

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/344750326>

Just a Joke? Can Sexist Comedy Harm Women's Cognitive Performance?

Article in *Psychology of Aesthetics Creativity and the Arts* · October 2020

DOI: 10.1037/aca0000369

CITATIONS

0

READS

460

4 authors:



Silvana Weber

University of Wuerzburg

23 PUBLICATIONS 156 CITATIONS

SEE PROFILE



Markus Appel

University of Wuerzburg

98 PUBLICATIONS 2,212 CITATIONS

SEE PROFILE



Melanie C. Steffens

Universität Koblenz-Landau

181 PUBLICATIONS 3,142 CITATIONS

SEE PROFILE



Vanessa Hirschhäuser

Universität Koblenz-Landau

1 PUBLICATION 0 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



"We promise full employment by 2025" - The effect of election pledges on turnout and vote choice (funded by Fritz Thyssen Stiftung) [View project](#)



2021 Würzburg Summer School "How narratives affect our lives: The psychology of narrative impact (SALON 2021)" [View project](#)

Just a Joke? Can Sexist Comedy Harm Women's Cognitive Performance?

*Silvana Weber*¹, *Markus Appel*¹, *Melanie C. Steffens*², & *Vanessa Hirschhäuser*²

¹ Julius-Maximilians-University Würzburg, Germany

² University of Koblenz-Landau, Germany

Manuscript accepted for publication in the journal

Psychology of Aesthetics, Creativity, and the Arts (10/15/2020).

© 2020, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/aca0000369

Corresponding author:

Silvana Weber, Dr., Institute of Human-Computer-Media, Julius-Maximilians-University Würzburg, Oswald-Külpe-Weg 82, 97074 Würzburg, Germany, Phone: +49-0931-31-88366, Email: silvana.weber@uni-wuerzburg.de

Co-Authors:

Markus Appel, Prof. Dr., Institute of Human-Computer-Media, Julius-Maximilians-University Würzburg, Germany, Email: markus.appel@uni-wuerzburg.de

Melanie C. Steffens, Prof. Dr., Psychological Institute, University of Koblenz-Landau, Germany, Email: steffens@uni-landau.de

Vanessa Hirschhäuser, M.Sc., Psychological Institute, University of Koblenz-Landau, Germany, Email: hirschhaeuser@uni-landau.de

Acknowledgements:

The authors would like to thank Markus Volkmar and our teams of student research assistants at both universities for their support in conducting the studies.

Abstract

The potential of sexist jokes in comedy to harm women has been a matter of intense public debate. Psychological research on sexist comedy is scarce and inconclusive. Theory on social identity threat suggests that communicating devaluation and negative group stereotypes impairs the performance of members of a targeted group: Do women exposed to sexism in stand-up comedy score worse in subsequent cognitive tasks compared to women's performance after watching non-sexist comedy? In four experiments, we examined women's performance on numerical and figural intelligence subtests after watching sexist comedy as compared to non-sexist comedy. In Experiment 1 ($n = 102$) and Experiment 4 ($n = 81$), the test performance of women who watched sexist stand-up comedy clips suffered. Experiments 2 ($n = 181$) and 3 ($n = 100$) showed mixed evidence (see online supplement). A mini meta-analysis reveals a small but significant negative overall effect ($d = -.27$). Self-reported perceived humor (state) and coping sense of humor (trait) did not consistently moderate the influence of sexist comedy. Insights gained from analyses of emotional responses, assessed both via self-reports and a facial coding software, were limited. We discuss implications for identity-threat theory and for using disparaging humor in comedy.

Keywords: Social Identity Threat; Humor; Sexism; Gender; Test Performance

Just a Joke? Can Sexist Comedy Harm Women's Cognitive Performance?

Comedy is a popular genre of the entertaining arts, its origin dating back to ancient Greek playwrights (e.g., Aristophanes, Menander, e.g., Konstan, 1995). Humorous communication often concerns social phenomena, which are commented on in a playful, yet informative way. As an aesthetic form of conveying a message and a creative act in itself, it elicits feelings of appreciation and positive emotions (Weisfeld, 2006). However, as early as in ancient Greek society, Plato and Aristotle suggested that people laugh at the misfortune of others (*Superiority Theory*; Morreall, 1987). From early on, comedy has been criticized for its potentially negative impact on individuals and societies (Plato, 375 BC / 1974). Today, sexist humor can be found in songs, TV shows, and movies. There have been intense discussions about comedians who have been condemned for playing around with stereotypes while trying to entertain on college campuses and in mainstream media (e.g., Flanagan, 2015; Marchese, 2019). One theoretical framework that could back a critical stance towards sexist or otherwise disparaging humor is *Social Identity Threat* (e.g., Lewis Jr. & Sekaquaptewa, 2016; Spencer et al., 2016; Steele, 1997), suggesting that devaluing a group or making negative group stereotypes salient can impair the ambitions and cognitive performance of members of the negatively stereotyped group. Given that today stereotypes and disparagement of women are often communicated in humorous formats (e.g., jokes, sitcoms, stand-up comedy), this form of communicating stereotypes could indeed be a source of social identity threat – despite the humorous notion – with negative consequences for women's cognitive performance and identification with male-dominated domains (Appel & Weber, 2017).

The psychological impact of stereotypes in sexist comedy has received little attention by empirical research under the social identity threat framework or otherwise. The aim of the present work was to examine the influence of real-world sexist comedy on women's cognitive performance. Four experiments were conducted (and meta-analyzed) to test the hypothesized

threat effect (two of them reported in the text and two in the online supplement). In addition, individual differences in the processing of humor were taken into account as potential moderators. Further, emotional responses were examined as potential moderators and mediators, assessed both via self-report measures and via coded facial expressions, which are reported in detail in the online supplement

(https://osf.io/5rtz7/?view_only=18ad9c5dae6e4253908b025a40656f62).

Social Identity Threat and Sexist Comedy

Sexist humor objectifies, stereotypes, victimizes, and degrades an individual due to his or her gender (LaFrance & Woodzicka, 1998). It commonly refers to jokes that are meant to be entertaining and amusing, but portray one sex, usually women, as inferior and lower in performance (Woodzicka & Ford, 2010). Theory and research on social identity threat suggest that when a group is negatively stereotyped or devalued, members of that very group (e.g., women, African Americans, or immigrants) experience stress and increase monitoring of themselves and others (Spencer et al., 2016; Steele, 1997). This reduces working memory capacity, which is then unavailable for any cognitive task. These processes can impair the building of abilities during the learning process, lead to task disengagement, and eventually lower performance in domains that the stereotypes apply to (Appel & Kronberger, 2012). In the long term, social identity threat can lead to disidentification from stereotyped domains and a skewed self-concept (Lewis Jr. & Sekaquaptewa, 2016; Steele et al., 2002) as well as negative consequences for personal relationships in the related domain (Martiny & Nikitin, 2019). Explicit sexist behavior (Logel et al., 2009) as well as subtle sexist cues (Adams et al., 2005) can result in social identity threat for women. The concept of social identity threat has large overlaps with the concept of stereotype threat, but is more encompassing: social identity threat includes not only situations in which negative stereotypes against a person's group are

communicated (i.e., stereotype threat), but also situations in which a person's group is otherwise despised or generally devalued (Appel & Weber, 2017).

In general, media displays of gender stereotypes such as women having low skills and being inferior to men are still widespread (e.g., Aubrey & Frisby, 2011; Eisend, 2010; Gerding & Signorielli, 2014; Kay & Furnham, 2013; Lauzen et al., 2008; Sink & Mastro, 2017). Previous studies have shown that media content, for instance, TV commercials, political ads, newspaper articles, and clips from TV shows can elicit social identity threat or stereotype threat among disparaged group members (e.g., Saleem et al., 2019; Schmader et al., 2015; Schmuck et al., 2017). A meta-analysis with 33 independent studies revealed a mean effect size of $d = -0.38$ in support of this assumption (Appel & Weber, 2017). For example, watching TV commercials in which women were shown in gender stereotypical roles (e.g., a female college student dreaming of becoming the homecoming queen) led to a decrease in math performance and women's leadership aspirations as compared to a neutral commercial (Davies et al., 2002; Davies et al., 2005). The exposure to recurring signals of non-belonging and to negative stereotypes over an extended period of time can result in chronic threat (Cook et al., 2012). The frequent stereotypic display of certain groups, such as communicated by sexist jokes in comedy, may contribute to a climate in which women suffer chronic experiences of the threat that is constantly lingering "in the air" (cf. Steele, 1997).

With the replication crisis in psychology, social identity threat and stereotype threat theory have been scrutinized (e.g., Flore et al., 2018; Flore & Wicherts, 2015; Pennington et al., 2019; Stoet & Geary, 2012). This has attracted attention to the circumstances under which the threat experience and related downstream effects can or cannot be observed. Our focus here is on comedy, given that much of the explicitly sexist content in the media can be found in comedy (Gray & Ford, 2013). In stark contrast to the intense debate at US campuses and the potential real-world consequences, little research has focused on the effects of

disparagement humor (i.e., humor that depreciates or maligns an individual or social group, Zillmann, 1983) on members of the disparaged group. Remarkably, (media) content that has been used in prior social identity threat studies has barely been humorous, except for one study involving a sexist cartoon (Oswald & Harvey, 2000-2001). The effects of cartoon exposure were non-significant overall, but in the absence of the instruction that “men and women perform equally” on the math test, the cartoon *increased* women’s math performance. Based on previous social identity threat theory and research, we would, on the contrary, expect sexist comedy to elicit threat, and thus, lead to *decreased* performance. However, the humorous format of comedy could also mitigate social identity threat, as humor has been theorized to effectively alleviate stress responses in performance situations (Ford et al., 2004; see also Freud, 1905/1960).

Depending on the social context, the acceptance and interpretation of sexist humor can vary: sexist jokes are perceived as less acceptable and more offensive in a work context than at a comedy club (Gray & Ford, 2013). (Sexist) humor can be used as a communication tool to demonstrate power, and can be a means, for instance, to preserve patriarchal structures in organizations (Kahn, 1989; see also Brunner & Costello, 2002). In turn, for women, sexist humor creates a hostile environment at work (Fitzgerald et al., 1995). The exposure to sexist humor increases men’s level of rape proclivity, sexism, and femininity ideology (e.g., Thomae & Viki, 2013; Wright et al., 2017). It also enhances men’s willingness to discriminate against women (Ford & Ferguson, 2004), and reduces their willingness to donate to a women’s organization (Ford et al., 2008).

Taken together, the literature on social identity threat suggests that disparagement humor against women – no matter if it is perceived to be funny or not – may potentially lead to threat effects (see Thomae & Pina, 2015), similar to, for example, portraying women in gender stereotypical roles (Davies et al., 2002). However, given the contested nature of social

identity threat effects, with meta-analyses confirming them on the one hand (e.g., Nguyen & Ryan, 2008) and failures to replicate them on the other (Flore et al., 2018), more research on sexism in humorous formats is needed.

Humor and Cognitive Performance

In contrast to the predictions based on the social identity threat framework, theory and research on humor suggest that the detrimental consequences of negative stereotypes could be absent in a humorous context. Unlike other mechanisms to cope with a threatening situation, humor may effectively alleviate stress responses (cf. *Relief Theory*; see Freud, 1905/1960), including the stress imposed by encountering negative stereotypes. Freud argued that humor as an adaptive mechanism defends the invulnerability of the ego. Hereby, the threat is cognitively still acknowledged, yet affective responses differ, as humor increases positive affect and the tendency to attribute failure externally (see Geisler & Weber, 2010). Further, research has shown that being exposed to (non-sexist) humorous stimuli before taking a math test can increase performance by decreasing state anxiety in the performance situation (Ford et al., 2012).

Definitions of humor and what makes things funny vary, both among lay people and scientists. Taking a cognitive approach to humor, *Incongruity Theory* (e.g., Hull et al., 2017; Suls, 1972) suggests that the cause of laughter is the perception (and resolution) of an inconsistency between what people expect to happen (e.g., based on their mental patterns) and what actually happens. A revision of Incongruity Theory, *Benign Violation Theory* (McGraw & Warren, 2010; Warren & McGraw, 2016), proposes that humor is elicited when a situation is simultaneously appraised as a *violation* (e.g., a threat to how one believes things should be) and as *benign* (i.e., it is alright, safe, or acceptable). Such violations include identity threats (e.g., insults, McGraw & Warren, 2014). Other research confirms that humor can be elicited by displays of hostility and disparagement (e.g., McCauley et al., 1983; Zillmann, 1983).

Usually, violations are not a reason to laugh; however, if they are perceived to happen in a safe context, for instance, a playful conversation or a stand-up comedy show, which includes cues suggesting that the situation should not be taken seriously, they may produce humor. In that case, violations are not perceived as a threat any more, but as funny (McGraw & Warren, 2010). Applied to sexist comedy, we suggest that an explicit communication of gender stereotypes or of the inferiority of women is a violation. Simultaneously, the stereotypes are communicated in a format that entails cues such as hyperboles, which suggest that what is said is not to be taken seriously, and therefore, acceptable (i.e., benign). As a result, the negative consequences of being exposed to gender stereotypes could be suspended (for related evidence, see Mallett et al., 2016; Woodzicka et al., 2015). However, individuals who do not approve the benign circumstances will not judge the jokes to be funny, while the sexist content remains. Thus, individual differences in responding to sexist humor could moderate its social identity threat effects.

