RESEARCH ARTICLE

Validating the semantic misattribution procedure as an implicit measure of gender stereotyping

Yang Ye*,‡ & Bertram Gawronski†,‡

* Ghent University, Belgium
† The University of Texas at Austin, USA
‡ The University of Western Ontario, Canada

Correspondence
Yang Ye, Department of Experimental-Clinical and Health Psychology, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium.
E-mail: yang.ye@ugent.be, yang.ye.psych@gmail.com

Received: 2 April 2017
Accepted: 30 August 2017

https://doi.org/10.1002/ejsp.2337

Keywords: gender stereotyping, implicit measures, misattribution, sexism, validity

Abstract

The current research tested the validity of the semantic misattribution procedure (SMP)—a variant of the affect misattribution procedure—as an implicit measure of gender stereotyping. In three studies (N = 604), prime words of gender-stereotypical occupations (e.g., nurse, doctor) influenced participants' guesses of whether unknown Chinese ideographs referred to male or female names in a stereotype-congruent manner. Priming scores of gender stereotyping showed high internal consistency and construct-valid correlations with explicit measures of sexism. Discriminant validity of gender stereotyping scores was tested by investigating relations with priming effects involving grammatical gender (e.g., mother, father). Evidence for discriminant validity was obtained when (1) trials from the two priming measures were presented in a blocked rather than interspersed manner and (2) the measure of stereotypical gender priming preceded the measure of grammatical gender priming. Overall, the SMP showed good psychometric properties and construct validity for the assessment of gender stereotyping.

A man and his son are in a car accident. The man dies instantly; the boy is critically wounded. The ambulance rushes him to the emergency room and the attendants wheel him quickly into the emergency room, on the slim chance he can be saved. The surgeon enters, takes one look at the boy and says, "I can't operate on this child. He's my son." How is this possible?

A few years ago, we presented the above question to undergraduate students in a social psychology class. Many of our students were unable to come up with the answer that seems so obvious in retrospect: the surgeon is the boy's mother. This anecdote illustrates that people are often unaware of how their judgments and behaviors are influenced by gender stereotypes (Deaux & Lewis, 1984; Jost & Kay, 2005).[1] Gender stereotypes can lead to sexism and discrimination against women (Cohen & Bunker, 1975; Glick, Zion, & Nelson, 1988; Heilman, 2001), in that stereotyping can undermine women's achievement in certain professions (e.g., science and math, Nosek et al., 2009) and contribute to gender inequality in various occupations (Cejka & Eagly, 1999; Glick & Fiske, 2001). As illustrated by the above anecdote, effects of gender stereotypes often occur automatically (e.g., Banaji & Greenwald, 1995; Banaji & Hardin, 1996), i.e., without intention, awareness, and in a manner that is difficult to control (Bargh, 1994; Moors & De Houwer, 2006). At the same time, people are often unwilling to overtly endorse gender stereotypes, which would represent a violation of social norms in Western societies. These issues have increased concerns about the usefulness of explicit self-report measures, which in turn contributed to the popularity of implicit measures to assess expressions of gender stereotypes (for a review, Gawronski & De Houwer, 2014).

Although various paradigms have been used to assess gender stereotyping (e.g., Banaji & Greenwald, 1995; Banaji & Hardin, 1996; Blair, Ma, & Lenton, 2001), the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998) is by far the most popular one among the available options. The IAT has been used to measure various types of gender stereotypical associations, including male–science versus female–arts (Nosek, Banaji, & Greenwald, 2002; Nosek et al., 2009; for a review, see Zitelny, Shalom, & Bar-Anan, 2017), male–strength versus female–weakness (Blair et al., 2001; Milne & Grafman, 2001), male–engineer versus female–teacher (White & White, 2006), and male–career versus female–household (Gawronski, Ehrenberg, Banse, Zukova, & Klauer, 2003). IAT-measured math–gender stereotypes were found to

[1]Interestingly, gender stereotypes seem to be even more powerful than stereotypical conceptions of parenthood. Among our students, we found a higher proportion of answers involving a couple of gay men who adopted the boy than answers involving a female doctor.

 Copyright © 2017 John Wiley & Sons, Ltd.

predict academic self-concepts, academic achievement, and enrollment preferences in 9-year old girls (Steffens, Jelenec, & Noack, 2010). Similarly, associating science with male relative to female in the IAT was found to predict gender-differences in 8th grade science and mathematics achievement at the national level across 34 countries (Nosek et al., 2009). Finally, the IAT has been widely used as a dependent measure in experimental studies to identify factors that either enhance or reduce gender stereotyping (e.g., Dasgupta & Asgari, 2004).

Despite the popularity of the IAT, the task has some important limitations. First, although IAT-based measures tend to be superior compared to other implicit measures in terms of their internal consistency (Gawronski & De Houwer, 2014), the task-structure of the IAT has been linked to various sources of systematic measurement error that can undermine its construct validity (Teige-Mocigemba, Klauer, & Sherman, 2010). Second, a substantial portion of studies has focused on the mere identification of a significant IAT effect to infer the prevalence of gender stereotyping (e.g., White & White, 2006) instead of testing construct-valid relations between gender stereotyping and criterion measures (for notable exceptions, see Nosek et al., 2009; Steffens et al., 2010). The lack of basic validation studies has limited the application of IAT-based measures of gender stereotyping and the interpretation of findings obtained in this research.

In the present work, we took a systematic approach to the validation of a new implicit measure of gender stereotyping: the *semantic misattribution procedure* (SMP). The SMP is a modified variant of the affect misattribution procedure (AMP; see Payne, Cheng, Govorun, & Stewart, 2005) designed to assess spontaneous behaviors resulting from the activation of semantic concepts rather than affective states (e.g., Deutsch & Gawronski, 2009; Gawronski & Ye, 2014; Imhoff, Schmidt, Bernhardt, Dierksmeier, & Banse, 2011; Sava et al., 2012). In addition to providing an alternative option for the assessment of gender stereotyping, the SMP has the potential to provide deeper insights for research on gender stereotyping and sexism, as well as for research on implicit measures more broadly.

Semantic Misattribution Procedure

The AMP (Payne et al., 2005), on which the SMP is based, has been widely applied in social psychology, showing good psychometric properties such as large effect sizes, high internal consistency, and evidence for construct validity (for a review, see Payne & Lundberg, 2014). On a typical AMP trial, participants are briefly presented with a prime stimulus (e.g., a positive or a negative image), which is followed by a backward-masked Chinese ideograph. Participants are asked to judge the Chinese ideograph as either visually more pleasant or visually less pleasant than average. The modal finding is that participants' judgments across trials tend to reflect their evaluative responses to the primes, despite explicit instructions not to let the primes influence their judgments of the Chines ideographs.

According to the misattribution account (Gawronski & Ye, 2014; Loersch & Payne, 2011; Payne, Hall, Cameron, & Bishara, 2010), priming effects in the AMP are mediated by the misattribution of spontaneous affective feelings elicited by the primes to ambiguous targets, such as Chinese ideographs. Gawronski and Ye (2014) further demonstrated that this misattribution process can operate not only on affective, but also semantic, attributes of the primes (see also Deutsch & Gawronski, 2009; Imhoff et al., 2011; Sava et al., 2012). In their research, participants were presented with male or female face primes and asked to guess whether the Chinese ideographs referred to a male or female name. The results showed that the gender of the face primes influenced participants' responses to the targets, in that participants were more likely to guess male (as opposed to female) when they were presented with a male face than when they were presented with a female face.

Similar to Gawronski and Ye's (2014) procedure, participants in the current research were asked to guess whether Chinese ideographs refer to male or female names. As primes, we used words representing *gender stereotypical occupations* (e.g., mechanic as a stereotypically male occupation, secretary as a stereotypically female occupation; see Banaji & Hardin, 1996). We predicted that responses in the task would be biased by the gender stereotypical occupations, such that participants would be more likely to respond *male* than *female* when they were primed with a stereotypically male occupation than when they were primed with a stereotypically female occupation, and vice versa. The overall size of such priming effects can be interpreted as a measure of *gender stereotyping*. Here, gender stereotyping refers to a behavioral effect (i.e., influence of primes on judgments of the targets), and is therefore agnostic with regard to the underlying mental representations of gender stereotypes as well as the processes by which these representations influence participants' responses on the task (see De Houwer, Gawronski, & Barnes-Holmes, 2013). Moreover, the SMP can be described as an *implicit measure* in the sense that the nature of the measured construct is implicit in the observed response (i.e., gender stereotyping is inferred from guesses regarding the meaning of Chinese ideographs after exposure to gender stereotypical occupations). In contrast, a measure would be an *explicit measure* to the extent that the nature of the measured construct is explicit in the observed response (e.g., gender stereotyping is inferred from agreement with statements ascribing gender-stereotypical attributes to women and men).

Rationale for Validation

The main goal of the current research was to establish the validity of our SMP-based measure of gender

stereotyping. Toward this end, we tested construct-valid patterns of relations between gender stereotyping in the SMP and explicit measures of sexism. In particular, we looked at three contemporary forms of sexism that have been distinguished from old-fashioned sexism (Swim, Aikin, Hall, & Hunter, 1995).

