Sexual Behavior*, 41, 673–682. <https://doi.org/10.1007/s10508-011-9855-9>
- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times they are a-changing...or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40, 353–363. <https://doi.org/10.1177/0361684316634081>
- Hammond, M. D., & Overall, N. C. (2013). Men’s hostile sexism and biased perceptions of intimate partners: Fostering dissatisfaction and negative behavior in close relationships. *Personality and Social Psychology Bulletin*, 39, 1585–1599. <https://doi.org/10.1177/0146167213499026>
- Hardie, J. H., Geist, C., & Lucas, A. (2014). His and hers: Economic factors and relationship quality in Germany. *Journal of Marriage and Family*, 76, 728–743. <https://doi.org/10.1111/jomf.12129>
- Hegarty, P., & Pratto, F. (2001). Sexual orientation beliefs: Their relationship to anti-gay attitudes and biological determinist arguments. *Journal of Homosexuality*, 41, 121–135. [https://doi.org/10.1300/J082v41n01\\_04](https://doi.org/10.1300/J082v41n01_04)
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women’s ascent up the organizational ladder. *Journal of Social Issues*, 57, 657–674. <https://doi.org/10.1111/0022-4537.00234>
- Heilman, M. E., & Okimoto, T. G. (2007). Why are women penalized for success at male tasks? The implied communality deficit. *Journal of Applied Psychology*, 92, 81–92. <https://doi.org/10.1037/0021-9010.92.1.81>
- Heilman, M. E., & Wallen, A. S. (2010). Wimpy and undeserving of respect: Penalties for men’s gender-inconsistent success. *Journal of Experimental Social Psychology*, 46, 664–667. <https://doi.org/10.1016/j.jesp.2010.01.008>
- Helms, H. M., Proulx, C. M., Klute, M. M., McHale, S. M., & Crouter, A. C. (2006). Spouses’ gender-typed attributes and their links with marital quality: A pattern analytic approach. *Journal of Social and Personal Relationships*, 23, 843–864. <https://doi.org/10.1177/2F0265407506068266>
- Hettinger, V. E., Hutchinson, D. M., & Bosson, J. K. (2014). Influence of professional status on perceptions of romantic relationship dynamics. *Psychology of Men & Masculinity*, 15, 470. <https://doi.org/10.1037/a0034034>
- Joel, S., Eastwick, P. W., & Finkel, E. J. (2017). Is romantic desire predictable? Machine learning applied to initial romantic attraction. *Psychological Science*, 28, 1478–1489. <https://doi.org/10.1177/0956797617714580>
- Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25, 881–919. <https://doi.org/10.1111/j.1467-9221.2004.00402.x>
- Kenny, D. A., & Cook, W. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustrations. *Personal Relationships*, 6, 433–448. <https://doi.org/10.1111/j.1475-6811.1999.tb00202.x>
- Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly*, 11, 83–96. <https://doi.org/10.1111/j.1471-6402.1987.tb00776.x>
- Ledermann, T., Macho, S., & Kenny, D. A. (2011). Assessing mediation in dyadic data using the actor partner interdependence model. *Structural Equation Modeling: a Multidisciplinary Journal*, 18, 595–612. <https://doi.org/10.1080/10705511.2011.607099>
- Lippa, R. A. (2007). The preferred traits of mates in a cross-national study of heterosexual and homosexual men and women: An examination of biological and cultural influences. *Archives of Sexual Behavior*, 36, 193–208. <https://doi.org/10.1007/s10508-006-9151-2>
- MacInnis, C. C., & Buliga, E. (2020). Don’t get above yourself: Heterosexual cross-class couples are viewed less favorably. *Psychology of Women Quarterly*, 44, 50–66. <https://doi.org/10.1177/0361684319878459>
- Marshall, T. C. (2010). Gender, peer relations, and intimate romantic relationships. In Chrisler J., McCreary D. (Eds.), *Handbook of gender research in psychology*. Springer. [https://doi.org/10.1007/978-1-4419-1467-5\\_12](https://doi.org/10.1007/978-1-4419-1467-5_12)