There are some indicators that individual differences in humor may influence the ability to cope with stereotyping content. Individuals who score high on *coping sense of humor*, defined as the tendency to respond with cheerfulness in stressful situations (Martin & Lefcourt, 1983; see also Martin, 1996), responded less anxiously in stressful situations. They appraise exams rather as a positive challenge and engage in more functional attribution styles of success and failure (Kuiper et al., 1993; see also Woodzicka & Ford, 2010). In addition, humor has been shown to be one strategy to cope with social identity threat, specifically (Ford et al., 2004). In that study, participants were informed that an upcoming test was diagnostic of mathematical ability and that it had shown gender differences in the past (threat condition), or participants received no such information (control condition). Although no humorous stimulus was presented, coping sense of humor moderated the effect: women who scored high in coping sense of humor were less affected by the performance-inhibiting effects of social

identity threat. They showed less anxiety and better performance in the potentially threatening situation.

The Current Research

With opposing theoretical assumptions based on social identity threat theory on one side and humor research on the other side, it is an open question whether media content that entails gender stereotypes but is communicated in a humorous format leads to negative consequences for members of the stigmatized group, in this case women. On the one hand, based on theory and research on social identity threat, we would expect to find decreased performance and lower domain identification in a sexist comedy as compared to a non-sexist comedy condition. On the other hand, the language used in humor and non-verbal cues give the threatening information conveyed an ironic notion that can make the situation benign. Consequently, comedy that entails gender stereotypes may not elicit social identity threat, if the benign circumstances are recognized; thus, a humorous response may alleviate the threat response. Therefore, it is possible that threat effects are not elicited when stereotypes are communicated in a humorous way.

The present research aims to examine potential social identity threat effects elicited by real-world stand-up comedy. To this aim, female participants were randomly assigned to watch either a comedy clip containing sexist humor or a control comedy clip without sexist jokes. Subsequently their cognitive performance in stereotypically male-dominated domains was assessed. The studies further examined individual differences in humor as predictors of cognitive performance in situations of social identity threat.

Hypotheses.

Guided by social identity threat theory, we assumed that women who are confronted with comedy including sexist jokes would perform worse in a cognitive performance test than women who are confronted with non-sexist comedy (Hypothesis 1 – main effect). We

expected that women who experience the sexist comedy as more humorous (state) show better performance than women who experience it as non-humorous (Hypothesis 2 – moderation effect 1). Further, based on Ford et al. (2004), coping sense of humor (trait) was expected to moderate the threat effect, as women who frequently use humor as a coping style were expected to be less affected by social identity threat (Hypothesis 3 – moderation effect 2).

Study Overview.

We conducted four lab experiments to examine our hypotheses. In all four experiments, we examined women's cognitive performance after watching a sexist comedy clip versus a control clip containing non-sexist comedy, but with varying designs and stimulus material. Experiments 1 and 4 are reported in the main paper, while Experiments 2 and 3 (though inconclusive) are reported in the online supplement to adhere to the complete reporting policy. Experiments 1-3 included perceived humor (state) and coping sense of humor (trait) as potential moderators. Experiments 2-4 also assessed additional moderating and mediating variables in order to explore the boundary conditions of the effect. Finally, we conducted a meta-analysis of all four experiments to estimate the overall magnitude of the obtained effect. In the paper and the online supplement, we follow a full transparency approach, and report all experiments, conditions, and variables examined. All four experiments were conducted in Germany, adhering to local ethical guidelines and data protection policies. Experiments 2 and 4 were pre-registered. Preregistration documents, the online supplement, and the material of all Experiments can be found in the Open Science Framework (https://osf.io/5rtz7/?view_only=18ad9c5dae6e4253908b025a40656f62).

Experiment 1

Stimuli

To induce social identity threat, a five-minute comedy clip of the comedian Mario Barth was used. Mario Barth is one of the best-selling comedians in Germany who regularly

includes disparagement humor against women in his stand-up shows, from which we showed one sequence. By making fun of women's shopping behavior and displaying it as their main interest in life, he implies that women are superficial and stupid. A comedy clip of similar length with the comedian Luke Mockridge that did not entail disparagement humor served as the control stimulus. As usual in comedy formats, there was laughter in the background. Stimulus material was pretested with a sample of female students ($n = 17$, repeated measurement) to ensure comparable levels of funniness and familiarity, but differences in perceived sexist humor (one item each, rated on an eight-point scale). All clips were freely available on *YouTube*.¹

Measures

Perceived Humor. Humor as a state was assessed with an ad hoc *Perceived Humor Scale* (following Swani et al., 2013). Participants evaluated whether they found the clip humorous with six items that were answered on a four-point scale (e.g., "I found the video funny"; 1 = "strongly disagree" to 4 = "strongly agree"). After excluding one item (i.e., "I believe the video was making a humor attempt"), internal consistency was excellent, $\alpha = .90$.

Coping Sense of Humor. Humor as a trait was measured with the *Coping Humor Scale* (Martin & Lefcourt, 1983). The scale assesses the tendency to use humor to cope with stressful situations with seven items (e.g., "I can usually find something to laugh or joke about even in trying situations") that are answered on a four-point scale (1 = "strongly disagree" to 4 = "strongly agree"; $\alpha = .67$).

Cognitive Performance. We assessed cognitive performance before (baseline) and after (post-test) watching the comedy. Numerical and figural subtests of the intelligence test *I-S-T 2000 R* (Liepmann et al., 2007) are designed as power tests and start with relatively low

¹ Detailed information on the stimulus material is provided in the online supplement. All video clips used in this research can be requested from the first author (i.e., videos, transcripts, and English translations).

difficulty, which then increases. We used ten items each (i.e., every other item) of Form A (baseline) and B (post-test) of the subtests that include arithmetical problems and figure selection tasks, as math and visual thinking are stereotypically considered male-dominated domains. As Form A and B were conceptualized as pseudo-parallel forms, learning effects cannot be ruled out after repeated testing. The test manual proposes ten minutes of working time per subtest for both arithmetical problems and figure selection tasks (for 20 items each in the original test); thus, the software was programmed to jump to the next page after exactly ten minutes. Participants were not required to answer all questions. In order to prevent people who were less motivated or disinterested from moving to the next page before even trying to solve the tasks, the “next”-button was hidden. The performance score (i.e., both subtests taken together) was corrected for guessing by subtracting the number of incorrect answers from the number of correct answers, for both points of measurement. Thus, negative scores were possible.

Participants and Procedure

An a priori sample size calculation (G*Power; Faul et al., 2007) for both main effects (between: comedy sexism; within: time of performance measurement) and the interaction effect was conducted. Based on the assumption of medium-sized effects ($f = 0.25$ with $\alpha = .05$ and $1 - \beta = .80$) and an estimated correlation of repeated measures of $r = .60$, a sample size of 104 participants was required for the between subjects effect, 34 participants were required for the within subjects effect, and 28 participants were required for detecting the interaction. We aspired the largest of the three required sample sizes.

The study was conducted in a pre-post control study design in a laboratory with a male research assistant. Interaction with the assistant was minimal, as the study was executed on a computer using a survey software. The cover story explained that the current study examines the association between media use and the ability to concentrate. Out of $n = 109$, seven

records had to be excluded because the participants could not complete the study due to technical difficulties and external disturbances. The final sample consisted of $n = 102$ female undergraduates (age: $M = 21.70$, $SD = 2.33$, range 18-27 years), who received course credit for their participation.

First, participants completed the Coping Humor Scale. Then, a first subset of items of the numerical and figural subtests of the intelligence test I-S-T 2000 R was administered (baseline). Next, participants were randomly assigned to watch one of the two videos (sexist vs. non-sexist) and rated how humorous they perceived the comedy. Afterwards, participants worked on a second subset of items of the numerical and figural subtests of the intelligence test I-S-T 2000 R (post-test). Finally, demographic data were collected and participants were asked about the suspected purpose of the study. Participants were thoroughly debriefed. In total, the study took about 40 minutes.

Results

Main Effect of the Treatment on Cognitive Performance

The results suggest that sexist comedy impaired test performance (see Figure 1: sexist condition: $n = 52$, $M_{baseline} = 5.06$, $SD_{baseline} = 5.63$, $M_{post} = 2.15$, $SD_{post} = 6.58$; control condition: $n = 50$, $M_{baseline} = 5.94$, $SD_{baseline} = 5.92$; $M_{post} = 6.00$, $SD_{post} = 6.03$). A 2 (point of measurement, repeated) \times 2 (treatment) ANOVA yielded a main effect for point of measurement, $F(1, 100) = 4.44$, $p = .038$, $d = .42$, and a significant interaction between point of measurement and video condition, $F(1, 100) = 4.83$, $p = .030$, $d = .44$.² Participants in the sexist comedy condition showed a decrease in performance, $p = .003$, 95% CI [1.03; 4.78], whereas participants in the neutral comedy condition did not, $p = .950$, 95% CI [-1.97; 1.85].

² This result includes seven participants who were suspicious of the experiment's goal and remarked a potential link between the sexist comedy and performance, as indicated in the debriefing question. Comparable results were obtained when these participants were excluded: main effect, $F(1, 93) = 6.16$, $p = .015$, $d = .51$; interaction, $F(1, 93) = 6.59$, $p = .012$, $d = .53$.

Preliminary Analyses of Humor Variables

Contradicting the pretest of the stimulus material, the neutral video was perceived to be funnier than the sexist video, $t_{Welch}(83.04) = 4.24, p < .001, d = .85$. Remarkably, there was more variance in the sexist condition ($M = 2.78, SD = 0.88$) than in the control condition ($M = 3.40, SD = 0.56$), Levene's test: $F = 15.11, p < .001$. Coping sense of humor did not differ between conditions, $t(100) = 0.89, p = .375, d = .18$. Both humor variables, state and trait, were unrelated for the total sample, $r = .05, p = .64$, and in both the control ($r = .02, p = .90$) and the sexist comedy condition ($r = .00, p = .99$).

Moderation of Perceived Humor (State)

All moderation analyses were conducted with the PROCESS-macro by Hayes (2013), using Model 1 (10'000 bootstraps). Treatment (dummy coded) and perceived humor (z-standardized) as well as the interaction between both variables served as predictors for difference scores of the cognitive performance measure. The interaction was non-significant, $B = 2.21, SE_B = 1.34, p = .10, 95\% \text{ CI } [-0.45; 4.87]$, but we still explored the data pattern. In partial support of the hypothesis, more perceived humor predicted better performance in the sexist video condition (simple slope: $B = 2.81, SE = 0.95, p = .004, 95\% \text{ CI } [0.93; 4.69]$), but not in the neutral video condition (simple slope: $B = 0.61, SE = 0.95, p = .52, 95\% \text{ CI } [-1.27; 2.49]$).

Moderation of Coping Sense of Humor (Trait)

A comparable analysis as above indicated that coping sense of humor (z-standardized) did not moderate the effect of the comedy clip, non-significant interaction: $B = 0.68, SE_B = 1.39, p = .62, 95\% \text{ CI } [-2.08; 3.44]$.

Discussion

This study examined the social identity threat effect among women if confronted with sexist comedy (vs. non-sexist comedy). Humor scores (state and trait) were included as

potential moderators, in order to explore the boundary conditions of the effect. Connecting social identity threat with humor research, we used humorous real-world stimulus material that contained either sexist or neutral jokes. Change scores from baseline suggest that women performed worse on a set of cognitive tests after exposure to comedy with sexist jokes than women exposed to comedy without sexist jokes. Watching sexist comedy led to a drop in performance whereas the performance remained unchanged in the control condition, with a small to medium effect for the interaction (Cohen, 1988). In contrast to findings of previous research (Ford et al., 2004), coping sense of humor and perceived humor did not moderate the effect. Even though the interaction was not significant, the simple slopes suggest that the funnier the sexist video was perceived, the smaller was the drop in performance from pre to post assessment. This indicates that perceiving sexist jokes as funny might alleviate the threat that they pose on women's gender identity. This finding is in line with Benign Violation Theory (McGraw & Warren, 2010), suggesting that something threatening (i.e., a violation) that is simultaneously judged as safe (i.e., benign), can be perceived as funny, and in turn, mitigate the threat response. Replicating the finding with other stimuli was deemed to be important.

Experiments 2-3

A new clip by a third comedian (called Hape Kerkeling) was added for the control conditions in Experiments 2 and 3 in order to arrive at similarly funny clips. Thus, Experiment 2 included two control conditions: the control video that had been used in Experiment 1 (Luke Mockridge) was, again, considered funnier than the clip in the experimental condition, and the new control video (Hape Kerkeling) was rated comparably funny as the experimental video. However, in Experiment 3, the two used clips (experimental video: Mario Barth; control video: Hape Kerkeling) differed regarding perceived funniness. Furthermore, in hindsight, we realized that the new control clip implemented in Experiments

2 and 3 could have subtly introduced gender-related social identity threat as well, as it may have been perceived to contain subtle disparagement humor (see online supplement for details on the stimulus material). In line with this thought, performance in that control condition resembled performance in the experimental condition. Due to this ambiguity that makes the findings inconclusive, we report those experiments in the online supplement, adhering to a full-disclosure policy. The results of the two experiments were also included in our mini meta-analysis reported below.