Benevolent sexism and *hostile sexism* are the two components of Glick and Fiske's (1996) ambivalent sexism inventory. The former is characterized by the endorsement of stereotypical beliefs that women are weak and in need of a male provider's support, while the latter is characterized by sexist antipathy against women in the traditional sense of prejudice. According to Glick and Fiske (1996, 2001), the two types of sexism converge in that both are associated with stereotypes about women. Yet, the two types of sexism diverge in that benevolent sexism involves positive feelings about women, whereas hostile sexism involves negative feelings. The third form of sexism included in the current work is *modern sexism* (Swim et al., 1995), which involves beliefs that discrimination of women is not a problem anymore, accompanied by a lack of support for policies to promote gender equality. According to Swim et al., the endorsement of modern sexism bolsters existing status differences between men and women and the status quo of gender inequality.

Although the three forms of contemporary sexism have been treated as distinct from old-fashioned sexism, they are different in the sense that they may show unique relations to gender stereotyping in the SMP. First, benevolent sexism (Glick & Fiske, 1996, 2001) involves beliefs about gender roles (e.g., 'women should be protected by men') and traits (e.g., 'women are wonderful but weak') that are consistent with the attributes ascribed to gender stereotypical occupations (e.g., doctor–male, nurse–female). Thus, to the extent that gender stereotyping in the SMP is systematically related to these beliefs, it should be positively correlated with benevolent sexism.

For hostile sexism (Glick & Fiske, 1996, 2001), we predicted a more complex relation with the SMP measure of gender stereotyping than that for benevolent sexism. On the one hand, hostile sexism involves negative sentiments toward women and beliefs that support these sentiments (e.g., 'men should be in superior roles, because they are superior'), which tend to be consistent with the attributes ascribed to gender stereotypical occupations. On the other hand, because overt expression of hostility toward women is deemed inappropriate in Western societies, it is likely moderated by self-presentational concerns such as the motivation to control prejudicial and stereotypical responses (e.g., Banse & Gawronski, 2003; Degner & Wentura, 2008; Dunton & Fazio, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Gawronski, Geschke, & Banse, 2003; Payne et al., 2005). On the Basis of these considerations, we predicted that hostile sexism should show a positive relation with gender stereotyping in the SMP when motivation to control is low, but not when motivation to control is high.

Finally, according to Swim et al. (1995), a central component of modern sexism is the belief that discrimination of women is not a problem anymore in the current society. Thus, whereas low levels of modern sexism are associated with perceptions of high discrimination against women, high levels of modern sexism are associated with perceptions of low discrimination against women. This aspect helps to distinguish between gender stereotyping and mere knowledge of gender inequality and discrimination. If the proposed priming effects of gender stereotypical occupations reflect mere knowledge of gender inequality and discrimination in gender-stereotypical occupations (see Eagly & Steffen, 1984), SMP scores should show a *negative* relation with modern sexism. That is, the more gender inequality participants perceive in the occupations presented in the SMP, the lower should be their modern sexism scores (with low scores reflecting acknowledgement of continued discrimination in these occupations). In contrast, if priming effects of gender stereotypical occupations reflect effects of gender stereotyping rather than mere knowledge of gender inequality and discrimination, SMP scores should be either unrelated or positively related to modern sexism.

Study 1

The main goal of Study 1 was to provide preliminary evidence for the construct validity of our SMP-based measure of gender stereotyping. We expected the SMP to show similar estimates of internal consistency as the AMP (see Payne & Lundberg, 2014) and reveal the hypothesized patterns of relations with the three explicit measures of sexism. In particular, we predicted SMP scores to be (1) positively correlated with benevolent sexism, (2) positively correlated with hostile sexism when motivation to control is low but not when it is high, and (3) *not* negatively correlated with modern sexism.

Method

Participants. A total of 200 undergraduates (138 women, 62 men, M_{age} = 18.3 years) at the University of Western Ontario in Canada participated for research credit of an introductory psychology course. Twenty-nine participants (all of Asian background) reported knowing the meaning of the Chinese ideographs that were used as target stimuli in the SMP. Data from these participants were excluded from analyses, resulting in a sample of 171 participants (114 women, 57 men, M_{age} = 18.3 years).

Procedure. Following the default procedure of the AMP (Payne et al., 2005), each SMP trial involved the presentation of a fixation cross for 500 ms, a prime word for 75 ms, a blank screen for 125 ms, a Chinese ideograph for 100 ms, and a black-and-white pattern mask. Participants were asked to guess whether the Chinese ideograph referred to a female

or a male name by pressing a right-hand key (*Numpad 5*) for male names and a left-hand key (*A*) for female names. The inter-trial interval was 1000 ms. Following Payne et al. (2005), participants were told that the words flashed before the Chinese ideographs can sometimes bias people's responses to the Chinese ideographs and that they should try their absolute best not to let the words bias their responses in any possible ways. The prime words included three stereotypically male occupations (i.e., *doctor*, *mechanic*, *engineer*) and three stereotypically female occupations (i.e., *nurse*, *secretary*, *receptionist*). Each prime was presented ten times, summing up to a total of 60 trials. The targets were 60 distinct Chinese ideographs from Payne et al. (2005). Trials were presented in random order, using ten blocks of six trials comprising the six prime words.

After the SMP, participants were asked to complete Glick and Fiske's (1996) scales for hostile and benevolent sexism, Swim et al.'s (1995) modern sexism scale, and a modified variant of Dunton and Fazio's (1997) motivation to control prejudiced reactions scale adapted for sexism. Responses were measured with 6-point rating scales with the end-points − 3 (*disagree strongly*) and +3 (*agree strongly*). Data were numerically recorded from 1 to 6. At the end of the study, participants reported demographic information and whether they know the meaning of any of the Chinese ideographs presented in the SMP.[2]

Results

Following the data analytic procedure for the AMP (Payne et al., 2005), we calculated two SMP scores reflecting the percentage of *male* responses for each of the two kinds of priming trials (i.e., stereotypically male occupation vs. stereotypically female occupation). The scores were then submitted to a repeated-measure ANOVA, which revealed a significant effect of Prime Type, $F(1, 170) = 70.60$, $p < .001$, $\eta_p^2 = .29$. Participants were more likely to guess *male* than *female* when they were primed with stereotypically male occupations ($M = 0.61$, $SD = 0.16$) than when they were primed with stereotypically female occupations ($M = 0.42$, $SD = 0.20$).

Expanding on these analyses, we calculated a single score of *stereotypical gender priming* by subtracting the SMP score from trials with stereotypically female primes from the SMP score from trials with stereotypically male primes. Thus, higher scores on this index reflect higher levels of gender stereotyping. The internal consistency of this score was estimated by calculating one priming score using the 30 trials of the first half of the SMP and another priming score using the 30 trials of the second half of the SMP. The SMP score of stereotypical gender priming showed satisfactory internal consistency with Cronbach's $\alpha = .82$.

As shown in Table 1a, women and men showed similar levels of stereotypical gender priming on the SMP ($p = .31$). In contrast, men showed significantly higher levels of benevolent sexism ($p = .01$), hostile sexism ($p < .001$), and modern sexism ($p = .001$), and significantly lower levels of motivation to control ($p = .003$) than women. Thus, to rule out spurious correlations resulting from gender differences on the four explicit measures, we calculated zero-order correlations (see lower-left triangle, Table 1a) as well as partial correlations (see upper-right triangle, Table 1a) controlling for participant gender (0 = female, 1 = male) for all measures in our study. As predicted, stereotypical gender priming showed a significant positive correlation with benevolent sexism regardless of whether correlations were analyzed at the zero-order level ($p = .006$) or at the level of partial correlations controlling for participant gender ($p = .009$). There was no significant correlation between stereotypical gender priming and modern sexism at the zero-order level ($p = .17$) and at the level of partial correlations controlling for participant gender ($p = .17$).

To test the predicted moderation of the relation between gender stereotypical priming and hostile sexism by motivation to control, we regressed standardized scores of hostile sexism onto standardized scores of stereotypical gender priming, motivation to control, and their interaction, controlling for participant gender (0 = female, 1 = male). The results showed that the product term representing the two-way interaction was a significant predictor of hostile sexism (Table 2). Further analyses revealed that, as predicted, stereotypical gender priming was a significant predictor of hostile sexism when motivation to control was low, $B = 0.30$, $SE = 0.10$, $t(167) = 2.99$, $p = .003$, but not when motivation to control was high, $B = -0.05$, $SE = 0.10$, $t(167) = -0.50$, $p = .62$ (Figure 1).

Discussion

SMP scores of gender stereotyping showed a large effect size in the predicted direction, satisfactory internal consistency, and construct-valid relations with explicit measures of sexism. Specifically, gender stereotypical priming showed a positive correlation with benevolent sexism and a positive relation with hostile sexism when motivation to control was low but not when it was high. The priming score did not show a negative correlation with modern sexism, suggesting that the SMP captures gender stereotyping rather than mere knowledge of gender inequality.

Study 2

The main goals of Study 2 were to (1) replicate the main findings of Study 1 and (2) test the discriminant validity

[2]The experimental materials, data, and analysis codes for all studies reported in this article are available at https://osf.io/vg9wf/.