- McNulty, J. K., Baker, L. R., & Olson, M. A. (2014). Implicit self-evaluations predict changes in implicit partner evaluations. *Psychological Science*, 25, 1649–1657. <https://doi.org/10.1177/0956797614537833>
- McNulty, J. K., Olson, M. A., Meltzer, A. L., & Shaffer, M. J. (2013). Though they may be unaware, newlyweds implicitly know whether their marriage will be satisfying. *Science*, 342, 1119–1120. <https://doi.org/10.1126/science.1243140>
- Meeussen, L., van Laar, C., & Verbruggen, M. (2019). Looking for a family man? Norms for men are toppling in heterosexual relationships. *Sex Roles*, 80, 429–442. <https://doi.org/10.1007/s11199-018-0946-0>
- Miller, P. J. E., Caughlin, J. P., & Huston, T. L. (2003). Trait expressiveness and marital satisfaction: The role of idealization processes. *Journal of Marriage and the Family*, 65, 978–995. <https://doi.org/10.1111/j.1741-3737.2003.00978.x>
- Morgenroth, T., & Heilman, M. E. (2017). Should I stay or should I go? Implications of maternity leave choice for perceptions of working mothers. *Journal of Experimental Social Psychology*, 72, 53–56. <https://doi.org/10.1016/j.jesp.2017.04.008>
- Murray, S. L. (1999). The quest for conviction: Motivated cognition in romantic relationships. *Psychological Inquiry*, 10, 23–34. [https://doi.org/10.1207/s15327965pli1001\\_3](https://doi.org/10.1207/s15327965pli1001_3)
- Muthén, L. K., & Muthén, B. O. (2008). *Mplus* (Version 5.1). Muthén & Muthén.



- Olson, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, *18*, 287–291. <https://doi.org/10.1111/2Fj.1467-9280.2007.01890.x>
- Organisation for Economic Cooperation and Development. (2018). *Employment outlook 2018 (statistical annex)*. [https://read.oecd-ilibrary.org/employment/oecd-employment-outlook-2018\\_empl\\_outlook-2018-en#page1](https://read.oecd-ilibrary.org/employment/oecd-employment-outlook-2018_empl_outlook-2018-en#page1)
- Park, B., Smith, J., & Correll, J. (2010). The persistence of implicit behavioral associations for moms and dads. *Journal of Experimental Social Psychology*, *46*, 809–815. <https://doi.org/10.1016/j.jesp.2010.04.009>
- Pew Research Center. (2013). *Breadwinner moms*. <https://www.pewsocialtrends.org/2013/05/29/breadwinner-moms/>
- Pierce, L., Dahl, M. S., & Nielsen, J. (2013). In sickness and in wealth: Psychological and sexual costs of income comparison in marriage. *Personality and Social Psychology Bulletin*, *39*, 359–374. <https://doi.org/10.1177/0146167212475321>
- Portegijs, W., & Van den Brakel, M. (2018). *Emancipatiemonitor 2018*. <https://digitaal.scp.nl/emancipatiemonitor2018/assets/pdf/emancipatiemonitor-2018-SCP.pdf>
- Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, *26*, 269–281. <https://doi.org/10.1111/1471-6402.t01-1-00066>
- Reis, H. T., & Gable, S. L. (2003). Toward a positive psychology of relationships. In C. L. M. Keyes, & J. Haidt (Eds.), *Flourishing: Positive psychology and the life well-lived* (pp. 129–159). American Psychological Association. <https://doi.org/10.1037/10594-006>
- Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, *57*, 637–655. <https://doi.org/10.1111/0022-4537.00233>
- Rudman, L. A. (1998). Self-promotion as a risk factor for women: The costs and benefits of counter stereotypical impression management. *Journal of Personality and Social Psychology*, *74*, 629–645. <https://doi.org/10.1037/0022-3514.74.3.629>
- Rudman, L. A., & Fairchild, K. (2004). Reactions to counter stereotypic behavior: The role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology*, *87*, 157–176. <https://doi.org/10.1037/0022-3514.87.2.157>