Experiment 4

The main aim of Experiment 4 was extending our findings to different sexist and non-sexist comedy clips. To obtain closer experimental control, two clips by the same comedian were selected after pretesting. Comedy by the comedian Luke Mockridge was shown in order to examine whether sexist jokes by a comedian rated popular in the target group can induce social identity threat. Additionally, we aimed at testing whether findings supported the hypothesis that emotional reactions, measured via self-reports, mediated the threat effect. No indirect effects were found (see online supplement for details). Moreover, to further examine the boundary conditions of the threat effect, the roles of two potential moderators were explored, that is, gender identification (Schmader, 2002) and endorsement of benevolent sexism, confining women to traditional social roles (i.e., agreement with statements such as “Women should be cherished and protected by men”, Glick & Fiske, 1996). Findings were inconclusive, with no evidence obtained supporting the hypotheses, and are reported in the supplement. This study was preregistered at aspredicted.org:

<http://aspredicted.org/blind.php?x=np449j>.

Stimuli

Two different video clips by the same comedian (Luke Mockridge) were used for the sexist and the control condition. The clips were pretested, ensuring comparable levels of

perceived funniness and differences in perceived sexism. In the sexist condition, women are objectified and devalued by focusing on their visual appearance. Further, a description of women's behavior during a date implies them being naive and superficial.

Measures

Cognitive Performance. A different numerical subtest (i.e., completing numerical series) of the I-S-T 2000 R was used (Liepmann et al., 2007). The 20 items of Form A were split up by every other item; thus, participants were asked to complete 10 items before and 10 items after watching the comedy clip. Due to repeated testing, learning effects are possible. Participants had 5 minutes per set to complete the task. Additionally, they were provided with a piece of paper to take notes.

Emotional Reaction, Gender Identification, and Benevolent Sexism. Details on these measures can be found in the online supplement.

Participants and Procedure

The same a-priori sample size calculation (G*Power; Faul et al., 2007) as in Experiment 1 applies. In total, $n = 102$ women participated, of which 21 were excluded (15 from the sexist and 6 from the control condition), because they reported that they knew the shown video clip, which in this case might have severely influenced the effect.³ Therefore, data of $n = 81$ women (age: $M = 23.42$, $SD = 3.43$, range 18-36 years) were analyzed (sexist condition: $n = 36$; control condition: $n = 45$). A sensitivity analysis (G*Power; Faul et al., 2007) shows that the effect size that could be detected regarding the focal between by within interaction (given $n = 81$ participants, $\alpha = .05$ and $1-\beta = .80$, $r = .60$ for the correlation between the repeated measures) amounts to $f = 0.14$. Thus, we were able to identify an interaction effect of small-to-medium size.

³ The full original clip ended with a very positive portrayal of women, which we had cut for the purpose of this study. All patterns of findings are comparable in the analyses including the total sample: main effect, $F(1, 100) = 23.98$, $p < .001$, $d = 0.98$; interaction, $F(1, 100) = 6.38$, $p = .013$, $d = 0.51$.

The experiment was conducted in a pre-post control study design in a laboratory with a female research assistant. The same cover story was used as in Experiment 1. Participants received either course credit or a small monetary compensation. After completing the first subset of the cognitive performance test, participants were randomly assigned to watch either the sexist or the neutral comedy clip. Afterwards, the second subset of the performance test was completed, followed by the evaluation of the emotional reaction and the assessment of gender identification and benevolent sexism. Additionally, demographic data (i.e., age, gender, sexual orientation, occupation), as well as familiarity with and attitude towards the comedian and the shown clip were assessed (binary answer format). Among the participants, 78% stated they generally liked the comedian. Finally, participants were thanked and debriefed. The study took about 25 minutes.

Results

Manipulation Checks

One-item manipulation checks showed that the participants did not consider the videos significantly different in funniness (5-point scale from 1 = “not funny at all” to 5 = “very funny”; sexist condition: $M = 3.28$, $SD = 1.19$; control condition: $M = 3.69$, $SD = 1.13$), $t(79) = 1.60$, $p = .12$, $d = 0.35$. Yet, the sexist video was rated more sexist (5-point scale from 1 = “not sexist at all” to 5 = “very sexist”, $M = 3.19$, $SD = 1.24$) than the control video ($M = 1.44$, $SD = 0.66$), $t_{Welch}(50.60) = -7.66$, $p < .001$, $d = 1.76$. Perceived funniness and perceived sexism tended to be negatively correlated, $n = 81$, $r = -.22$, $p = .052$.

Main Effect of the Treatment on Cognitive Performance

The results indicate an effect of the experimental condition on post-comedy performance (sexist condition: $M_{baseline} = 5.50$, $SD_{baseline} = 2.97$, $M_{post} = 5.53$, $SD_{post} = 2.95$; control condition: $M_{baseline} = 5.42$, $SD_{baseline} = 2.44$; $M_{post} = 6.71$, $SD_{post} = 2.94$). A 2 (point of measurement, repeated) \times 2 (treatment) ANOVA yielded a main effect of point of

measurement, $F(1, 79) = 9.92, p = .002, d = .71$, and the expected significant interaction between point of measurement and video condition, $F(1, 79) = 9.10, p = .003, d = .68$. Data suggest a negative effect of exposure to comedy with sexist jokes on performance (see Figure 1): participants in the control condition showed a pre-post increase in completing the numerical series test, $p < .001, 95\% \text{ CI } [0.73; 1.84]$, whereas participants in the sexist comedy condition did not, $p = .93, 95\% \text{ CI } [-0.59; 0.65]$. Put differently, women's cognitive performance was more negative after watching sexist comedy than the control clip, $F(2, 78) = 9.21, p = .003, d = .69$ (controlling for pre-exposure performance).

Discussion

The results of Experiment 4 confirm our main hypothesis. Women in a control comedy condition outperformed women, who were exposed to sexist comedy. The cognitive test we used differed from the one used in Experiment 1; we found higher post- vs. pre-test scores in the control condition (i.e., a learning effect), whereas the scores in the sexist comedy condition remained nearly unchanged. As stated above, due to the pseudo-parallel forms such learning effects can occur. We argue that these effects were inhibited in the threat condition.

From a methodological point of view, using two video clips of the same comedian for both the sexist and the control condition eliminated some interpretation problems.

Additionally, as research has found, inducing social identity threat through subtle cues can be more detrimental for women's performance than explicit cues (Nguyen & Ryan, 2008).

Opting to use a comedy clip from Luke Mockridge for the sexist condition, who is generally known to be less offensive in comparison to Mario Barth, whose sexist jokes are very blatant, may have resulted in more intense threat effects (see Discussion). The findings also demonstrate that participants' overall liking of a comedian who uses disparagement humor (amongst other sorts of jokes) does not eliminate the sexist comedy's damaging effects.

Mini Meta-Analysis

Our central goal was to examine the impact of sexist comedy (vs. non-sexist comedy) on the cognitive performance of women. As the results of the four experiments differed in that respect (for Experiments 2 and 3 see online supplement), we meta-analyzed the data to get a clearer picture (see Goh et al., 2016). All four experiments entailed one group that was exposed to a sexist comedy clip and one or two groups that were exposed to a non-sexist comedy clip. Cognitive performance served as a dependent variable across experiments (but operationalizations varied as indicated).

The mini meta-analysis followed the recommendations by Goh and colleagues (2016). We used the software *Comprehensive Meta-Analysis* (Borenstein et al., 2005) to conduct the analysis. Cohen's d was chosen as the effect size metric and we used the fixed effects model. For Experiments 1, 2, and 4, data on pre- and post-exposure performance were entered, for Experiment 3 only post-exposure performance scores were available. The results of both control conditions in Experiment 2 were pooled. Across all experiments, the difference between the sexist and the non-sexist condition amounted to $d = -.27$, 95% CI $[-.459; -.081]$, $p = .005$. Taking together the results from our four lab experiments ($n = 464$), this finding indicates that cognitive performance was impaired by being exposed to sexist (vs. non-sexist) comedy.

General Discussion

In recent years, sexism in comedy has become a contested issue. One of the arguably most relevant questions regarding sexist comedy is the question of its psychological effects. What are the consequences of being exposed to sexist comedy? Prior research has shown that sexist jokes may increase men's hostility towards women (Ford et al., 2015). Yet, previous research has put little emphasis on women and their behavior after being exposed to sexist jokes. The current series of studies addressed this research lacuna from a social identity threat

perspective. This work advances theory and research on social identity threat by combining it with humor research (i.e., Benign Violation Theory), and thus, extending the phenomenon to a new and relevant applied context. Our investigation contributes to prior theory and research on sexist humor and to media-elicited threat effects. The studies profit from an experimental design, allowing us to draw causal conclusions. We conducted four experiments to test the assumption that sexist comedy undermines women's cognitive performance – and we examined the role of humor. Social identity threat effects were obtained in Experiment 1 and Experiment 4, while Experiments 2 and 3 yielded null results (see online supplement). Because results varied across the four experiments, a mini meta-analysis was conducted. This meta-analysis suggests that – overall – watching sexist comedy impaired women's cognitive performance on stereotypically male-associated tests.

Social identity threat has become a controversial topic in recent years (e.g., Flore et al., 2018; Flore & Wicherts, 2015; Stoet & Geary, 2012). A major next step in this research program is to test the theory in fields of theoretical and applied relevance. In our experiments, we found support for our assumption of social identity threat effects due to sexist comedy exposure. However, the results were heterogeneous across studies, and the average effect sizes were small. The null findings in Experiments 2-3 could either be due to an unlucky choice of control conditions, or alternatively, another demonstration that social identity threat effects are hard to replicate. We therefore perceive our findings as a starting point for future research, rather than an end point. The potentially detrimental effects of sexism in comedy and other entertaining arts has received far less attention in experimental research than warranted, given the heated public debate.

The Ambivalent Role of Humor

Humor may play an ambivalent role in the context of sexism. While it is an aesthetic form of communication on the one hand, it may, on the other hand, also be perceived to

downplay the severity of the content. The presented comedy clips included unexpected and abrupt shifts in perspective, which elicit laughter (cf. Incongruity Theory). However, not everyone may acknowledge the humor in stereotypic presentations of women, as displayed by Mario Barth (Experiments 1-3) or Luke Mockridge (Experiment 4). Therefore, based on Benign Violation Theory, we focused on humor as a way to cope with potentially threat eliciting, sexist, yet humorous media content. The tongue-in-cheek communication of comedy imparts that the sender of the message distances him or herself from the content. Recipients may consider that the protagonists or media professionals involved do not actually believe in the stereotype, which makes the situation benign, and thus, may not provoke a threat response. Yet, if the benignancy of the situation is not recognized by the recipient, the stereotypic display of women may bear a threat to in-group members (i.e., female recipients). Thus, the individual difference variables coping sense of humor (trait) and perceived humor (state) were included as potential moderators.

We could not replicate the finding by Ford and colleagues (2004), as coping sense of humor did not moderate the threat effects. Notably, the reliability of the scale was relatively low across studies. Finding a better way to assess humor as a coping style should be an objective of future research. In the given context, other humor styles, such as self-enhancing or self-defeating humor (as assessed in the *Humor Styles Questionnaire*, Martin et al., 2002) might give broader insight into the role of trait humor in potentially threatening performance situations. Additionally, further explorations of individual differences in how individuals perceive and respond to sexist comedy may provide deeper insight into its potential detrimental effects. It stands to reason that people who self-select to watch humor that disparages their group are less likely to experience threat effects, as they acknowledge the benign situation. However, in the current research, participants did not self-select whether they wanted to view sexist or non-sexist comedy, because they were randomly assigned to the

experimental conditions. Benevolent sexism was considered as a potential moderating factor in Experiment 4. The results align with previous research which showed that people who are high in sexism report a higher appreciation of sexist humor (e.g., Eyssel & Bohner, 2007; Greenwood & Isbell, 2002), such as women with high levels of benevolent sexism did not experience a threat effect. This exploratory finding should be taken into account in future research.

In line with findings of previous research, perceived humor (state) predicted cognitive performance in Experiment 1, suggesting that the funnier women perceived any of the presented comedy clips (independent of experimental condition), the better they performed in a subsequent test. This result was replicated in Experiment 3, but not in Experiment 2. However, the results of the moderation analyses of perceived humor were inconclusive. Even though the interactions were not significant in Experiment 1 (and in Experiment 3), the simple slopes revealed that in the sexist comedy condition in Experiment 1 (and a similar tendency in Experiment 3), the drop in performance was reduced if the comedy was perceived as funny. This pattern of results could not be replicated in Experiment 2. Despite using advanced measurements to assess emotions (i.e., a software-based analysis of facial expressions, see online supplement), Experiment 3 could not shed light on the complex relationship between perceived humor and performance. Including other emotions into the model also did not contribute to a clearer picture, neither as moderators (Experiment 3) nor as mediators (Experiment 4).

Limitations and Future Research Directions

A first limitation pertains to culture. The current studies were conducted with German speaking, female university students. Replications in other countries, with different samples, and with a larger variety of comedians are encouraged. While men might show social identity threat effects after being exposed to sexist comedy against men, they could also experience

reversed effects after watching sexist comedy against women (i.e., *stereotype lift*; see Appel & Weber, 2017). Examining cognitive consequences of sexist comedy for men could rule out alternative explanations of our results such as priming effects.

As usual in lab experiments, we can only provide evidence for the general existence of the effect in a rather artificial situation (i.e., taking a cognitive performance test after watching comedy). It remains open how much exposure is necessary to undermine women's motivation or domain identification in general, and how long-ranging the effect might be. Similarly, one would need to systematically test which aspects of cognitive performance suffer and which do not.