Table 1 Correlations and descriptive statistics of stereotypical gender priming (SMP_s), grammatical gender priming (SMP_g), benevolent sexism (BEN), hostile sexism (HOS), modern sexism (MSS), and motivation to control prejudiced reactions (MCPR), Studies 1–3

	1.	*2.*	*3.*	*4.*	*5.*	
a. Results from Study 1						
1. SMP_s	*.82*	.20**	.11	.05	.14	
2. BEN	.21**	*.76*	.37***	−.02	−.04	
3. HOS	.13	.41***	*.83*	.09	−.18*	
4. MSS	.07	.03	.17*	*.74*	−.08	
5. MCPR	.12	−.09	−.25***	−.13	*.73*	
M_{women} (n = 114)	0.18 (0.29)	3.52 (0.73)	3.14 (0.82)	3.09 (0.68)	3.70 (0.62)	
M_{men} (n = 57)	0.23 (0.32)	3.84 (0.81)	3.79 (0.82)	3.49(0.83)	3.41 (0.49)	
F_{gender}(1, 169)	1.05	6.77**	24.19***	11.22***	9.13**	
	1.	*2.*	*3.*	*4.*	*5.*	*6.*
b. Results from Study 2						
1. SMP_s	*.86*	.77***	.16*	.20*	.03	.04
2. SMP_g	.77***	*.80*	.18*	.10	−.02	.03
3. BEN	.17*	.18*	*.81*	.23**	.11	.10
4. HOS	.21*	.10	.24**	*.82*	.26***	−.18*
5. MSS	.05	−.02	.13	.29***	*.80*	−.15
6. MCPR	.04	.03	.09	−.18*	−.16*	*.76*
M_{women} (n = 119)	0.29 (0.42)	0.36 (0.35)	3.36 (0.91)	3.16 (0.80)	2.80 (0.81)	3.64 (0.67)
M_{men} (n = 46)	0.39 (0.38)	0.37 (0.40)	3.58 (0.75)	3.49 (1.03)	3.23 (0.88)	3.55 (0.52)
F_{gender}(1, 163)	1.80	0.01	2.18	4.78*	8.88**	0.62
c. Results from the grammatical priming first condition, Study 3						
1. SMP_s	*.82*	.64***	.16	.14	.15	−.07
2. SMP_g	.64***	*.85*	.23**	.19*	.18*	−.14
3. BEN	.16	.23**	*.89*	.28**	.24**	.02
4. HOS	.15	.17	.28**	*.93*	.72***	−.33***
5. MSS	.16	.16	.25**	.77***	*.89*	−.43***
6. MCPR	−.08	−.12	−.004	−.43***	−.51***	*.79*
M_{women} (n = 40)	0.24 (0.35)	0.44 (0.38)	2.97 (1.23)	2.33 (1.01)	2.29 (1.03)	3.87 (0.62)
M_{men} (n = 90)	0.29 (0.38)	0.42 (0.44)	3.12 (1.03)	3.29 (1.15)	3.27 (1.07)	3.30 (0.64)
F_{gender}(1, 128)	0.42	0.07	0.53	20.74***	23.90***	21.61***
d. Results from the stereotypical priming first condition, Study 3						
1. SMP_s	*.85*	.53***	.15	.28***	.32***	−.27**
2. SMP_g	.53***	*.72*	.004	.15	.21*	−.16
3. BEN	.18*	.01	*.88*	.35***	.31***	−.10
4. HOS	.31***	.15	.41***	*.94*	.68***	−.45***
5. MSS	.34***	.21*	.35***	.70***	*.92*	−.44***
6. MCPR	−.28**	−.16	−.12	−.46***	−.46***	*.82*
M_{women} (n = 56)	0.13 (0.42)	0.23 (0.40)	2.62 (1.09)	2.26 (1.13)	2.33 (1.35)	3.66 (0.82)
M_{men} (n = 82)	0.23 (0.36)	0.25 (0.30)	3.12 (0.94)	3.20 (1.11)	2.97 (1.08)	3.48 (0.68)
F_{gender}(1, 136)	2.32	0.16	8.49**	23.28***	9.50**	1.96

Note: Standard deviations are presented in parentheses; Cronbach's αs (in italics) are presented on the diagonal of the correlation matrices; Zero-order correlation coefficients are displayed on the lower-left side of the diagonal; Partial correlation coefficients controlling for participant gender (0 = female, 1 = male) are displayed on the upper-right side of the diagonal.
*$p < .05$,
**$p < .01$,
***$p < .001$.

of gender stereotyping in the SMP. Toward this end, Study 2 used the same procedure and materials as Study 1 while additionally including SMP trials with grammatically male or female words as primes (e.g., king, queen). Conceptually, priming effects on these kinds of trials do not reflect gender stereotyping, as the gender connotations of the primes are determined by grammatical rather than stereotypical features. To distinguish between the two kinds of priming effects, we use the term *stereotypical gender priming* to refer to priming effects of gender stereotypical occupations and the term *grammatical gender priming* to refer to priming effects of words involving grammatical gender. It is important to distinguish between the two kinds of priming effects, because only the former, but not the latter, reflects genuine effects of gender stereotyping. Thus, any overlap between the two scores should reflect measurement variance rather than construct variance in the degree of gender stereotyping (see Wentura & Degner, 2010). Assuming that our findings reflect genuine effects of gender stereotyping (rather than stereotype-independent tendencies to show a gender priming effect

Table 2 Results of regression analyses regressing hostile sexism onto participant gender (0 = female, 1 = male), stereotypical gender priming (SMP_s), grammatical gender priming (SMP_g), motivation to control prejudiced reaction (MCPR), and their interactions, Studies 1–3

			Study 3	
	Study 1	Study 2	Grammatical priming first	Stereotypical priming first
Model 1				
Intercept	−0.18*	−0.08	−0.34*	−0.41***
Gender	0.61***	0.29	0.51***	0.62***
SMP_s	0.12	0.21**	0.12	0.15*
MCPR	−0.19**	−0.21**	−0.33***	−0.36***
SMP_s × MCPR	−0.17*	−0.15$^+$	0.05	−0.15*
Model 2				
Intercept		−0.09	−0.35*	−0.42***
Gender		0.34*	0.54***	0.68***
SMP_g		0.11	0.14	0.07
MCPR		−0.19*	−0.32***	−0.39***
SMP_g × MCPR		−0.14$^+$	0.06	−0.14*
Model 3				
Intercept		−0.07	−0.35*	−0.41***
Gender		0.27	.53***	.63***
SMP_s		0.31**	.04	.15$^+$
SMP_g		−0.13	.12	−0.01
MCPR		−0.21**	−0.31***	−0.36***
SMP_s × MCPR		−0.11	.11	−0.10
SMP_g × MCPR		−0.05	.06	−0.07

Note: Results from regression analyses with benevolent and modern sexism measures as dependent measures showed no moderating effect of MCPR. The results of these analyses can be either reproduced using the data and analysis codes that are available in the archive, or are available from the authors upon request.
*p < .05;
**p < .01;
***p < .001;
$^+$p = .07.

in the SMP), we expected the scores from the two measures to diverge. In particular, we predicted that only stereotypical gender priming scores, but not grammatical gender priming scores, would show the predicted relations with explicit measures of sexism. Moreover, stereotypical gender priming should still show the predicted relations with explicit measures of sexism after controlling for grammatical gender priming.

Method

Participants. A total of 194 undergraduates (140 women, 54 men, M_{age} = 21.5 years) at the University of Western Ontario in Canada participated for research credit for an introductory psychology course. The study used a 2 (Prime Gender: male vs. female) × 2 (Prime Type: stereotypical vs. grammatical) within-subjects design. Twenty-nine participants (27 with Asian background) reported knowing the meaning of the Chinese ideographs used as target stimuli. Data from these participants were excluded from analyses, leaving us with a sample of 165 participants (119 women, 46 men, M_{age} = 21.6 years).

Procedure. Participants completed an SMP task that included two types of trials: (1) *stereotypical priming* trials with the six stereotypically male or female occupations from Study 1 as primes, and (2) *grammatical priming* trials with six grammatically male (i.e., *king*, *waiter*, *father*) or grammatically female (*i.e.*, *queen*, *waitress*, *mother*) nouns as primes. Three neutral words (i.e., *door*, *coffee*, *table*) were included as filler primes.[3] The procedural details of the SMP were identical to Study 1. Each of the 15 primes was presented six times, summing up to a total of 90 trials. The target stimuli were 90 distinct Chinese ideographs from Payne et al.'s (2005) AMP. Trials were presented in random order, using six blocks of 15 trials comprising the 15 prime words. After the SMP, participants were asked to complete the explicit measures as in Study 1. At the end of the study, participants reported demographic information and whether they knew the meaning of any of the Chinese ideographs presented in the SMP.

Results

Means. Following the procedures in Study 1, we calculated four SMP scores reflecting the percentage of *male* responses for each of the four kinds of priming trials: stereotypically male, stereotypically female, grammatically male, and grammatically female. The resulting scores were submitted to a 2 (Prime Gender) × 2 (Prime Type) repeated-measures ANOVA, which revealed a significant main effect of Prime Gender, $F(1, 164) = 142.74$, $p < .001$, $\eta_p^2 = .465$. Moreover, the analysis revealed a significant two-way interaction

[3]Responses on the control trials were not included in the analyses.