- Rudman, L. A., & Glick, P. (1999). Implicit gender stereotypes and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of managers. *Journal of Personality and Social Psychology*, *77*, 1004–1010. <https://doi.org/10.1037/0022-3514.77.5.1004>
- Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin*, *26*, 1315–1328. <https://doi.org/10.1177/0146167200263001>
- Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice toward female leaders. *Journal of Experimental Social Psychology*, *48*, 165–179. <https://doi.org/10.1016/j.jesp.2011.10.008>
- Rudman, L. A., & Phelan, J. E. (2007). The interpersonal power of feminism: Is feminism good for romantic relationships? *Sex Roles*, *57*, 787–799. <https://doi.org/10.1007/s11199-007-9319-9>
- Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The investment model scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships*, *5*, 357–391. <https://doi.org/10.1111/j.1475-6811.1998.tb00177.x>
- Sanchez, D. T., Crocker, J., & Boike, K. R. (2005). Doing gender in the bedroom: Investing in gender norms and the sexual experience. *Personality and Social Psychology Bulletin*, *31*, 1445–1455. <https://doi.org/10.1177/0146167205277333>
- Schmader, T., Major, B., Eccleston, C. P., & McCoy, S. K. (2001). Devaluing domains in response to threatening intergroup comparisons: Perceived legitimacy and the status value asymmetry. *Journal of Personality and Social Psychology*, *80*, 782. <https://doi.org/10.1037/0022-3514.80.5.782>
- Stake, J. E., & Eisele, H. (2010). Gender and personality. In Chrisler J., McCreary D. (Eds.), *Handbook of gender research in psychology* (pp. 19–40). Springer.
- Swanborn, P. (2010). Case study research: What, why and how? *Sage Publications*. <https://doi.org/10.4135/9781526485168>
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*, 440–463. <https://doi.org/10.1037/a0018963>
- Twenge, J. M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex Roles*, *36*, 305–325. <https://doi.org/10.1007/BF02766650>
- Twenge, J. M. (2009). Status and gender: The paradox of progress in an age of narcissism. *Sex Roles*, *61*, 338–340. <https://doi.org/10.1007/s11199-009-9617-5>
- Van der Lippe, T., & Glebbeek, A. (2003). *Time competition survey*. Utrecht University/University of Groningen.
- Vandello, J. A., Bosson, J. K., Cohen, D., Burnaford, R. M., & Weaver, J. R. (2008). Precarious manhood. *Journal of Personality and Social Psychology*, *95*, 1325–1339. <https://doi.org/10.1037/a0012453>
- Vink, M., Derks, B., Ellemers, N., & Van der Lippe (2022a). *All is nice and well unless she outshines him: Higher social status benefits women's well-being and relationship quality but not if they surpass their male partner*. [Manuscript submitted for publication]. Department of Social, Health, and Organizational Psychology, Utrecht University.
- Vink, M., van der Lippe, T., Derks, B., & Ellemers, N. (2022b). Does national context matter when women surpass their partner in status? *Frontiers in Psychology*, *12*, 1–13. [670439]. <https://doi.org/10.3389/fpsyg.2021.670439>
- Wallen, A. S., Morris, M. W., Devine, B. A., & Lu, J. G. (2017). Understanding the MBA gender gap: Women respond to gender norms by reducing public assertiveness but not private effort. *Personality and Social Psychology Bulletin*, *43*, 1150–1170. <https://doi.org/10.1177/0146167217708574>
- Wilcox, W. B., & Nock, S. L. (2006). What's love got to do with it? Equality, equity, commitment and women's marital quality. *Social Forces*, *84*, 1321–1345. <https://doi.org/10.1353/sof.2006.0076>
- Zhang, H. (2015). Wives' relative income and marital quality in urban China: Gender role attitudes as a moderator. *Journal of Comparative Family Studies*, *46*, 203–220. <https://doi.org/10.3138/jcfs.46.2.203>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.