Another limitation is the use of different comedians for the sexist versus non-sexist comedy conditions. On the one hand, both comedians presented in Experiment 1 won the recognized German Comedy Award and are established comedians in the German media. On the other hand, however, they also differ in several respects, most importantly, regarding the type of comedy that they are known for. Mario Barth, who is widely known since the 2000s, is popular as a comedian who jokes about gender stereotypes, as the titles of his shows already suggest (e.g., "Men are primitive, but happy", 2006). In contrast, Luke Mockridge is a younger comedian, who entered the public comedy stage in the 2010s. In his shows, he talks about growing up in the 90s and uses rather puerile humor. That being said, using clips of different comedians (Experiments 1-3) involved various sources of variance. This problem was resolved in Experiment 4 by using the same comedian for both the sexist and the non-sexist comedy condition.

Despite employing sophisticated methods beyond self-report measures (see online supplement), we could not clarify the role of state humor. If humor really reduced the negative influence of sexism and alleviated the threat that the violation poses, cognitive performance should be better after hearing funny sexist jokes as compared to unfunny sexist

jokes. Instead of measuring individual differences in the humor response, funniness could be experimentally manipulated to examine the causal influence of humor. Further, a baseline condition in which participants watch a non-funny instead of a funny video in between the pre and post measures could clarify matters. On a related note, some research has compared the perception of derogatory humor with that of the same statements in the absence of humor. It was examined whether jokes were rated more offensive and confrontation-worthy than statements with the same content. Sexist statements were perceived as more offensive than sexist jokes (Woodzicka et al., 2015). Similarly, future research on social identity threat should compare humorous versus non-humorous disparagement. Whereas the latter control condition is desirable to test whether humor can mitigate the effect of sexist content, it would be hard to design a satisfactory control condition comparable to the comedy clips we presented. Future research following up on our set of studies could instead use jokes to test whether exposure to them leads to less detrimental effects than the same content conveyed in a non-humorous fashion. This would be important evidence concerning Benign Violation Theory, to assess how humorous disparagement may be different from non-humorous disparagement. However, for the present purposes, it was sufficient to demonstrate that sexism, though presented in comedy, can impair women's cognitive performance.

Some jokes in the real world focus on stereotypes that are closely related with potential dependent variables of cognitive performance (e.g., women are bad at math and can't think logically), whereas other jokes express a general devaluation of women (e.g., sexual objectification). In our studies, we only included the latter, and thus, rather focused on social identity threat, while the former might target more concretely stereotype threat. Based on theory and research, we assume that both types of sexist jokes may impair women's cognitive performance, yet future research is needed to extend the findings by examining different joke content.

Evidence suggests that sexist jokes are better accepted when a member of the depreciated group tells them (Ford, 2000; Thai et al., 2019). Based on Benign Violation Theory, the gender of the comedian (i.e., the communicator of the sexist comedy) could affect whether the situation is considered more or less benign, and thus, influence the level of disparagement and threat elicited by the stimulus. We assume that watching a female comedian telling sexist jokes would not impair performance to the same extent as a male comedian telling sexist jokes. However, this remains speculative. In this context, it should also be considered that men are stereotypically considered funnier than women, even though empirical findings on this matter are controversial (meta-analysis by Greengross et al., 2020; see also Hooper et al., 2016; Mickes et al., 2011).

Conclusion

Overall, four studies including 464 female participants, suggest that watching sexist comedy can impair women's cognitive performance as compared to non-sexist comedy. We showed that this applies to situations in which a cognitive performance test is given right after being confronted with sexist comedy. Going beyond this situational approach, reiterating sexist jokes in the media or on live comedy stages could contribute to an overall sexist climate that could have further detrimental consequences for women (i.e., chronic threat effects, spillover effects). Any public discussion on the boon or bane of using stereotypes in comedy (e.g., Flanagan, 2015; Marchese, 2019) can profit from empirical research examining its influence both on members of the group that is made fun of and others. The present findings suggest that sexist humor can have manifest negative consequences on those made fun of, even if the performance is meant to be just funny and entertaining.

References

- Adams, G., Garcia, D. M., Purdie-Vaughns, V., & Steele, C. M. (2005). The detrimental effects of a suggestion of sexism in an instruction situation. *Journal of Experimental Social Psychology, 42*, 602–615. <https://doi.org/10.1016/j.jesp.2005.10.004>
- Appel, M., & Kronberger, N. (2012). Stereotype threat and the achievement gap: Stereotype threat prior to test taking. *Educational Psychology Review, 24*, 609–635. <https://doi.org/10.1007/s10648-012-9200-4>
- Appel, M., & Weber, S. (2017). Do mass mediated stereotypes harm members of negatively stereotyped groups? A meta-analytical review on media-generated stereotype threat and stereotype lift. *Communication Research. Ahead-of-Print* <https://doi.org/10.1177/0093650217715543>
- Aubrey, J. S., & Frisby, C. M. (2011). Sexual objectification in music videos: A content analysis comparing gender and genre. *Mass Communication & Society, 14*, 475–501. <https://doi.org/10.1080/15205436.2010.513468>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive Meta-Analysis* (Version 2.2.027) [Computer software]. Biostat.
- Brunner, P. W., & Costello, M. L. (2002). Where's the joke? The meaning behind sexual humor. *Advancing Women in Leadership*. <http://web.advancingwomen.com/aw1/spring2002/BRUNN~37.HTM>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum. <https://doi.org/10.1016/B978-0-12-179060-8.50012-8>
- Cook, J. E., Purdie-Vaughns, V., Garcia, J., & Cohen, G. L. (2012). Chronic threat and contingent belonging: Protective benefits of values affirmation on identity development. *Journal of Personality and Social Psychology, 102*, 479–496. <https://doi.org/10.1037/a0026312>

- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, *28*, 1615-1628. <https://doi.org/10.1177/014616702237644>
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, *88*, 276–287. <https://doi.org/10.1037/0022-3514.88.2.276>
- Eisend, M. (2010). A meta-analysis of gender roles in advertising. *Journal of the Academy of Marketing Science*, *38*, 418–440. <https://doi.org/10.1007/s11747-009-0181-x>
- Eyssel, F., & Bohner, G. (2007). The rating of sexist humor under time pressure as an indicator of spontaneous sexist attitudes. *Sex Roles*, *57*(9-10), 651-660. <https://doi.org/10.1007/s11199-007-9302-5>
- 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>
- Fitzgerald, L. F., Swan, S., & Fischer, K. (1995). Why didn't she just report him? The psychological and legal implications of women's response to sexual harassment. *Journal of Social Issues*, *51*, 117–137. <https://doi.org/10.1111/j.1540-4560.1995.tb01312.x>
- Flanagan, C. (2015). That's Not Funny! Today's college students can't seem to take a joke. *The Atlantic*. <http://www.theatlantic.com/magazine/archive/2015/09/thats-not-funny/399335/>

- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*, 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>
- Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: a registered report. *Comprehensive Results in Social Psychology, 3*, 140–174. <https://doi.org/10.1080/23743603.2018.1559647>
- Ford, T. E. (2000). Effects of sexist humor on tolerance of sexist events. *Personality and Social Psychology Bulletin, 26*, 1094–1107. <https://doi.org/10.1177/01461672002611006>
- Ford, T. E., Boxer, C. F., Armstrong, J., & Edel, J. R. (2008). More than “just a joke”: The prejudice-releasing function of sexist humor. *Personality and Social Psychology Bulletin, 34*(2), 159–170. <https://doi.org/10.1177/0146167207310022>
- Ford, T. E., & Ferguson, M. A. (2004). Social consequences of disparagement humor: A prejudiced norm theory. *Personality and Social Psychology Review, 8*, 79–94. https://doi.org/10.1207/S15327957PSPR0801_4
- Ford, T. E., Ferguson, M. A., Brooks, J. L., & Hagadone, K. M. (2004). Coping sense of humor reduces effects of stereotype threat on women’s math performance. *Personality and Social Psychology Bulletin, 30*, 643–653. <https://doi.org/10.1177/0146167203262851>
- Ford, T. E., Ford, B., Boxer, C. M., & Armstrong, J. A. (2012). Effect of humor on state anxiety and math performance. *Humor, 25*, 59–74. <https://doi.org/10.1515/humor-2012-0004>

- Ford, T. E., Richardson, K., & Petit, W. E. (2015). Disparagement humor and prejudice: Contemporary theory and research. *Humor, 28*, 171–186.
<https://doi.org/10.1515/humor-2015-0017>
- Freud, S. (1905/1960). Jokes and their relation to the unconscious. In *The Standard Edition of the complete works of Sigmund Freud* (Vol. 8, pp. 225–233). Hogarth.
- Geisler, F. C. M., & Weber, H. (2010). Harm that does not hurt: Humor in coping with self-threat. *Motivation and Emotion, 34*, 446–456. <https://doi.org/10.1007/s11031-010-9185-6>
- Gerding, A., & Signorielli, N. (2014). Gender roles in tween television programming: A content analysis of two genres. *Sex Roles, 70*, 43–56. <https://doi.org/10.1007/s11199-013-0330-z>
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491–512.
<https://doi.org/10.1037//0022-3514.70.3.491>
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*, 535–549. <https://doi.org/10.1111/spc3.12267>
- Gray, J., & Ford, T. E. (2013). The role of social context in the interpretation of sexist humor. *Humor, 26*, 277–293. <https://doi.org/10.1515/humor-2013-0017>
- Greengross, G., Silvia, P. J., & Nusbaum, E. C. (2020). Sex differences in humor production ability: A meta-analysis. *Journal of Research in Personality, 84*, 103886.
<https://doi.org/10.1016/j.jrp.2019.103886>
- Greenwood, D., & Isbell, L. M. (2002). Ambivalent sexism and the dumb blonde: Men's and women's reactions to sexist jokes. *Psychology of Women Quarterly, 26*(4), 341–350.
<https://doi.org/10.1111/1471-6402.t01-2-00073>

- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.
<https://doi.org/10.1111/jedm.12050>
- Hooper, J., Sharpe, D., & Roberts, S. G. B. (2016). Are men funnier than women, or do we just think they are?. *Translational Issues in Psychological Science*, 2, 54–62.
<https://doi.org/10.1037/tps0000064>
- Kahn, W. (1989). Toward a sense of organizational humor: Implications for organizational diagnosis and change. *The Journal of Applied Behavioral Science*, 25, 45–63.
<https://doi.org/10.1177/0021886389251004>
- Kay, A., & Furnham, A. (2013). Age and sex stereotypes in British television advertisements. *Psychology of Popular Media Culture*, 2, 171–186. <https://doi.org/10.1037/a0033083>
- Konstan, D. (1995). *Greek comedy and ideology*. Oxford University Press.
- Kuiper, N. A., Martin, R. A., & Olinger, L. J. (1993). Coping humour, stress, and cognitive appraisals. *Canadian Journal of Behavioural Science*, 25, 81–96.
<https://doi.org/10.1037/h0078791>
- LaFrance, M., & Woodzicka, J. A. (1998). No laughing matter: Women's verbal and nonverbal reactions to sexist humor. In J. Swim & C. Stagnor (Eds.), *Prejudice: The target's perspective* (pp. 61–80). Academic Press. <https://doi.org/10.1016/B978-012679130-3/50038-7>
- Lauzen, M. M., Dozier, D. M., & Horan, N. (2008). Constructing gender stereotypes through social roles in prime-time television. *Journal of Broadcasting & Electronic Media*, 52, 200–214. <https://doi.org/10.1080/08838150801991971>
- Lewis Jr, N. A., & Sekaquaptewa, D. (2016). Beyond test performance: A broader view of stereotype threat. *Current Opinion in Psychology*, 11, 40-43.
<https://doi.org/10.1016/j.copsy.2016.05.002>

- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R* (extended ed.). Hogrefe.
- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology, 96*, 1089–1103.
<https://doi.org/10.1037/a0015703>
- Mallett, R. K., Ford, T. E., & Woodzicka, J. A. (2016). What did he mean by that? Humor decreases attributions of sexism and confrontation of sexist jokes. *Sex Roles, 75*(5-6), 272–284. <https://doi.org/10.1007/s11199-016-0605-2>
- Marchese, D. (2019, September 30). Bill Maher on the perils of political correctness. *The New York Times Magazine*.
<https://www.nytimes.com/interactive/2019/09/30/magazine/bill-maher-interview.html>
- Martin, R. A. (1996). The situational humor response questionnaire (SHRQ) and coping humor scale (CHS): A decade of research findings. *Humor, 9*, 251–272.
<https://doi.org/10.1515/humr.1996.9.3-4.251>
- Martin, R. A., & Lefcourt, H. M. (1983). Sense of humor as a moderator of the relation between stressors and moods. *Journal of Personality and Social Psychology, 45*, 1313–1324. <https://doi.org/10.1037/0022-3514.45.6.1313>
- Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality, 37*, 48–75. [https://doi.org/10.1016/S0092-6566\(02\)00534-2](https://doi.org/10.1016/S0092-6566(02)00534-2)
- Martiny, S. E., & Nikitin, J. (2019). Social identity threat in interpersonal relationships: Activating negative stereotypes decreases social approach motivation. *Journal of Experimental Psychology: Applied, 25*, 117–128. <https://doi.org/10.1037/xap0000198>

McCauley, C., Woods, K., Coolidge, C., & Kulick, W. (1983). More aggressive cartoons are funnier. *Journal of Personality and Social Psychology, 44*, 817–823.

<https://doi.org/10.1037/0022-3514.44.4.817>

McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science, 21*, 1141–1149.

<https://doi.org/10.1177/0956797610376073>

McGraw, A. P. & Warren, C. (2014). Benign violation theory. In S. Attardo (Ed.), *Encyclopedia of Humor Studies* (pp. 75–77). Sage Publications.