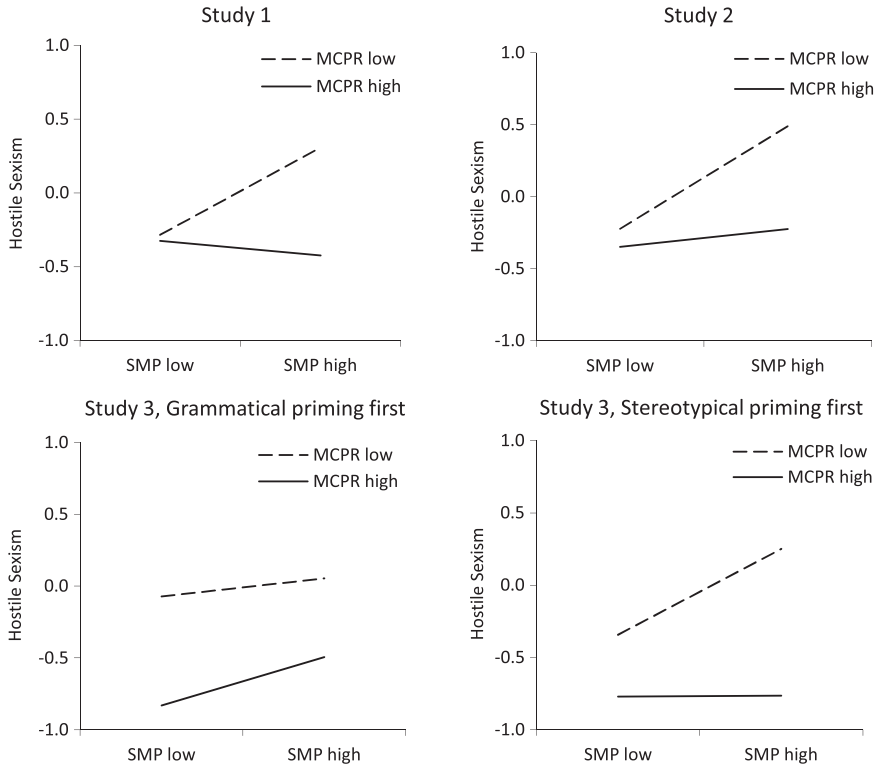

Fig. 1: Hostile sexism as a function of stereotypical gender priming (SMP), moderated by motivation to control prejudiced reactions (MCPR), Studies 1–3. SMP low = −1 *SD*; SMP high = +1 *SD*; MCPR low = −1 *SD*; MCPR high = +1 *SD*

between Prime Gender and Prime Type, $F(1, 164) = 4.74$, $p = .031$, $\eta_p^2 = .028$, suggesting a larger priming effect of grammatical gender primes than stereotypical gender primes. Separate analyses for the two types of primes revealed that participants were more likely to guess *male* when they were primed with a stereotypically male occupation ($M = 0.69$, $SD = 0.22$) than when they were primed with a stereotypically female occupation ($M = 0.37$, $SD = 0.24$), $F(1, 164) = 96.85$, $p < .001$, $\eta_p^2 = .371$. Moreover, participants were more likely to guess *male* when they were primed with a grammatically male noun ($M = 0.70$, $SD = 0.18$) than when they were primed with a grammatical female noun ($M = 0.34$, $SD = 0.24$), $F(1, 164) = 164.51$, $p < .001$, $\eta_p^2 = .501$.

Priming scores of *stereotypical gender priming* and *grammatical gender priming* were calculated in line with the procedures in Study 1. The SMP score of stereotypical gender priming showed satisfactory internal consistency with Cronbach's $\alpha = .86$; the SMP score of grammatical gender priming had an internal consistency of Cronbach's $\alpha = .80$. Table 1b presents the descriptive statistics for the two priming scores and the explicit measures of sexism by participant gender. Comparisons of means revealed that men showed significantly higher levels of hostile sexism ($p = .03$) and modern sexism ($p = .003$) than women. Yet, men and women did not significant differ in terms of stereotypical gender priming ($p = .18$), grammatical gender priming ($p = .92$), benevolent sexism ($p = .14$), and motivation to control ($p = .43$).

Correlations. Table 1b presents the zero-order correlations (see lower-left triangle) and partial correlations (see upper-right triangle) controlling for participant gender (0 = female, 1 = male) for all measures in the study. The results showed a significant correlation between stereotypical and grammatical gender priming at the level of zero-order correlations as well as at the level of partial correlations controlling for participant gender (ps $< .001$). Moreover, both priming scores showed significant zero-order correlations with benevolent sexism (ps = .03 and .02, respectively), which remained significant after controlling for participant gender (ps = .04 and .02, respectively). Neither priming score showed a significant zero-order correlation with modern sexism (ps = .53 and .80, respectively) and motivation to control (ps = .66 and .73, respectively). A divergence between the two priming scores was observed only for hostile sexism: Whereas stereotypical gender priming and hostile sexism showed a significant zero-order correlation ($p = .006$) and partial correlation controlling for participant gender ($p = .01$), grammatical gender priming and hostile sexism were not significantly related at the level of zero-order correlations ($p = .22$) and partial correlations ($p = .22$).

To further examine the relations between gender priming and explicit measures of sexism, we calculated partial correlations between each priming index and the three explicit measures controlling for the respective other priming index and participant gender (see Table 3). The results showed that, after controlling for grammatical gender priming and participant gender,

stereotypical gender priming still showed a significant positive correlation with hostile sexism (p = .01). The relation to modern sexism remained non-significant (p = .39). However, the significant zero-order correlation with benevolent sexism became non-significant (p = .71). When controlling for stereotypical gender priming, grammatical gender priming was still not significantly correlated with hostile sexism (p = .25) and modern sexism (p = .40). Yet, similar to stereotypical gender priming, its significant zero-order correlation with benevolent sexism became non-significant (p = .23).

Regression. To test the moderating effect of motivation to control on the relation between each priming index and hostile sexism, we submitted each of the two priming indices to the regression model of Study 1 (see Table 2). Replicating the findings of Study 1, hostile sexism was predicted by a marginally significant two-way interaction between stereotypical gender priming and motivation to control (p = .06). Further analyses revealed a significant positive relation between stereotypical gender priming and hostile sexism when motivation to control was low (at −1 *SD*), B = 0.36, SE = 0.11, $t(161) = 3.25$, $p = .001$, but not when motivation to control was high (at +1 *SD*), B = 0.06, SE = 0.11, $t(161) = 0.57$, p = .57 (see Figure 1). A similar pattern emerged for grammatical gender priming, which revealed a marginally significant two-way interaction between grammatical gender priming and motivation to control in the prediction of hostile sexism (p = .07). Further analyses showed that grammatical gender priming predicted hostile sexism when motivation to control was low (at −1 *SD*), B = 0.25, SE = 0.11, $t(161) = 2.28$, p = .02, but not when motivation to control was high (at +1 *SD*), B = −0.03, SE = 0.11, $t(161) = -0.26$, p = .80.

To further explore potential redundancies between the two priming scores, we regressed hostile sexism simultaneously onto the two priming indices, motivation to control, and the two product terms representing the two-way interactions between motivation to control and each priming index, controlling for participant gender (see Table 2). In this model, stereotypical gender priming was a significant positive predictor (p = .009), while grammatical gender priming was a non-significant predictor (p = .26). Yet, the two-way interaction between stereotypical gender priming and motivation to control was non-significant (p = .42). The same was true for the two-way interaction between grammatical gender priming and motivation to control (p = .71).

Table 3 Partial correlations between stereotypical gender priming (SMP_s) or grammatical gender priming (SMP_g) with benevolent sexism (BEN), hostile sexism (HOS), modern sexism (MSS), and motivation to control prejudiced reaction (MCPR) controlling for participant gender and the respective other priming score, Studies 2 and 3

	Study 2		Study 3			
			Grammatical priming first		Stereotypical priming first	
	SMP_s	SMP_g	SMP_s	SMP_g	SMP_s	SMP_g
BEN	.03	.09	.02	.17	.17*	−.09
HOS	.20*	−.09	.02	.13	.24**	.00
MSS	.07	−.07	.05	.11	.26**	.04
MCPR	.03	−.01	.03	−.13	−.22*	−.02

*$p < .05$,
**$p < .01$,
***$p < .001$.

Discussion

The main goals of Study 2 were to (1) replicate the main findings of Study 1 and (2) test the discriminant validity of gender stereotyping in the SMP. Toward this end, Study 2 used the same procedure and materials as Study 1, additionally including SMP trials with grammatically male or female words as primes. Although Study 2 successfully replicated the main findings of Study 1, the findings were mixed with regard to the discriminant validity between stereotypical gender priming and grammatical gender priming. On the one hand, the two priming scores were highly correlated, revealed similar patterns of relations with benevolent sexism and modern sexism, and revealed the same interaction with motivation to control in the prediction of hostile sexism. These results suggest that our findings with the SMP might be driven by stereotype-unrelated measurement variance rather than genuine effects of gender stereotyping. On the other hand, there was some evidence for discriminant validity, in that stereotypical gender priming showed a significant positive relation with hostile sexism after controlling for grammatical gender priming, although this relation was no longer moderated by motivation to control. In contrast, when controlling for stereotypical gender priming, grammatical gender priming showed no significant relation with hostile sexism.

The obtained overlap between the two kinds of priming effects poses a challenge to the SMP as a measure of gender stereotyping, because it suggests that our findings reflect stereotype-independent tendencies to show a gender priming effect in the SMP rather than genuine effects of gender stereotyping. However, before drawing premature conclusions, it seems important to rule out alternative explanations for the observed level of overlap between the two priming scores. One potential factor might be the interspersed presentation of the two kinds of priming trials, which may direct participants' attention to common features of the two kinds of primes (see Gawronski & Ye, 2015). Specifically, it is possible that exposure to words with a grammatical gender (e.g., king, queen) directs attention to gender as a salient feature, which may enhance gender stereotypical interpretations of otherwise gender-neutral occupations (e.g., doctor, nurse). Such spillover effects could lead to inflated levels of overlap between the two kinds of priming scores, and thereby attenuate their discriminant validity. Study 3 aimed to test this possibility.