Mickes, L., Walker, D., Parris, P., Mankoff, R., & Christenfield, N. (2011). Who's funny: Gender stereotypes, humor production and memory bias. *Psychonomic Bulletin & Review, 19*, 108–112. <https://doi.org/10.3758/s13423-011-0161-2>

Morreall, J. (1987). *The philosophy of laughter and humor*. State University of New York Press.

Nguyen, H. H., & Ryan, A. M. (2008). Does stereotype threat affect cognitive ability test performance of minorities and women? A meta-analytic review of experimental evidence. *Journal of Applied Psychology, 93*, 1314–1335.

<https://doi.org/10.1037/a0012702>

Oswald, D. L., & Harvey, R. D. (2000-2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology, 19*, 338–356.

<https://doi.org/10.1007/s12144-000-1025-5>

Pennington, C. R., Litchfield, D., McLatchie, N., & Heim, D. (2019). Stereotype threat may not impact women's inhibitory control or mathematical performance: Providing support for the null hypothesis. *European Journal of Social Psychology, 49*, 717–734.

<https://doi.org/10.1002/ejsp.2540>

Plato (375 BC / 1974). *The Republic*. Hardmondsworth: Penguin.

- Saleem, M., Wojcieszak, M. E., Hawkins, I., Li, M., & Ramasubramanian, S. (2019). Social identity threats: How media and discrimination affect Muslim Americans' identification as Americans and trust in the US government. *Journal of Communication, 69*, 214–236. <https://doi.org/10.1093/joc/jqz001>
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38*, 194–201. <https://doi.org/10.1006/jesp.2001.1500>
- Schmader, T., Block, K., & Lickel, B. (2015). Social identity threat in response to stereotypic film portrayals: Effects on self-conscious emotion and implicit ingroup attitudes. *Journal of Social Issues, 71*, 54–72. <https://doi.org/10.1111/josi.12096>.
- Schmuck, D., Matthes, J., & Paul, F. H. (2017). Negative stereotypical portrayals of Muslims in right-wing populist campaigns: Perceived discrimination, social identity threats, and hostility among young Muslim adults. *Journal of Communication, 67*, 610–634. <https://doi.org/10.1111/jcom.12313>
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>
- Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629. <https://doi.org/10.1037/0003-066X.52.6.613>
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 37, pp. 397–407). Academic Press. [https://doi.org/10.1016/S0065-2601\(02\)80009-0](https://doi.org/10.1016/S0065-2601(02)80009-0)

- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement?. *Review of General Psychology, 16*, 93–102.
<https://doi.org/10.1037/a0026617>
- Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons.
In J. Goldstein & P. McGhee (Eds.), *Psychology of Humor* (pp. 81–100). Academic Press.
- Swani, K., Weinberger, M. G., & Gulas, C. S. (2013). The impact of violent humor on advertising success: A gender perspective. *Journal of Advertising, 42*, 308–319.
<https://doi.org/10.1080/00913367.2013.795121>
- Thai, M., Borgella, A. M., & Sanchez, M. S. (2019). It's only funny if we say it: Disparagement humor is better received if it originates from a member of the group being disparaged. *Journal of Experimental Social Psychology, 85*, 103838.
<https://doi.org/10.1016/j.jesp.2019.103838>
- Thomae, M. & Pina, A. (2015). Sexist humor and social identity: the role of sexist humor in men's in-group cohesion, sexual harassment, rape proclivity, and victim blame. *Humor, 28*, 187–204. <https://doi.org/10.1515/humor-2015-0023>
- Thomae, M., & Viki, G. T. (2013). Why did the woman cross the road? The effect of sexist humor on men's rape proclivity. *Journal of Social, Evolutionary, and Cultural Psychology, 7*, 250–269. <https://doi.org/10.1037/h0099198>
- Warren, C., & McGraw, A. P. (2016). Differentiating what is humorous from what is not. *Journal of Personality and Social Psychology, 110*(3), 407–430.
<http://dx.doi.org/10.1037/pspi0000041>
- Weisfeld, G. E. (2006). Humor appreciation as an adaptive esthetic emotion. *Humor, 19*(1), 1–26. <https://doi.org/10.1515/HUMOR.2006.001>

- Wright, C. L., DeFrancesco, T., Hamilton, C., & Vashist, N. (2018). "Boy's Club:" Examining sexist humor on types of sexism and femininity ideology using two research approaches. *Humor*, 31, 129–150. <https://doi.org/10.1515/humor-2017-0108>
- Woodzicka, J. A., & Ford, T. E. (2010). A framework for thinking about the (not-so-funny) effects of sexist humor. *Europe's Journal of Psychology*, 6, 174–195. <https://doi.org/10.5964/ejop.v6i3.217>
- Woodzicka, J. A., Mallett, R. K., Hendricks, S., & Pruitt, A. V. (2015). It's just a (sexist) joke: Comparing reactions to sexist versus racist communications. *Humor*, 28, 289–309. <https://doi.org/10.1515/humor-2015-0025>
- Zillmann, D. (1983). Disparagement humor. In P. E. McGhee & J. H. Goldstein (Eds.), *Handbook of Humor Research* (pp. 85–107). Springer. https://doi.org/10.1007/978-1-4612-5572-7_5

Figures

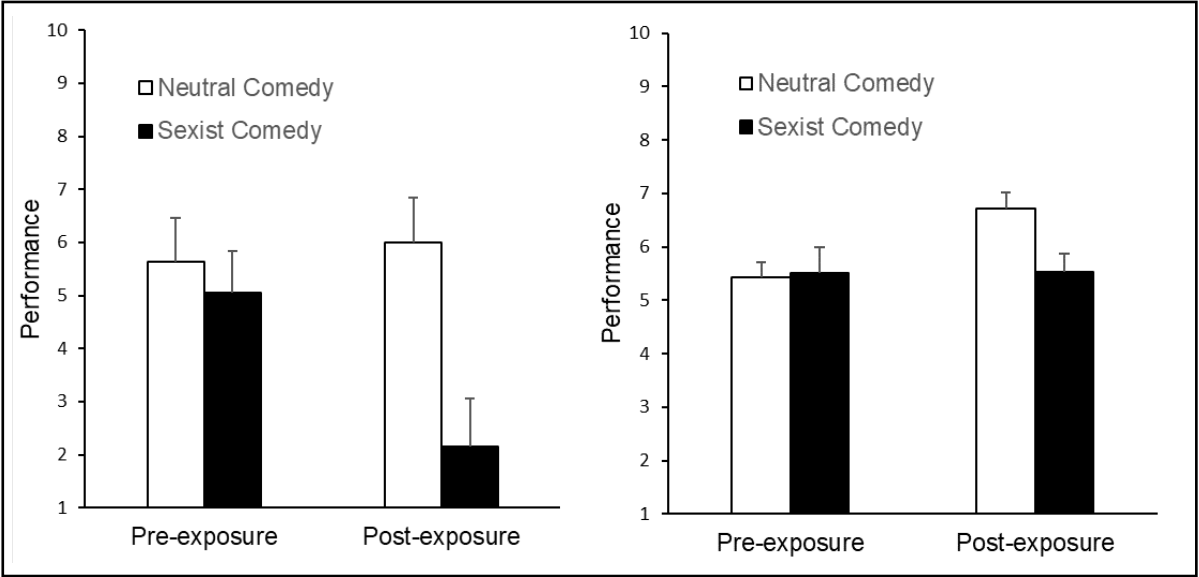


Figure 1. Interaction between Point of Measurement and Video Condition (control vs. sexist comedy) in Experiment 1 (left panel) and Experiment 4 (right panel).

----- **Online Supplement** -----

Supplementary Material**Just a Joke? Can Sexist Comedy Harm Women's Cognitive Performance?**

| | |
|--|----|
| Supplement 1: Stimuli..... | 3 |
| Supplement 2: Experiment 2 | 4 |
| Supplement 3: Experiment 3 | 13 |
| Supplement 4: Humor (state and trait) as a predictor for cognitive performance..... | 18 |
| Supplement 5: Gender identification and benevolent sexism as predictors for cognitive performance (Experiment 4) | 19 |
| Supplement 6: Emotional reactions as mediating variables (Experiment 4)..... | 22 |
| References | 26 |

Tables

Table S1: Overview of the Stimuli Used in the four Experiments

Table S2: Baseline (Pre-Exposure) and Post-Exposure Cognitive Performance Test Scores by Experimental Condition in Experiment 2.

Table S3: Experiment 2: Moderation Analysis with Experimental Condition and Perceived Humor as the Predictor and Cognitive Performance as the Criterion, using the PROCESS-macro by Hayes (2013), Model 1 (10'000 bootstraps).

Table S4: Moderation Analysis with Experimental Condition and Identification with Women, Identification with Feminists, and Perceived Femininity as the Predictor, and Cognitive Performance as the Criterion (Experiment 4)

Table S5: Mediation Analysis with the Emotional Reactions Anger, Disgust, Hostility, Surprise, and Amusement as Mediators, Experimental Condition as the Predictor, and Cognitive Performance as the Criterion (Experiment 4)

Figures

Figure S1. Interaction between Experimental Condition and Perceived Humor (Experiment 2). Sexist Comedy: Mario Barth; First Control: Luke Mockridge; Second Control: Hape Kerkeling.

Figure S2: Mediation of Emotional Reactions (Anger, Disgust, Hostility, Surprise, and Amusement) between Experimental Condition and Change in Performance (Experiment 4)

Supplement 1: Stimuli

Table S1

Overview of the Stimuli Used in the Four Experiments

| Exp. | Sexist condition | Control condition (1) | Control condition (2) |
|------|------------------|---|---|
| 1 | Comedian | Mario Barth | Luke Mockridge |
| | Show | „Männer sind primitiv, aber glücklich“ (2007) | “Der Wahnsinn in den 90ern – Luke Mockridge - TV total” (2014) |
| | Retrieved | July 4, 2015 from https://www.youtube.com/watch?v=jCimA8iUbl4&feature=youtu.be [Video File]. | July 4, 2015 from https://www.youtube.com/watch?v=fNfUZhTiDBQ [Video File]. |
| 2 | Comedian | See Experiment 1 | See Experiment 1 |
| | Show | | Hape Kerkeling „Wieder auf Tour – Live“ (2005) |
| | Retrieved | | November 2, 2016 from https://www.youtube.com/watch?v=HCzqgDYxkY&feature=youtu.be [Video File]. |
| 3 | See Experiment 1 | - | See Experiment 2 |
| 4 | Comedian | Luke Mockridge | Luke Mockridge |
| | Show | “Luke Mockridge über Tinder und die Welt der Frauen - TV total” (2015) | “Luke Mockridges Karnevalsbilanz - TV total” (2014) |
| | Retrieved | August 3, 2018 from https://www.youtube.com/watch?v=Y87zYKPw7B8&t=283s [Video File]. | July 31, 2018 from https://www.youtube.com/watch?v=qI3lwgKMI_M&t=308s [Video File]. |

Note. All video clips used in this research can be requested from the first author.

Supplement 2: Experiment 2

Experiment 2 was an extended replication of the first study. Further, we examined two exploratory research questions: Is there a difference in women's gender identification after watching sexist comedy compared to non-sexist comedy? Does watching sexist comedy decrease women's domain identification with math? This study was preregistered under [aspredicted.org](http://aspredicted.org/blind.php?x=rt5dt8): <http://aspredicted.org/blind.php?x=rt5dt8>.

Stimuli

Concerning the experimental condition and the first control condition, the same stimulus material as in Experiment 1 was used. Due to the differences in perceived humor in Experiment 1 between the sexist and the non-sexist comedy clip, we included a second control group, using another non-sexist clip (Comedian: Hape Kerkeling), which was pretested to be comparably funny to the experimental condition.

Measures

Self-Reported Humor. The same measures as in Experiment 1 were used: the *Perceived Humor Scale* (five items, $\alpha = .90$) to assess humor as a state and the *Coping Humor Scale* (seven items, $\alpha = .69$) to assess humor as a trait.

Cognitive Performance. The same numerical and figural subtests of the I-S-T 2000 R were administered.

Gender Identification. We assessed women's identification with their gender with the *Gender Identification Scale* (Schmader, 2002). Four items are answered on a five-point scale (e.g., "Being a woman is an important part of my self-image"; 1 = "strongly disagree" to 5 = "strongly agree"). The scale's reliability was acceptable, $\alpha = .66$.

Domain Identification. Participants were asked to indicate their interest in a hypothetical seminar as part of their university studies. They read two vignettes, one indicating that the focus of the seminar lies on "verbal ability and linguistic analyses", while

the second one focuses on “numbers and statistical analyses”. For each vignette, participants could indicate their interest on a 7-point scale (1 = “I find it not interesting at all” to 7 = “I find it highly interesting”).

Participants and Procedure

Based on the previous experiment, we expected a small to moderate effect. As this study was an extended replication of Experiment 1, which only had two experimental groups ($n = 102$, cf. conditions 1 and 2 in the current study), we decided to include 20% more participants per group to increase power, resulting in $n = 180$ women in total, i.e., $n = 60$ per experimental condition.

Again, the computer-based experiment used a pre-post control study design in a laboratory with a male research assistant. Participants were 188 women, but seven did not finish the study. The final sample consisted of $n = 181$ female undergraduates (age: $M = 21.12$, $SD = 4.18$, range 18-53 years), who received course credit or a small monetary remuneration.

After completing the Coping Humor Scale and filling in the first subset of the cognitive performance test, participants watched one out of the three videos (random assignment: sexist vs. control 1 vs. control 2) and rated how humorous they perceived the video. Then, they worked on the second subset of the cognitive performance test, filled in the Gender Identification Scale, and answered the Domain Identification items. Before being debriefed, demographic data, and familiarity with and attitude towards the comedian (on a 5-point scale from 1 = “strongly disagree” to 5 = “strongly agree”) were administered. Mario Barth ($M = 2.69$, $SD = 1.23$) was rated significantly less popular among the participants than Luke Mockridge ($M = 4.07$, $SD = 1.11$) and Hape Kerkeling ($M = 3.18$, $SD = 1.08$), $F(2, 176) = 22.41$, $p < .001$, $d = 1.00$. In total, the study took about 45 minutes.

Results

Main Effect of the Treatment on Cognitive Performance

The results suggest that the comedy condition influenced test performance (see Table S2 for descriptive statistics). A 2 (point of measurement, repeated) \times 3 (treatment) ANOVA yielded a main effect for point of measurement, $F(1, 178) = 11.88, p = .001, d = .52$, and a significant interaction between point of measurement and experimental condition, $F(2, 178) = 4.62, p = .011, d = .45$.⁴ Unexpectedly, participants in the sexist comedy condition (Mario Barth), $p = .161, 95\% \text{ CI } [-2.07; 0.34]$, and in the second non-sexist comedy condition (Hape Kerkeling), $p = .849, 95\% \text{ CI } [-1.28; 1.05]$, showed no change in performance, whereas performance of participants in the first non-sexist comedy condition (Luke Mockridge) increased, $p < .001, 95\% \text{ CI } [-3.78; -1.43]$.

⁴ This result includes five participants who were suspicious of the experiment's goal. Similar results were obtained when these participants were excluded: main effect, $F(1, 173) = 14.76, p < .001, d = .59$; interaction, $F(2, 173) = 4.63, p = .011, d = .46$.

Table S2

Baseline (Pre-Exposure) and Post-Exposure Cognitive Performance Test Scores by Experimental Condition in Experiment 2

| | <i>M (SD)</i> | <i>Min</i> | <i>Max</i> |
|--|---------------|------------|------------|
| Sexist condition (Mario Barth) | | | |
| <i>n</i> | 58 | | |
| Baseline | 4.71 (5.33) | -8 | 16 |
| Post-Exposure | 5.57 (4.61) | -11 | 15 |
| Difference Score | 0.86 (4.88) | -12 | 13 |
| Non-sexist condition 1 (Luke Mockridge) | | | |
| <i>n</i> | 61 | | |
| Baseline | 3.44 (5.46) | -10 | 17 |
| Post-Exposure | 6.05 (4.38) | -2 | 18 |
| Difference Score | 2.61 (4.27) | -7 | 13 |
| Non-sexist condition 2 (Hape Kerkeling) | | | |
| <i>n</i> | 62 | | |
| Baseline | 4.39 (4.79) | -5 | 16 |
| Post-Exposure | 4.50 (4.60) | -7 | 14 |
| Difference Score | 0.11 (4.81) | -11 | 11 |

Notes. Min = Lowest score, Max = Highest score. Difference Score = post minus pre score.

Preliminary Analyses of Humor Variables

Regarding perceived humor, the three videos differed significantly from each other, $F(2, 178) = 14.13, p < .001, d = .80$. The first control video was perceived to be funnier than the sexist comedy clip ($p < .001$) and the second control video ($p < .001$), while the sexist comedy clip and the second control video did not differ ($p = .88$). Again, descriptive statistics indicated more variance (Levene's test: $F(2,178) = 3.28, p = .019$) in the sexist condition ($M = 2.99, SD = 0.77$) and in the second control condition ($M = 2.98, SD = 0.72$) than in the first control condition ($M = 3.59, SD = 0.55$). Coping sense of humor did not differ between the three conditions, $F(2, 178) = 2.21, p = .113, d = .31$. Both humor variables were unrelated for the total sample, $r = -.07, p = .38$, and in the three subgroups.

Main Effect on Gender and Domain Identification

Participants showed more interest in the verbal than in the mathematical domain, irrespective of experimental condition. A 2 (domain, repeated) \times 3 (treatment) ANOVA yielded a main effect for domain, $F(1, 178) = 172.08, p < .001, d = 1.97$, but no significant interaction between domain and video condition, $F(2, 178) = 0.14, p = .866, d = .10$. Likewise, a univariate ANOVA revealed no difference regarding participants' gender identification, $F(2, 178) = 0.51, p = .60, d = .16$, after watching the comedy clips. Neither perceived humor nor coping sense of humor moderated any of these effects.

Moderation of Perceived Humor (State)

Perceived humor moderated the effect of the comedy clips. Treatment (dummy coded) and perceived humor (z-standardized) as well as the interactions between the variables served as predictors for difference scores of the cognitive performance measure. Detailed results are displayed in Table S3. R^2 significantly increased due to the interaction, $\Delta R^2 = 0.04, F(2, 175) = 3.69, p = .027$. The interaction between the experimental treatment and perceived humor is displayed in Figure S1. The simple slopes revealed that perceived humor predicted performance in the sexist video condition, $B = -1.71, SE = 0.58, p = .004, 95\% \text{ CI} [-2.85; -0.56]$, but not in the two non-sexist conditions, $B = 0.67, SE = 0.77, p = .38, 95\% \text{ CI} [-0.85; 2.20]$, and $B = 0.03, SE = 0.60, p = .96, 95\% \text{ CI} [-1.16; 1.21]$. Notably, the effect of perceived humor on performance in the sexist comedy condition was in the opposite direction compared to Experiment 1: the less funny the sexist comedy was perceived, the better the performance.

Table S3

Experiment 2: Moderation Analysis with Experimental Condition and Perceived Humor as the Predictor and Cognitive Performance as the Criterion, using the PROCESS-macro by Hayes (2013), Model 1 (10'000 bootstraps)

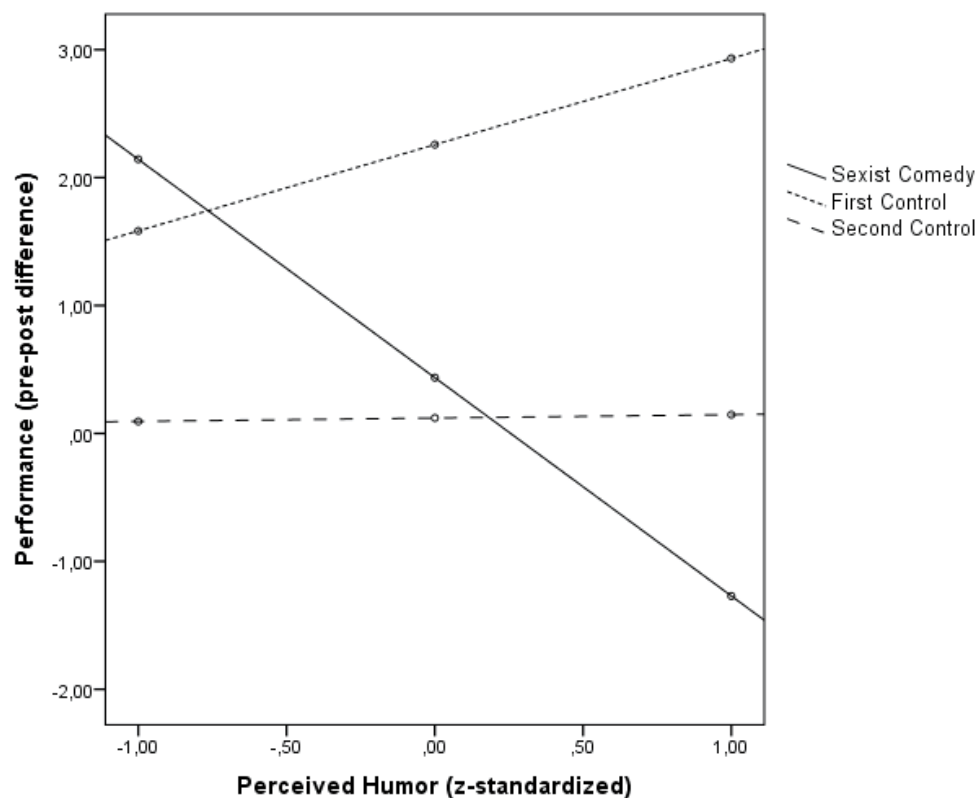
| | <i>B</i> | <i>SE_B</i> | <i>t</i> | <i>p</i> | <i>LLCI</i> | <i>ULCI</i> |
|---------------------------------------|----------|-----------------------|----------|----------|-------------|-------------|
| Constant | 0.44 | 0.62 | 0.71 | .481 | -0.78 | 1.66 |
| Perceived Humor | -1.71 | 0.58 | -2.94 | .004 | -2.85 | -0.56 |
| Sexist Comedy vs. Control 1 (Dummy 1) | 1.82 | 0.94 | 1.93 | .055 | -0.04 | 3.68 |
| Sexist Comedy vs. Control 2 (Dummy 2) | -0.32 | 0.86 | -0.37 | .715 | -2.02 | 1.39 |
| Dummy 1 × Perceived Humor | 2.38 | 0.97 | 2.46 | .015 | 0.47 | 4.29 |
| Dummy 2 × Perceived Humor | 1.73 | 0.83 | 2.08 | .039 | 0.09 | 3.38 |

$R = 0.31$, $R^2 = 0.10$, $F(5, 175) = 3.80$, $p = .003$.

Note. Sexist Comedy: Mario Barth; Control 1: Luke Mockridge; Control 2: Hape Kerkeling.

Figure S1

Interaction between Experimental Condition and Perceived Humor (Experiment 2). Sexist Comedy: Mario Barth; First Control: Luke Mockridge; Second Control: Hape Kerkeling



Moderation of Coping Sense of Humor (Trait)

Coping sense of humor did not moderate the effect of the comedy clips. The difference scores of the cognitive performance measure were regressed on treatment (dummy coded) and coping sense of humor (z-standardized) and the interactions between the variables. The complete model was significant, $R = 0.25$, $R^2 = 0.06$, $F(5, 175) = 2.40$, $p = .039$. Yet, adding the interaction did not explain additional variance, $\Delta R^2 = 0.01$, $F(2, 175) = 0.78$, $p = .460$. There was no main effect for coping sense of humor, $B = -0.15$, $SE_B = 0.56$, $p = .79$, 95% CI [-1.25; 0.95], a main effect for the first dummy (sexist comedy vs. control 1), $B = 1.82$, $SE_B = 0.87$, $p = .038$, 95% CI [0.10; 3.54], but not for the second dummy (sexist comedy vs. control 2), $B = -0.75$, $SE_B = 0.86$, $p = .38$, 95% CI [-2.45; 0.95]. The interactions were not significant, $B = 0.80$, $SE_B = 0.85$, $p = .35$, 95% CI [-0.89; 2.48], and $B = 0.97$, $SE_B = 0.85$, $p = .25$, 95% CI [-0.70; 2.64], respectively.

Discussion

The results of Experiment 2, in which another control condition was included to ensure comparable funniness of the stimulus material, provide a mixed picture. Women exposed to sexist comedy performed worse (post vs. pre-exposure) than women in the first control condition, whereas no difference to the second control condition was observed. In this study, the effect was not as distinct as in Experiment 1, leading to no clear conclusion whether watching comedy with or without disparaging humor against women can elicit social identity threat. In contrast to Experiment 1, participants in the first control condition showed a learning effect from pre to post assessment, while this effect was inhibited in the sexist condition, and also not apparent in the second control condition.

Neither gender identification nor domain identification were affected by watching sexist comedy. Stereotype threat theory and research identified gender identification as an important predictor for the threat effect to occur (Schmader, 2002). Further, it has been shown

that women who are less identified with their gender find sexist humor more amusing (Kochersberger et al., 2014). Thus, gender identity might rather be an individual-difference variable that affects outcomes elicited by sexist humor than a dependent variable (see Experiment 4).

As in Experiment 1, perceived humor predicted performance in the sexist comedy condition. However, remarkably, the effect was reversed, indicating that the less humorous participants perceived the sexist video, the more they improved their performance. Previous research has shown that explicit stereotype threat cues were less detrimental for women's performance than subtle cues (Nguyen & Ryan, 2008). As Mario Barth, whose sexist jokes are rather blatant, was rated less popular among the participants than Luke Mockridge and Hape Kerkeling, it may have been easier to "brush off" the obviously sexist jokes by a disliked comedian. If the more popular comedian Luke Mockridge had told sexist jokes, effects might have looked differently (see Experiment 4).

The contradicting results of Experiments 1 and 2 call for a more detailed inspection of perceived humor and its role in the proposed social identity threat effect. Perceived humor is likely a reflection of several responses to a humorous stimulus, including, but not limited to the funniness of the stimulus. Negative emotions due to the sexist nature of a stimulus might come into play as well. A negative mood can lead to less efficient task solving and lower performance in a learning phase (Brand et al., 2007). *Resource Allocation Theory* (Ellis & Ashbrook, 1988, 1989) suggests that emotions can disrupt performance as they reduce working memory capacity to process the task at hand. Negative or irrelevant emotions are suggested to be particularly disruptive, as they trigger task-irrelevant thoughts during task completion. Sexist and neutral jokes elicited differences in ratings of perceived humor, yet similarly triggered Duchenne smiling (LaFrance & Woodzicka, 1998), suggesting that the measurement of facial expressions can reveal more subtle emotions beyond self-report

measures. In Experiment 3, in order to examine the emotional boundary conditions of the social identity threat effect, we assessed emotional responses in a more sophisticated way.