Study 3

The main goal of Study 3 was to re-examine the discriminant validity of our gender stereotyping SMP measure controlling for potential artifacts from the interspersed presentation of the two kinds of priming trials in Study 2 (i.e., stereotypical gender primes vs. grammatical gender primes). Toward this end, we presented the two kinds of priming trials in two separate blocks. Counter-balancing the order of the two blocks also allowed us to directly test the specified hypothesis of potential spillover effects. To the extent that the overlap between the two kinds of priming scores was inflated by the presentation of words with a grammatical gender, the observed overlap between the two priming scores should be higher when the grammatical gender priming measure is completed before the stereotypical gender priming measure than when it is completed after the stereotypical gender priming measure. Thus, the stereotypical gender priming measure should show higher levels of construct validity when it is completed before the grammatical gender priming measure than when it is completed after the grammatical gender priming measure.

Method

Participants. Participants were recruited through the online platform Prolific. Of the 285 participants who started the study, 272 completed all tasks of the study and received £1 as compensation (96 women, 174 men, 2 unspecified; M_{age} = 33.1 years; 270 reported being U.S. residents; 265 reported English as first language). At the end of the study, participants were asked to report their knowledge of the meanings of the Chinese ideographs by selecting one of four response options: *No, I do not know any* (n = 253), *Yes, but only a few* (n = 17), *Yes, quite a lot* (n = 1), and *Yes, almost all of them* (n = 1). Two participants who selected *Yes, quite a lot* or *Yes, almost all of them* were excluded from the analyses.[4] Two participants with unspecified gender were also excluded from the analyses, leaving us with a sample of 268 participants (96 women, 172 men, M_{age} = 33.2 years). The study used a 2 (Prime Gender: male vs. female, within Ss) × 2 (Prime Type: stereotypical vs. grammatical, within Ss) × 2 (Block Order: stereotypical priming first vs. grammatical priming first, between Ss) mixed design. Participants were randomly assigned to one of the two between-subjects conditions of Block Order.

Procedure. The SMP was similar to the one in Study 2, except that grammatical priming trials and stereotypical priming trials were presented in two separate blocks instead of in a single block. For the sake of consistency, the three neutral primes (i.e., *door, coffee, table*) from Study 2 were included in both blocks. In each block, each prime was presented six times, totaling 54 trials per block. For the target stimuli, we used 100 distinct Chinese ideographs from Payne et al. (2005). Within each block, trials were presented in random order. Because many keyboards do not have a separate number pad, participants were asked to press the *E* key for female names and *I* key for male names. The instructions for the SMP were streamlined for online testing.

After the SMP, participants were asked to complete the explicit measures as in Study 1. At the end of the study, participants reported demographic information and whether they know the meaning of any of the Chinese ideographs presented in the SMP. For exploratory purposes, participants were additionally asked if they would vote for Hillary Clinton or Donald Trump in the 2016 U.S. presidential election if they had to choose between the two candidates. The study was conducted in August 2016, approximately two months before the 2016 U.S. presidential election.

Participants completed all tasks online using the program Inquisit 5 Web by Millisecond Software. After providing their informed consent, participants were redirected from a webpage at Prolific to a webpage at Millisecond, where they were asked to download the software used to run the study. The program started automatically after the download was completed. Participants completed the tasks using their own computers. After completion of the study, participants were re-directed to a webpage at Prolific to receive their payment.

Results

Means. As in Study 2, we calculated four SMP scores reflecting the percentage of *male* responses for each of the four kinds of priming trials (i.e., stereotypically male, stereotypically female, grammatically male, grammatically female), which were submitted to a 2 (Prime Gender) × 2 (Prime Type) × 2 (Block Order) mixed ANOVA. The analysis revealed a significant main effect of Prime Gender, $F(1, 266) = 184.86$, $p < .001$, $\eta_p^2 = .41$, a significant two-way interaction between Prime Gender and Order of Measure, $F(1, 266) = 10.46$, $p = .001$, $\eta_p^2 = .038$, a significant two-way interaction between Prime Gender and Prime Type, $F(1, 266) = 22.72$, $p < .001$, $\eta_p^2 = .079$, and a significant three-way interaction between all three factors, $F(1, 266) = 5.90$, $p = .016$, $\eta_p^2 = .022$.

To decompose this three-way interaction, we calculated scores of grammatical and stereotypical gender priming following the procedure in Study 2. In the two Block Order conditions (stereotypical priming first vs. grammatical priming first), the SMP score of stereotypical gender priming showed satisfactory internal consistencies with Cronbach's α = .82 and .85, respectively, and the SMP score of grammatical gender priming showed internal consistencies of Cronbach's α = .85 and .72, respectively. In all four conditions

[4]Excluding the data from the 17 participants who reported knowing the meanings of "only a few" Chinese ideographs did not change the reported findings in terms of statistical significance.

defined by Prime Type and Block Order, cell means of the priming scores were significantly larger than zero (all $ps < .001$), indicating that participants were more likely to guess *male* on trials with "male" primes than on trials with "female" primes. The two types of priming scores were submitted to a 2 (Prime Type) × 2 (Block Order) mixed ANOVA. The two-way interaction between Prime Type and Block Order was statistically identical to the three-way interaction in the previous analysis, $F(1, 266) = 5.90$, $p = .016$, $\eta_p^2 = .022$. Further analyses revealed that in the grammatical priming first condition, scores of grammatical gender priming ($M = 0.42$, $SD = 0.42$) were significantly larger than scores of stereotypical gender priming ($M = 0.27$, $SD = 0.37$), $F(1, 266) = 25.13$, $p < .001$, $\eta_p^2 = .086$. In the stereotypical priming first condition, scores of grammatical gender priming ($M = 0.24$, $SD = 0.34$) were only marginally larger than scores of stereotypical gender priming ($M = 0.19$, $SD = 0.39$), $F(1, 266) = 2.83$, $p = .095$, $\eta_p^2 = .01$.

We further examined gender differences in the two priming scores and explicit measures of sexism in each Block Order condition (see Table 1c and Table 1d).[5] Scores of stereotypical and grammatical gender priming did not differ between men and women (in the grammatical priming first condition, $ps = .52$ and .79, respectively; in the stereotypical priming first condition, $ps = .13$ and .69, respectively). In the grammatical priming first condition, men reported similar levels of benevolent sexism ($p = .47$), higher levels of hostile sexism ($p < .001$) and modern sexism ($p < .001$), and lower levels of motivation to control ($p < .001$) than women. In the stereotypical priming first condition, men reported higher levels of benevolent sexism ($p = .004$), hostile sexism ($p < .001$) and modern sexism ($p = .002$), and similar levels of motivation to control ($p = .16$) than women.

Correlations. Results from the grammatical priming first condition ($n = 130$) are reported in Table 1c and those from the stereotypical priming first condition ($n = 138$) are reported in Table 1d. Zero-order correlations and partial correlations controlling for participant gender (female = 0, male = 1) are reported in the lower-left and upper-right triangles of the two tables, respectively.

First, the two priming scores showed zero-order correlations of $r = .64$ ($p < .001$) in the grammatical priming first condition (Table 1c) and $r = .53$ ($p < .001$) in the stereotypical priming first condition (Table 1d). Although the two correlations were not significantly different, $z = 1.36$, $p = .17$, both were significantly lower than the correlation of $r = .77$ in Study 2, $z = -2.22$, $p = .02$ in the grammatical priming first condition and $z = -3.69$, $p < .001$ in the stereotypical priming first condition.

In the grammatical priming first condition (Table 1c), stereotypical gender priming showed a marginal positive correlation with benevolent sexism ($p = .06$), hostile sexism ($p = .09$), and modern sexism ($p = .07$), and a non-significant correlation with MCPR ($p = .34$). Grammatical gender priming showed significant positive correlations with benevolent sexism ($p = .007$) and marginal positive correlations with hostile sexism ($p = .05$) and modern sexism ($p = .07$), and a non-significant correlation with motivation to control ($p = .17$).

In the stereotypical priming first condition (Table 1d), stereotypical gender priming showed a significant positive correlation with benevolent sexism ($p = .04$), replicating corresponding findings in Studies 1 and 2. There was also a significant positive correlation with hostile sexism ($p < .001$), replicating corresponding findings in Study 2. Unlike the findings in Study 1 and 2, stereotypical gender priming showed a significant positive correlation with modern sexism ($p < .001$) and a significant negative correlation with motivation to control ($p = .001$). In contrast, grammatical gender priming showed a non-significant correlation with benevolent sexism ($p = .88$), a marginal positive correlation with hostile sexism ($p = .07$), a significant positive correlation with modern sexism ($p = .01$), and a marginal negative correlation with motivation to control ($p = .06$).[6]

As in Study 2, we further calculated partial correlations between each priming score and the four explicit measures controlling for the respective other priming score and participant gender. Table 3 shows that, in the grammatical priming first condition, stereotypical gender priming showed non-significant correlations with all four explicit measures (all $ps > .73$) after controlling for grammatical gender priming and participant gender. Grammatical gender priming showed a marginal positive correlation with benevolent sexism ($p = .053$) and non-significant correlations with the other explicit measures (all $ps > .13$) after controlling for stereotypical gender priming and participant gender. In contrast, in the stereotypical priming first condition, stereotypical gender priming showed significant positive correlations with benevolent sexism ($p = .04$), hostile sexism ($p = .004$), and modern sexism ($p = .003$), and a significant negative correlation with motivation to control ($p = .01$) after controlling for grammatical gender priming and participant gender. Grammatical gender priming, in contrast, was not significantly correlated with any of the explicit measures

[5]An omnibus ANOVA with Block Order and participant gender as independent variables revealed significant main effects of participant gender for benevolent sexism ($p = .015$), hostile sexism ($p < .001$), modern sexism ($p < .001$), and motivation to control ($p < .001$), and non-significant main effects of participant gender for stereotypical gender priming ($p = .13$) and grammatical gender priming ($p = .98$). No significant interactions between Block Order and participant gender were found (all $ps > .19$), except for motivation to control ($p = .032$).