Supplement 3: Experiment 3

Previous research has shown that women who are exposed to sexist jokes are more likely to report anger, surprise, and disgust (LaFrance & Woodzicka, 1998). To examine the role of different emotions during the process in more detail, we included a continuous assessment of emotional responses to the stimulus in Experiment 3. As situational humor manifests physically in a higher likeliness to activate the smile muscle (Danzer et al., 1990), we used an automated facial coding software. The experimental design was adapted to the altered methodological approach.

Stimuli

We used the stimulus material rated comparably funny in Experiment 2: Mario Barth functioned as the sexist comedian and Hape Kerkeling as the non-sexist comedian.

Measures

Self-Reported Humor. As in Experiments 1 and 2, we used the *Perceived Humor Scale* (five items, $\alpha = .86$) to assess self-reported humor as a state, and the *Coping Humor Scale* (seven items, $\alpha = .60$) to assess humor as a trait.

Emotional Facial Expressions. Participants were recorded while they watched the comedy clip. Under the assumption that basic emotions correspond with particular facial expressions, the automated facial coding software (*FaceReader*, Noldus, 2013) assesses the category and the intensity of facial expressions of six basic emotions (i.e., happy, sad, angry, surprised, scared, disgusted) based on the *Facial Action Coding System* (FACS; Ekman & Friesen, 1978) as well as a neutral state. In its analyses, the software assigns a value from 0 (indicating the emotion is absent) to 1 (indicating the emotion is fully present) for every frame of the videos of the participants' facial expressions. We calibrated the software to account for participants' physiognomic differences (i.e., *participant calibration*) and calculated the average score of each emotion for the duration of the comedy clip. The FaceReader software

has been shown to be a reliable and valid measurement of emotional responses to various media stimuli (e.g., Chentsova-Dutton & Tsai, 2010; Lewinski et al., 2014).

Cognitive Performance. Participants' cognitive performance was assessed with three figural subtests of the intelligence test *I-S-T 2000 R* (Liepmann et al., 2007). We used the subtests that include figure selection tasks, mental rotation tasks, and figural matrices (ten items each, upper half of form A, i.e., the more difficult items). All items were multiple-choice questions with five answer options, out of which only one was correct. Participants had five minutes to complete each group of tasks; then, the survey software was programmed to jump to the next page. Again, the performance score was corrected for guessing based on the multiple-choice format: $\text{correct answers} - \left(\frac{\text{incorrect answers}}{4} \right)$

Participants and Procedure

Under the assumption of an effect size of $d = .50$, $\alpha = 0.05$, and a power of $1 - \beta = .80$, the optimal sample size includes a total number of $N = 102$ women. Participants were 103 female university students, yet $n = 3$ people had to be excluded due to disturbances during the data collection process. The final sample consisted of 100 women (age: $M = 21.85$, $SD = 4.00$, range 18-49 years) who received course credit or a small monetary remuneration. They were thoroughly informed about the procedure, gave their written consent to being recorded, and were positioned for optimal recording. All ethical guidelines and data protection policies were met.

First, participants completed the Coping Humor Scale. Then, they were randomly assigned to watch one of the two videos and rated how humorous they perceived the video. Subsequently, they completed the cognitive performance test and answered the demographic questions. Finally, participants were asked to put on a happy, sad, surprised, angry, fearful, disgusted, and neutral face, which was needed for the calibration of the FaceReader analyses, and then thanked and debriefed.

Results

Main Effect of the Treatment on Cognitive Performance

No social identity threat effect was observed. Participants' test performance in the sexist video condition ($N = 49$, $M = 11.53$, $SD = 5.29$) and in the non-sexist condition ($N = 51$, $M = 11.86$, $SD = 4.47$) did not differ, $t(98) = 0.34$, $p = .74$, $d = -0.07$.

Preliminary Analyses of Humor Variables

Coping sense of humor did not differ between both conditions, $t(98) = -0.28$, $p = .78$, $d = -0.06$, but regarding perceived humor, the non-sexist comedy ($M = 2.99$, $SD = 0.60$) was rated somewhat funnier than the sexist comedy ($M = 2.73$, $SD = 0.78$), $t_{Welch}(89.75) = 1.90$, $p = .06$, $d = 0.38$. Again, the results suggest inhomogeneity of variances (Levene's test: $F = 5.82$, $p = .018$). The two scales were not significantly correlated for the total sample, $r = .04$, $p = .69$, and in the two subgroups.

Moderation of Perceived Humor (State)

Perceived humor did not moderate the effect. Treatment (dummy coded), perceived humor (z-standardized), and the interaction served as predictors for cognitive performance. The interaction was not significant, $B = -0.64$, $SE_B = 1.02$, $p = .53$, 95% CI [-2.66; 1.39]. The simple slopes revealed a tendency for more perceived humor predicting better performance in the sexist video condition, $B = 1.42$, $SE = 0.80$, $p = .080$, 95% CI [-0.18; 3.02], but not in the non-sexist condition, $B = 0.78$, $SE = 0.63$, $p = .214$, 95% CI [-0.46; 2.03].

Moderation of Coping Sense of Humor (Trait)

Coping sense of humor did not moderate the effect. Cognitive performance was regressed on treatment (dummy coded), coping sense of humor (z-standardized), and the interaction. Like in Experiments 1 and 2, the interaction was not significant, $B = -1.15$, $SE_B = 0.99$, $p = .24$, 95% CI [-3.11; 0.80].

Moderation of Emotional Facial Expressions

Treatment (dummy coded) and the FaceReader score for happy (z-standardized) as well as their interaction served as predictors for cognitive performance. The interaction was not significant, $B = -0.72$, $SE_B = 0.98$, $p = .47$, 95% CI [-2.66; 1.23]. None of the other FaceReader scores of the emotional facial expressions moderated the effect. Overall, neutral ($M = 0.34$, $SD = 0.18$) and happy ($M = 0.18$, $SD = 0.12$) were the most prevalent facial expressions while watching the comedy clips. Self-reported perceived humor was significantly correlated with the facial expressions happy ($r = .46$, $p < .001$), neutral ($r = -.23$, $p = .023$), and disgusted ($r = .23$, $p = .022$). The two groups did not differ regarding any of the emotional facial expressions, except for disgust, $t_{Welch}(50.91) = 2.87$, $p = .006$, $d = 0.57$, which was more prevalent in the non-sexist control condition ($M = 0.02$, $SD = 0.05$) than in the sexist condition ($M = 0.00$, $SD = 0.00$). However, due to the very low variance, this result should not be interpreted before replication. Further, there was no difference regarding the overall valence of the emotional facial expressions, $t(98) = -1.15$, $p = .25$, $d = -0.23$.

Discussion

In contrast to the two previous studies, the results of Experiment 3 revealed no significant effect of the comedy condition. Deviating from Experiments 1 and 2, we employed a two-group experimental design. The first two experiments benefit from the repeated-measures design, which subtracts the between-participants variance, and thus, reveals a more accurate indicator of the true effect. Thus, choosing a different experimental design might have led to the non-significant main effect.

Furthermore, none of the suggested moderators explained under which boundary conditions the social identity threat effect occurs. On a critical note, facial expressions of emotions mainly function as interpersonal communication between individuals. The situation (sitting in front of a computer) had no interactive element. To ensure a good recording of

participants' faces, they had been instructed to sit in the same spot, not to move forward or backward, and so on, which made the situation rather artificial. This could have led to a change in participants' reactions to the comedy clips, and thus, to a biased assessment of the continuous emotional facial responses. In Experiment 4, we therefore turned to self-reported emotions and tested whether emotions mediated (instead of moderated) the threat effect (see Schmader et al., 2008, for an integrated process model of stereotype threat which suggests a mediating role of emotions).

A possible explanation for the (non-significant) effects in Experiments 2 and 3 could be that the control clip was perceived to contain subtle disparagement humor, too. In hindsight, this could be the case, because the (gay) comedian Hape Kerkeling commented on his age and weight. He also collected responses from female audience members and made fun of them (yet not playing on stereotypes). Therefore, Experiment 4 used different comedy clips.

Supplement 4: Humor (state and trait) as a predictor for cognitive performance**Experiment 1**

Independent of the experimental condition, two separate linear regression analyses revealed that perceived humor (z-standardized) significantly predicted an improvement in performance (difference score of pre and post assessment), $B = 1.71$, $SE_B = 0.69$, $\beta = 0.24$, $p = .015$, while coping sense of humor (z-standardized) did not, $B = 0.96$, $SE_B = 0.69$, $\beta = 0.14$, $p = .165$.

Experiment 2

Independent of the experimental condition, two separate linear regression analyses revealed that neither perceived humor (z-standardized), $B = -0.07$, $SE_B = 0.35$, $\beta = -0.02$, $p = .837$, nor coping sense of humor (z-standardized), $B = 0.29$, $SE_B = 0.36$, $\beta = 0.06$, $p = .415$, predicted change in performance.

Experiment 3

Independent of the experimental condition, two separate linear regression analyses revealed that perceived humor (z-standardized) significantly predicted performance, $B = 1.02$, $SE_B = 0.48$, $\beta = 0.21$, $p = .037$, while coping sense of humor (z-standardized) did not, $B = -0.17$, $SE_B = 0.49$, $\beta = -0.03$, $p = .734$. Similar to Experiment 1, the more humorous the videos were perceived, the better the performance.

Experiment 4

Independent of the experimental condition, two separate linear regression analyses revealed that neither perceived funniness (z-standardized), $B = 0.16$, $SE_B = 0.22$, $\beta = 0.08$, $p = .470$, nor perceived sexism (z-standardized), $B = 0.31$, $SE_B = 0.22$, $\beta = 0.16$, $p = .159$, predicted change in performance.

Supplement 5: Gender identification and benevolent sexism as predictors for cognitive performance (Experiment 4)

Women's individual levels of benevolent sexism correlate positively with negative emotional reactions to sexist humor, such as disgust (LaFrance & Woodzicka, 1998). Further, benevolent sexism can negatively affect women's cognitive performance (Dardenne et al., 2007). We suggest that women with high levels of benevolent sexism show a larger decrease in performance in the sexist condition as compared to women in the control condition.

Measures

Gender Identification. Gender identification was operationalized in two ways. First, four items each (Doosje et al., 1995; van Breen et al., 2017) were answered on a seven-point scale (e.g., "I identify with other women/feminists"; 1 = "strongly disagree" to 7 = "strongly agree") to assess identification with women ($\alpha = .61$) and with feminists ($\alpha = .94$) separately. Second, perceived femininity was measured using the *Traditional Masculinity-Femininity Scale* (Kachel et al., 2016). Six items were answered on a seven-point scale (e.g., "I consider myself ..."; 1 = "very masculine" to 7 = "very feminine"; $\alpha = .76$).

Benevolent Sexism. Benevolent sexism was measured using the 11-item subscale from the *Ambivalent Sexism Inventory* (Glick & Fiske, 1996; e.g., "Women should be cherished and protected by men", 1 = "do not agree at all" to 7 = "strongly agree"; $\alpha = .87$).

Results

Gender Identification as a Predictor for Cognitive Performance

Three separate linear regression analyses revealed that, independent of the experimental condition, neither identification with women (z-standardized), $B = 0.03$, $SE = 0.22$, $\beta = 0.01$, $p = .90$, nor with feminists (z-standardized), $B = 0.15$, $SE = 0.22$, $\beta = 0.08$, $p = .51$, nor perceived femininity (z-standardized), $B = -0.36$, $SE = 0.22$, $\beta = 0.18$, $p = .10$, predicted change in performance.

Moderation of Gender Identification

Identification with women moderated the effect of the comedy clips. Treatment (dummy coded), identification with women (z-standardized), and the interaction served as predictors for cognitive performance. The interaction caused a significant increase in R^2 , $\Delta R^2 = 0.045$, $F(1,77) = 4.10$, $p = .046$. The simple slopes for identification with women revealed a tendency to predict performance in the non-sexist condition, $B = 0.52$, $SE = 0.30$, $p = .087$, 95% CI [-0.08; 1.12], but not in the sexist condition, $B = -0.32$, $SE = 0.29$, $p = .27$, 95% CI [-0.89; 0.25]. Descriptively, contrary to what stereotype threat theory suggests (Schmader, 2002), in the sexist comedy condition, the more the participants identified with women, the better their performance tended to be. In the non-sexist condition, the effect pointed in the opposite direction: the higher the identification, the lower the performance tended to be. Neither identification with feminists, nor perceived femininity moderated the threat effect of the comedy clips (for detailed results see Table S4).

Moderation of Benevolent Sexism

Benevolent sexism did not significantly moderate the effect of the comedy clips. Treatment (dummy coded) and benevolent sexism (z-standardized) as well as the interaction between the variables served as predictors of cognitive performance. The interaction caused no significant increase in R^2 , $\Delta R^2 = 0.016$, $F(1,77) = 1.37$, $p = .25$. The simple slopes revealed that the threat effect was statistically significant for women with low, $B = 1.75$, $SE = 0.59$, $p = .004$, 95% CI [0.57 2.93], and intermediate levels of benevolent sexism, $B = 1.26$, $SE = 0.42$, $p = .004$, 95% CI [0.43; 2.10], but not for those with high levels of benevolent sexism, $B = 0.77$, $SE = 0.59$, $p = .20$, 95% CI [-0.41; 1.95]. This pattern speaks against our hypothesis.