[6]Results of partial correlations showed similar patterns as zero-order correlations in both Block Order conditions. Although the significance levels of correlation coefficients differed between the two Block Order conditions, tests of differences on the same types of correlation coefficients between the two conditions revealed no significant differences (all $ps > .05$).

(all $ps > .30$) after controlling for stereotypical gender priming and participant gender.

Regression. We first regressed standardized scores of hostile sexism onto standardized scores of each priming index, motivation to control, Block Order (0 = grammatical priming first, 1 = stereotypical priming first) and all possible interactions between these three variables, controlling for participant gender. For stereotypical gender priming, we found a three-way interaction between stereotypical gender priming, motivation to control, and Block Order, $B = -0.20$, $SE = 0.10$, $t(267) = -1.96$, $p = .05$, indicating that the two-way interaction between stereotypical gender priming and motivation to control was qualified by Block Order. For grammatical gender priming, we also found a significant three-way interaction between grammatical gender priming, motivation to control, and Block Order, $B = -0.20$, $SE = 0.10$, $t(267) = -2.00$, $p = .05$, suggesting that the two-way interaction between grammatical gender priming and motivation to control was also qualified by Block Order.

In the second step, we regressed, in each Block Order condition, hostile sexism onto each priming index, motivation to control, and their two-way interaction while controlling for participant gender (see Table 2). For stereotypical gender priming, the two-way interaction between stereotypical gender priming and motivation to control was not significant in the grammatical priming first condition ($p = .49$), but significant in the stereotypical priming first condition ($p = .03$). Further analyses revealed that, in the stereotypical priming first condition, stereotypical gender priming significantly was positively related to hostile sexism when motivation to control was low (at -1 *SD*), $B = 0.30$, $SE = 0.10$, $t(134) = 2.97$, $p = .004$, but not when motivation to control was high (at +1 *SD*), $B = 0.003$, $SE = 0.10$, $t(134) = 0.03$, $p = .98$ (see Figure 1). In the grammatical priming first condition, stereotypical gender priming did not significantly predict hostile sexism either when motivation to control was low (at -1 *SD*), $B = 0.06$, $SE = 0.10$, $t(126) = 0.63$, $p = .53$, or when motivation to control was high (at +1 *SD*), $B = 0.17$, $SE = 0.12$, $t(126) = 1.43$, $p = .16$ (see Figure 1).

For grammatical gender priming, the two-way interaction between the grammatical gender priming and motivation to control was a non-significant predictor of hostile sexism in the grammatical priming first condition ($p = .42$), but a significant predictor in the stereotypical priming first condition ($p = .030$). Further analyses revealed that, in the stereotypical priming first condition, grammatical gender priming was positively related to hostile sexism when motivation to control was low (-1 *SD*), $B = 0.22$, $SE = 0.09$, $t(134) = 2.29$, $p = .02$, but not when it was high (+1 *SD*), $B = -0.07$, $SE = 0.10$, $t(134) = -0.71$, $p = .48$. In the grammatical priming first condition, grammatical gender priming did not significantly predict hostile sexism when motivation to control was low (at -1 *SD*), $B = 0.08$, $SE = 0.10$, $t(126) = 0.78$, $p = .44$, and marginally predicted hostile sexism when motivation to control was high (at +1 *SD*), $B = 0.20$, $SE = 0.12$, $t(126) = 1.74$, $p = .08$.

In the third and final step, we regressed, in each Block Order condition, hostile sexism simultaneously on both priming indices, motivation to control, and two two-way interaction terms representing the interactions between each priming index and motivation to control, while controlling for participant gender (see Table 2). In the grammatical priming first condition, none of the two priming indices, or their interactions with motivation to control, were significant predictors of hostile sexism ($p > .26$). In the stereotypical priming first condition, stereotypical gender priming was a marginally significant predictor of hostile sexism ($p = .07$). Grammatical gender priming did not significantly predict hostile sexism ($p = .90$). The two-way interactions between stereotypical gender priming and motivation to control ($p = .24$) and between grammatical gender priming and motivation to control ($p = .36$) were non-significant.

Voting intention. Of the 268 participants included in the analysis, 183 (103 and 80 in the stereotypical and grammatical priming first conditions, respectively) reported that they would vote for Hillary Clinton and 85 (35 and 50 in the stereotypical and grammatical priming first conditions, respectively) reported that they would vote for Donald Trump in the 2016 U.S. presidential election. The responses were transformed into a dummy-coded variable of voting intention (0 = voting for Hillary Clinton, 1 = voting for Donald Trump).

To examine whether the two priming scores predict voting intentions, we first regressed voting intention on stereotypical gender priming and grammatical gender priming, Block Order (0 = grammatical priming first, 1 = stereotypical priming first), and the two-way interactions between each priming index and Block Order, controlling for participant gender (0 = female, 1 = male). The analysis revealed a significant effect of Block Order, $B = -0.67$, $SE = 0.28$, Wald = 5.63, $p = .02$, $OR = 0.51$, a marginally significant effect of participant gender, $B = 0.57$, $SE = 0.30$, Wald = 3.57, $p = .06$, $OR = 1.76$, and more importantly, a marginally significant interaction between stereotypical gender priming and Block Order, $B = 0.67$, $SE = 0.36$, Wald = 3.47, $p = .06$, $OR = 1.95$. No other significant effects were found (all $ps > .36$).

To specify this interaction, we regressed voting intention onto stereotypical gender priming, grammatical gender priming, and participant gender within each Block Order condition. In the grammatical priming first condition, stereotypical gender priming was not a significant predictor of voting intention, $B = -0.22$, $SE = 0.24$, Wald = 0.80, $p = .37$, $OR = 0.81$. The same was true for grammatical gender priming, $B = -0.003$, $SE = 0.24$, Wald = 0.00, $p = .99$, $OR = 1.00$, and participant gender, $B = 0.70$, $SE = 0.42$, Wald = 2.79, $p = .10$, $OR = 2.01$. In the stereotypical priming first condition, in contrast, stereotypical gender priming positively predicted voting intention, $B = 0.67$, $SE = 0.26$, Wald = 6.55, $p = .01$,

OR = 1.96, while grammatical gender priming, *B* = −0.007, *SE* = 0.25, Wald = 0.001, *p* = .98, *OR* = 0.99, and participant gender, *B* = 0.42, *SE* = 0.43, Wald = 0.96, *p* = .33, *OR* = 1.53, did not predict voting intention.[7]

Discussion

The main goal of Study 3 was to re-examine the overlap between stereotypical gender priming and grammatical gender priming in the SMP by controlling for potential artifacts from the interspersed presentation of the two kinds of priming trials in Study 2. Toward this end, we presented stereotypical gender primes and grammatical gender primes in two separate blocks in counter-balanced order. The two kinds of priming effects still showed considerable overlap, suggesting that a significant portion of variance in the stereotypical gender priming measure reflects stereotype-independent measurement factors. Yet, overlap between the two kinds of priming effects was much lower with the blocked design of Study 3 compared to the interspersed design in Study 2, suggesting that the unexpectedly high correlation between priming scores was inflated by the interspersed presentation of priming trials.

Although the order of the two SMPs did not affect the zero-order correlation between priming effects in the two tasks, Study 3 revealed some evidence that task order influenced the construct validity of stereotypical gender priming scores. Consistent with our hypothesis of spillover effects from the grammatical gender SMP, construct and discriminant validity of the stereotypical gender SMP was higher when participants completed the stereotypical gender SMP before the grammatical gender SMP.

Specifically, when participants completed the stereotypical gender SMP first, stereotypical gender priming showed construct-valid relations with benevolent sexism and hostile sexism, which remained significant after controlling for grammatical gender priming. Although grammatical gender priming showed similar patterns of relations with the two explicit measures, these relations did not hold after controlling for stereotypical gender priming. Moreover, regression analyses for hostile sexism replicated the findings of Study 2. When stereotypical and grammatical gender priming scores were examined in separate models, the relation of each priming score with hostile sexism was moderated by motivation to control. Yet, when both priming scores were included in the same model, stereotypical gender priming predicted hostile sexism after controlling for grammatical gender priming (although this effect was not moderated by motivation to control). In contrast, grammatical gender priming did not predict hostile sexism after controlling for stereotypical gender priming. Finally, our exploratory analyses regarding voting intentions indicated that stereotypical gender priming predicted the intention to vote for Donald Trump versus Hillary Clinton after controlling for grammatical gender priming, the latter of which showed no significant relation with voting intentions in the same model.

The pattern of results was rather different when participants completed the grammatical gender SMP first. In this case, the stereotypical gender priming scores showed reduced construct validity and reduced discriminant validity with regard to their overlap with the measure of grammatical gender priming. This pattern was most evident in the analysis of partial correlations and the regression analysis, indicating the absence of construct-valid relations between stereotypical gender priming and explicit measures of sexism measures after controlling for grammatical gender priming. When participants completed the grammatical gender SMP first, stereotypical gender priming also did not predict the intention to vote for Donald Trump versus Hillary Clinton after controlling for grammatical gender priming. Although grammatical gender priming showed significant relations with explicit measures of sexism, these relations did not hold after controlling for stereotypical gender priming. Overall, these findings are consistent with the hypothesis that interspersed presentations of the two types of priming trials inflated the overlap between grammatical gender priming and stereotypical gender priming. When spillover effects from grammatical priming trials were eliminated (i.e., when the stereotypical gender priming task is completed before the grammatical gender priming task), there was evidence for the construct and discriminant validity of stereotypical gender priming scores, although overlap to grammatical gender priming scores was not fully eliminated.

Despite this conclusion, it is worth noting that Study 3 did not replicate some of the findings from Study 1 and Study 2. Whereas stereotypical gender priming was uncorrelated with modern sexism in Studies 1 and 2, the two measures were positively correlated in Study 3. However, the positive correlation in Study 3 is still consistent with our prediction that the two measures should *not* be negatively correlated. If stereotypical gender priming would show a negative correlation with modern sexism, one would have to conclude that effects of stereotypical gender priming reflect mere knowledge of gender inequality rather genuine effects of gender stereotyping. Moreover, whereas stereotypical gender priming was unrelated to motivation to control in Studies 1 and 2, the two measures were negative correlated in Study 3. The latter finding suggests that responses in the SMP are not immune to intentional control

[7]When we added the three explicit measures of sexism as predictors, none of the priming indices were significant predictors (all *p*s > .10) of voting intention in either Block Order condition. Instead, benevolent sexism was the only significant predictor (*p* = .008) in the grammatical priming first condition, while modern sexism was the only significant predictor (*p* < .001) in the stereotypical priming first condition. However, these results should be treated with caution, because the SMP and the explicit measures of sexism were not counter-balanced in the current study, implying a confound between type of measure and proximity to the voting intention measure. For this reason, results from these analyses are not reported in detail. The full results are available from the authors upon request.

(see also Teige-Mocigemba, Penzl, Becker, Henn, & Klauer, 2016). Yet, differences between samples (Canadian undergraduates in Studies 1 and 2 vs. U.S. residents in Study 3) and testing environments (laboratory in Studies 1 and 2 vs. online in Study 3) might have contributed to these discrepancies as well.

General Discussion

The main goal of the current research was to validate a new implicit measure of gender stereotyping based on the SMP, a modified variant of the AMP designed to assess spontaneous behaviors resulting from the activation of semantic concepts rather than affective states (Deutsch & Gawronski, 2009; Gawronski & Ye, 2014; Imhoff et al., 2011; Sava et al., 2012). Across three studies, SMP scores of stereotypical gender priming consistently showed large effect sizes, satisfactory internal consistency, and construct-valid relations with explicit measures of contemporary sexism. Discriminant validity of stereotypical gender priming scores was tested by investigating relations with priming effects involving grammatical gender. Evidence for discriminant validity was obtained when (1) the two kinds of priming trials were presented in a blocked rather than interspersed manner and (2) the measure of stereotypical gender priming preceded the measure of grammatical gender priming. Finally, preliminary evidence for predictive validity is provided by the finding that stereotypical gender priming scores predicted the intention to vote for Donald Trump versus Hillary Clinton in the 2016 U.S. presidential campaign. Overall, these findings suggest that the SMP is a promising tool for measuring gender stereotyping.

Our SMP-based measure has several advantages over existing IAT-based measures of gender stereotyping (see Zitelny et al., 2017). As explained in detail by Teige-Mocigemba et al. (2010), the blocked task-structure of the IAT has been linked to various sources of systematic measurement error that can undermine its construct validity. The SMP avoids these limitations by relying on a much simpler procedure based on the notion of sequential priming. Yet, like any other implicit measure, the SMP has its own limitations, one being the substantial proportion of method variance in SMP scores (see Studies 2 and 3). Because performance on different measures can be driven by distinct processes, effects obtained with one measure may not generalize to other measures to the extent that these effects are driven by method-related processes (e.g., Deutsch & Gawronski, 2009; Gawronski & Bodenhausen, 2005; Gawronski, Cunningham, LeBel, & Deutsch, 2010). Thus, to ensure accurate interpretations of findings obtained with a given measure, it is recommended to replicate these findings with alternative measures based on a different underlying mechanism (Gawronski, Deutsch, LeBel, & Peters, 2008). To the extent that a given finding replicates across measures, it seems reasonable to assume that this finding reflects a genuine effect involving the construct of interest. However, if different measures show different outcomes for the same study design, it seems likely that these outcomes are driven by method-related rather than construct-related processes.

Gender Stereotyping and Sexism

In addition to providing evidence for the validity of the SMP as an implicit measure of gender stereotyping, the current findings provide deeper insights into the relation between gender stereotyping and sexist attitudes. The positive relations of gender stereotyping scores to both hostile and benevolent sexism suggest that gender stereotyping is systematically related to sexist beliefs regardless of the evaluative connotation of these beliefs. As we mentioned earlier, hostile sexism represents overt negativity against women, whereas benevolent sexism reflects discriminatory positive attitudes toward women (Glick & Fiske, 1996). Moreover, both types of sexism have been linked to stereotypical beliefs. Whereas hostile sexism involves beliefs that women are unfit for certain powerful roles in society, benevolent sexism involves patronizing beliefs that confine women to stereotype-congruent roles. The current findings suggest that gender stereotypes can give rise to both negative and positive forms of sexism and discrimination. If gender stereotyping was exclusively related to negative beliefs about women, it should have shown a positive correlation with hostile sexism and a negative correlation with benevolent sexism.

More broadly, the current findings also provide deeper insights into the relation between implicit and explicit measures in the area of sexism and gender stereotyping. Previous research on racial attitudes suggests that responses on implicit measure can reflect mental contents that participants are unwilling to report, whereas responses on explicit measures are subject to motivational control (e.g., Dunton & Fazio, 1997; Fazio et al., 1995; Payne et al., 2005). Consistent with this idea, the relation between hostile sexism and gender stereotyping in the SMP was moderated by motivation to control prejudiced reactions. That is, gender stereotyping was positively related to hostile sexism when motivation to control was low, but not when it was high. This moderation stands in contrast to earlier claims by Glick and Fiske (1996), who suggested that people are less concerned about appearing sexist than they are about appearing racist. The current findings suggest that motivation to control may play a similar role in the expression of sexist beliefs, at least when these beliefs involve negative views about women.

Interestingly, motivation to control did not influence the expression of sexist beliefs involving discriminatory positive views about women. In the current work, gender stereotyping in the SMP was positively related to benevolent sexism, and this relation was unaffected by motivation to control prejudiced reactions. A potential interpretation of this finding is that sexist beliefs involving positive views about women do not represent a form of prejudice in the eyes of people who hold such

 Copyright © 2017 John Wiley & Sons, Ltd.

views. Hence, concerns about expressing prejudiced beliefs may not affect the expression of beliefs related to benevolent sexism, because the latter beliefs are not regarded as prejudice. Together, these assumptions imply that motivation to control should moderate the relation between gender stereotyping and hostile sexism, but not the relation between gender stereotyping and benevolent sexism, consistent with the findings in the current studies.

Future Directions

The current research was concerned with a particular type of stereotypes: gender stereotypes involving different occupations. The central idea underlying this work is that people may spontaneously think of men when they are exposed to certain kinds of occupations (e.g., doctor, mechanic, engineer) and of women when they are exposed to other kinds of occupations (e.g., nurse, secretary, receptionist). This idea was captured by the sequence of primes and targets in the SMP, in that the primes involved gender stereotypical occupations that may activate thoughts of men and women, which in turn may bias gender-related judgments of otherwise gender-neutral targets. Yet, gender stereotyping can come in various other forms that are different from the setup in the current SMP. For example, researchers may be interested in whether exposure to men and women activates thoughts of different social roles (e.g., career vs. household) or different trait concepts (e.g., strong vs. weak). Although the current work focused specifically on occupational gender stereotyping, it provides a foundation for applications of the SMP to capture alternative forms of gender stereotyping. For example, it seems possible to construct an SMP that includes male and female faces as primes and a judgmental task involving guesses about whether the Chinese ideographs refer to words related to career or household. Similarly, one could construct an SMP with male and female face primes and judgments about the perceived strength or weakness of the Chinese ideographs. Thus, the current work opens the door for a wide range of applications of the SMP to measure other forms gender stereotyping.