Table S4

Moderation Analysis with Experimental Condition and Identification with Women, Identification with Feminists, and Perceived Femininity as the Predictor, and Cognitive Performance as the Criterion (Experiment 4)

| | <i>B</i> | <i>SE_B</i> | <i>t</i> | <i>p</i> | <i>LLCI</i> | <i>ULCI</i> |
|---|----------|-----------------------|----------|----------|-------------|-------------|
| Identification with women | | | | | | |
| Constant | -1.33 | 0.28 | -4.81 | <.001 | -1.88 | -0.78 |
| Identification with women | 0.52 | 0.30 | 1.73 | .087 | -0.08 | 1.12 |
| Sexist Comedy vs. Control | 1.27 | 0.41 | 3.07 | .003 | 0.45 | 2.09 |
| Condition × Identification with women | -0.84 | 0.41 | -2.03 | .046 | -1.66 | -0.01 |
| <i>R</i> = 0.39, <i>R</i> ² = 0.15, <i>F</i> (3, 77) = 4.54, <i>p</i> = .006 | | | | | | |
| Identification with feminists | | | | | | |
| Constant | -1.29 | 0.28 | -4.62 | <.001 | -1.85 | -0.74 |
| Identification with feminists | 0.28 | 0.26 | 1.05 | .30 | -0.25 | 0.80 |
| Sexist Comedy vs. Control | 1.27 | 0.42 | 3.01 | .004 | 0.43 | 2.10 |
| Condition × Identification with feminists | -0.32 | 0.44 | -0.73 | .47 | -1.19 | 0.56 |
| <i>R</i> = 0.34, <i>R</i> ² = 0.12, <i>F</i> (3, 77) = 3.37, <i>p</i> = .023 | | | | | | |
| Perceived Femininity | | | | | | |
| Constant | -1.23 | 0.28 | -4.35 | <.001 | -1.79 | -0.67 |
| Perceived Femininity | -0.42 | 0.32 | -1.30 | .20 | -1.07 | 0.23 |
| Sexist Comedy vs. Control | 1.17 | 0.42 | 2.77 | .007 | 0.33 | 2.02 |
| Condition × Perceived Femininity | 0.27 | 0.43 | 0.64 | .52 | -0.58 | 1.13 |
| <i>R</i> = 0.35, <i>R</i> ² = 0.13, <i>F</i> (3, 77) = 3.68, <i>p</i> = .016 | | | | | | |

Discussion

Only identification with women moderated the effect of sexist humor on cognitive performance. However, the pattern was not as expected: in the non-sexist condition, the more participants identified with their group, the lower was their performance, whereas descriptively, in the sexist condition, the more participants identified with women, the better was their performance. These findings should not be interpreted before replication.

Supplement 6: Emotional reactions as mediating variables (Experiment 4)

Measures

Emotional Reaction. Participants were asked to evaluate their emotional reaction after watching the video on a seven-point scale (e.g., “I felt angry”; 1 = “do not agree at all” to 7 = “strongly agree”). The examined emotions were anger, disgust, hostility, surprise, and amusement (LaFrance & Woodzicka, 1998).

Results

Mediation by Emotional Reaction

The mediation analyses were conducted with the PROCESS-macro by Hayes (2013), using model 4 (5,000 bootstraps). Treatment (dummy coded) and emotional reaction (z-standardized) served as predictors for difference scores of the cognitive performance measures. Contrary to the assumption that emotions mediate the threat effect (Schmader et al., 2008), none of the regressions using anger, disgust, hostility, surprise, or amusement as mediators showed either a partial or full mediation. Although all mediations suggested significant direct and total effects, no indirect effects could be found (see Table S5 and Figure S2).

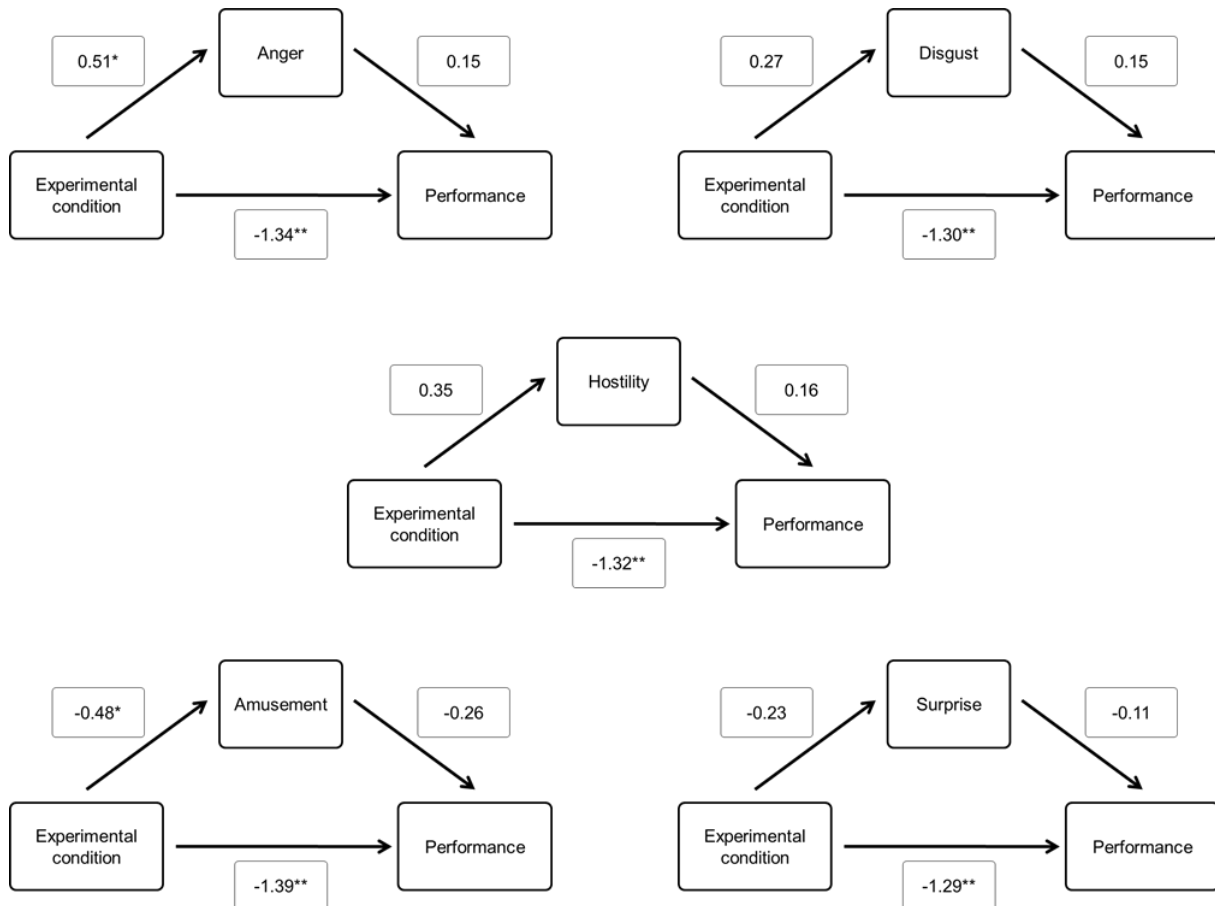
Table S5

Mediation Analysis with the Emotional Reactions Anger, Disgust, Hostility, Surprise, and Amusement as Mediators, Experimental Condition as the Predictor, and Cognitive Performance as the Criterion (Experiment 4)

| Effect | β | p | 95% CI | |
|-------------------------------------|---------|------|--------|-------|
| | | | Lower | Upper |
| Emotional Reaction Anger | | | | |
| Total | -1.26 | .003 | -2.093 | -0.43 |
| Direct | -1.34 | .003 | -2.20 | -0.68 |
| Indirect (Mediation) | -0.074 | | -0.18 | 0.45 |
| Emotional Reaction Disgust | | | | |
| Total | -1.26 | .003 | -2.09 | -0.64 |
| Direct | -1.30 | .003 | -2.15 | -0.66 |
| Indirect (Mediation) | 0.042 | | -0.11 | 0.25 |
| Emotional Reaction Hostility | | | | |
| Total | -1.26 | .003 | -2.09 | -0.43 |
| Direct | -1.32 | .003 | -2.16 | -0.47 |
| Indirect (Mediation) | -0.056 | | -0.081 | 0.37 |
| Emotional Reaction Surprise | | | | |
| Total | -1.26 | .003 | -2.09 | -0.43 |
| Direct | -1.29 | .003 | -2.13 | -0.44 |
| Indirect (Mediation) | 0.024 | | -0.081 | 0.21 |
| Emotional Reaction Amusement | | | | |
| Total | -1.26 | .003 | -2.09 | -0.43 |
| Direct | -1.39 | .002 | -2.24 | -0.53 |
| Indirect (Mediation) | -0.13 | | -0.043 | 0.51 |

Figure S2

Mediation of Emotional Reactions (Anger, Disgust, Hostility, Surprise, and Amusement) between Experimental Condition and Change in Performance (Experiment 4)

**Discussion**

Previous research suggested that negative emotions, which are caused by sexist humor, could play an important role concerning task solving and learning processes (e.g., Brand et al., 2007; see also LaFrance & Woodzicka, 1998). In Experiment 4, a different methodological approach was taken than in Experiment 3. Emotions were assessed via self-reports. In contrast to the moderation hypothesis in Experiment 3, mediation analyses could have exposed connections between sexist humor, emotions, and cognitive performance. However, the mediation analyses with self-reported emotions showed no evidence for the aforementioned relations. On a critical note, measuring emotions can be difficult, especially in

the context of sexist humor. Social norms, individual abilities to access emotions, as well as social desirability processes can interfere with the reliable evaluation of emotions.

References

- Brand, S., Reimer, T., & Opwis, K. (2007). How do we learn in a negative mood? Effects of a negative mood on transfer and learning. *Learning & Instruction, 17*, 1–16.
<https://doi.org/10.1016/j.learninstruc.2006.11.002>
- Chentsova-Dutton, Y. E., & Tsai, J. L. (2010). Self-focused attention and emotional reactivity: The role of culture. *Journal of Personality and Social Psychology, 98*, 507–519. <https://doi.org/10.1037/a0018534>
- Danzer, A., Dale, J. A., & Kliens, H. L. (1990). Effect of exposure to humorous stimuli on induced depression. *Psychological Reports, 66*, 1027–1036.
<https://doi.org/10.2466/pr0.1990.66.3.1027>
- Dardenne, B., Dumont, M., & Bollier, T. (2007). Insidious dangers of benevolent sexism: Consequences for women's performance. *Journal of Personality and Social Psychology, 93*, 764–779. <https://doi.org/10.1037/0022-3514.93.5.764>
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology, 31*, 410–436. <https://doi.org/10.1006/jesp.1995.1018>
- Ekman, P., & Friesen, W. V. (1978). *Manual for the Facial Action Code*. Consulting Psychologists Press.
- Ellis, H. C., & Ashbrook, P. W. (1988). Resource allocation model of the effects of depressed mood states on memory. In K. Fiedler & J. Forgas (Eds.), *Affect, cognition, and social behavior: New evidence and integrative attempts* (pp. 25–43). Hogrefe.
- Ellis, H. C., & Ashbrook, P. W. (1989). The "state" of mood and memory research: A selective review. *Journal of Social Behavior and Personality, 4*, 1–21.

- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491–512. <https://doi.org/10.1037//0022-3514.70.3.491>
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press. <https://doi.org/10.1111/jedm.12050>
- Kachel, S. M., Steffens, M. C., & Niedlich, C. (2016). Traditional masculinity and femininity: Validation of a new scale assessing gender roles. *Frontiers in Psychology, 7*:956. <https://doi.org/10.3389/fpsyg.2016.00956>
- Kochersberger, A., Ford, T., Woodzicka, J., Romero-Sanchez, M., & Carretero-Dios, H. (2014). The role of identification with women as a determinant of amusement with sexist humor. *Humor, 27*, 441–460. <https://doi.org/10.1515/humor-2014-0071>
- LaFrance, M., & Woodzicka, J. A. (1998). No laughing matter: Women's verbal and nonverbal reactions to sexist humor. In J. Swim & C. Stagnor (Eds.), *Prejudice: The target's perspective* (pp. 61–80). Academic Press. <https://doi.org/10.1016/B978-012679130-3/50038-7>
- Lewinski, P., Fransen, M. L., & Tan, E.S.H. (2014). Predicting advertising effectiveness by facial expressions in response to amusing persuasive stimuli. *Journal of Neuroscience, Psychology, and Economics, 7*, 1–14. <https://doi.org/10.1037/npe0000012>
- Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (extended ed.)*. Hogrefe.
- Nguyen, H. H., & Ryan, A. M. (2008). Does stereotype threat affect cognitive ability test performance of minorities and women? A meta-analytic review of experimental evidence. *Journal of Applied Psychology, 93*, 1314–1335. <https://doi.org/10.1037/a0012702>

- Noldus (2013). *FaceReader: Tool for automatic analysis of facial expression* (Version 5.0) [Apparatus and software]. Noldus Information Technology B.V.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38, 194–201.
<https://doi.org/10.1006/jesp.2001.1500>
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115, 336–356.
<https://doi.org/10.1037/0033-295X.115.2.336>
- van Breen, J. A., Spears, R., Kuppens, T., & de Lemus, S. (2017). A multiple identity approach to gender: Identification with women, identification with feminists, and their interaction. *Frontiers in Psychology*, 8:1019.
<https://doi.org/10.3389/fpsyg.2017.01019>