Yet, any research using such modified SMPs would benefit from prior validation studies like the current ones. A particularly important issue in this regard is evidence for discriminant validity. The present research represents one of the few cases that have examined the discriminant validity of an implicit measure with a procedurally similar measure that taps onto a conceptually distinct construct (see also Back, Schmukle, & Egloff, 2009). Although we found evidence for unique variance in measures of stereotypical gender priming and grammatical gender priming, the two kinds of priming scores showed considerable overlap. Moreover, although the degree of overlap was moderated by procedural factors (i.e., blocked vs. interspersed presentations of the two kinds of primes; order of the two kinds of measures when using blocked presentations), the overlap was not eliminated under conditions that reduced spillover effects from grammatical gender priming trials to stereotypical gender priming trials. Thus, any relation of stereotypical gender priming scores to a criterion measure may reflect either (1) construct-related effects of gender stereotyping or (2) construct-independent effects of method-related factors (e.g., content-independent differences in the tendency to show a priming effect in the SMP; see Wentura & Degner, 2010). In addition to ruling out the involvement of construct-unrelated factors in the relation between gender stereotyping and explicit measures of sexism, the current work suggests a potential way to tackle this issue in future research with the SMP. By including an additional SMP capturing construct-independent effects of method-related factors, it is possible to control for such factors and thereby provide more compelling evidence that the obtained relations to criterion measures reflect genuine effects of the to-be-measured construct. Yet, such additional SMPs should be administered after the critical SMPs for the to-be-measured construct to avoid potential spillover effects of the kind obtained in the current research.

More broadly, our findings highlight the necessity to establish the construct validity of implicit measures, especially for new applications of an existing paradigm with novel materials (e.g., primes, targets, categories). It is common practice to assume construct validity of such applications based on their face validity. Yet, the current findings suggest that this practice can be problematic, because findings obtained with a measure designed to capture a particular construct (e.g., stereotypical gender priming) might also be obtained with a procedurally similar measure that assesses a theoretically distinct construct (e.g., grammatical gender priming). For example, in the domain of racial attitudes, it seems important to establish the discriminant validity of an implicit measure of spontaneous evaluations of racial categories (e.g., Black vs. White) vis-à-vis applications of the same measure to assess spontaneous evaluations of concepts that are unrelated to racial categories (e.g., flowers vs. insects). To the extent that the two measures produce the same pattern of results, it would seem premature to interpret the obtained effect in terms of racial attitudes.[8]

Limitations

Although the SMP has some advantages over other implicit measures such as the IAT, it also has several limitations. First, findings obtained with the SMP might be more sensitive to extreme scores compared to those obtained with other implicit measures. Bar-Anan and Nosek (2014) showed that excluding

[8]In research using the IAT (Greenwald et al., 1998), such a pattern may occur when the obtained effect is driven by differences in inhibitory control rather than racial attitudes (see Sherman et al., 2008).

participants with the 10% most extreme scores had a stronger impact on the AMP compared to other implicit measures. Whereas for most other measures internal consistencies and correlations with explicit measures remained unaffected, the AMP showed reduced internal consistencies and reduced correlations with explicit measures after exclusion of the 10% most extreme scores.

To test whether the SMP is similarly affected by the exclusion of extreme scores, we re-ran the analyses in all three studies after excluding participants who showed the 10% most extreme scores on the stereotypical and grammatical priming indices (5% on each end of the distribution).[9] Aside from a few (relatively small) deviations in the size and significance levels of some of the reported results, the results were largely consistent with the original analyses. Moreover, the results were only partly consistent with Bar-Anan and Nosek's (2014) finding: Although the internal consistencies of SMP scores tended to be slightly lower after exclusion of participants with extreme scores, their correlations with explicit measures remained largely unaffected. Thus, more research is needed before drawing inferences about shortcomings of the SMP from empirically demonstrated shortcomings of the AMP.

A second limitation is that SMP scores, like the scores of many other implicit measures, are based on differences between two individual component scores (e.g., the percentage of *male* responses on two different types of trials). There are some well-known problems with the use of difference scores in psychological research (for a review, see Edwards, 2002), such as confounded effects due to the failure to account for the absolute value of each component score. Future research using novel statistical methods (e.g., polynomial regression; see Edwards, 2002) may help to address the shortcomings of difference scores in research using the SMP and other implicit measures (e.g., by modeling the levels of congruence and incongruence of two component scores).

A third concern is that priming effects in the SMP might reflect the intentional application of prime features in judgments of the targets rather than misattribution of mental contents activated by the primes (see Bar-Anan & Nosek, 2014). Although previous research supports the presumed role of misattribution processes in the AMP (e.g., Gawronski & Ye, 2014; Payne et al., 2013), this evidence does not necessarily speak against an additional contribution of intentional processes. One potential way to rule out intentional processes in the SMP is to direct participants' attention away from the relevant features of the primes. To the extent that priming effects remain intact under such conditions, it seems less likely that intentional processes contributed to the observed outcomes. In line with this idea, Gawronski and Ye (2015) presented participants with Black and White faces of either young or old age in an AMP to measure spontaneous race and age bias. In one condition, participants were asked to keep a mental tally of how many Black and White faces are presented throughout the task. In another condition, participants were asked to keep a mental tally of how many young and old faces are presented throughout the task. Consistent with earlier findings (Gawronski et al., 2010), Gawronski and Ye (2015) found reliable priming effects of race and age bias regardless of whether participants were instructed to pay attention to the race or the age of the face primes. More importantly, priming scores of race bias were significant related to self-reported intentional use of the primes in judging the targets only when participants were instructed to pay attention to race, but not when they were instructed to pay attention to age. Conversely, priming scores of age bias were significant related to self-reported intentional use of the primes only when participants were instructed to pay attention to age, but not when they were instructed to pay attention to race. Similar procedures may help to rule out intentional processes in the SMP.

A final issue concerns the high internal consistencies for the grammatical gender SMP in the current studies. Although high internal consistencies are desirable for measures designed to capture individual differences in the construct of interest, high internal consistency can be a matter of concern when there is no reason to expect systematic variation across participants. Assuming that all participants understood the semantic meaning of the grammatical gender primes in the current studies (e.g., king, queen), there should be no systematic variation in priming scores across participants, leading to low internal consistency of the priming scores. Thus, the fact that the grammatical gender SMP showed internal consistencies that were comparable to the stereotypical gender SMP suggests that participants systematically differ in their tendency to show a priming effect in the SMP. A potential interpretation of this tendency is that participants differ in the degree to which they apply activated mental contents in judgments of the ambiguous targets, which may be independent of the nature of the activated contents. Such a construct-independent tendency explains not only the high internal consistency of the grammatical gender SMP, but also the high correlations between two kinds of priming scores in the present research. Conceptually, individual differences in the tendency to apply activated mental contents can be considered as a source of method variance (Campbell & Fiske, 1959) that should be controlled in studies using the SMP. For example, if a given factor influences the tendency to apply activated mental contents in the SMP, the observed changes in SMP scores may be misinterpreted as reflecting a genuine change in the activation of the mental contents. Such misinterpretations can be prevented by measuring and controlling for method variance, similar to the use of a grammatical gender SMP in the current studies.

[9]A summary of the results with extreme score exclusion in comparison to the results without such exclusion, as well as the complete results of the analyses with extreme score exclusion, can be found in the Supporting Information.

 Copyright © 2017 John Wiley & Sons, Ltd.

Conclusion

Much is to be learned about gender stereotyping, especially with the increasing concerns about contemporary forms of sexism. In the present research, we aimed to validate a new measure of gender stereotyping based on the SMP. Our findings suggest that the SMP is a promising tool for the assessment of gender stereotyping, while highlighting the importance of testing and establishing the construct validity of implicit measures.

Author Note

The research reported in this article has been supported by grants from Social Sciences and Humanities Research Council of Canada (Canada Research Chairs Program #215983) and Fonds Wetenschappelijk Onderzoek (Methusalem Grant BOF16/MET_V/002 of Ghent University).

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.

References

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*, 533–548.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, *68*, 181–198.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136–141.

Banse, R., & Gawronski, B. (2003). Die Skala Motivation zu vorurteilsfreiem Verhalten: Skaleneigenschaften und Validierung. *Diagnostica*, *49*, 4–13.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*, 668–688.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, *25*, 413–423.

Cohen, S. L., & Bunker, K. A. (1975). Subtle effects of sex role stereotypes on recruiters' hiring decisions. *Journal of Applied Psychology*, *60*, 566–572.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*, 642–658.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, *24*, 252–287.

Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, *46*, 991–1004.

Degner, J., & Wentura, D. (2008). The extrinsic affective Simon task as an instrument for indirect assessment of prejudice. *European Journal of Social Psychology*, *38*, 1033–1043.

Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, *45*, 101–114.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology*, *46*, 735–754.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow, & N. W. Schmitt (Eds.), *Advances in measurement and data analysis* (pp. 350–400). San Francisco: Jossey-Bass.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology*, *89*, 672–685.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion*, *24*, 1008–1025.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 283–310). New York: Cambridge University Press.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, *24*, 218–225.

Gawronski, B., Ehrenberg, K., Banse, R., Zukova, J., & Klauer, K. C. (2003). It's in the mind of the beholder: The impact of stereotypic associations on category-based and individuating impression formation. *Journal of Experimental Social Psychology*, *39*, 16–30.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the

construal of individuating information. *European Journal of Social Psychology*, *33*, 573–589.

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin*, *40*, 3–15.

Gawronski, B., & Ye, Y. (2015). Prevention of intention invention in the affect misattribution procedure. *Social Psychological and Personality Science*, *6*, 101–108.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*, 491–512.

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*, 109–118.

Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex discrimination in hiring decisions? *Journal of Personality and Social Psychology*, *55*, 178–186.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*, 657–674.

Imhoff, R., Schmidt, A. F., Bernhardt, J., Dierksmeier, A., & Banse, R. (2011). An inkblot for sexual preference: A semantic variant of the affect misattribution procedure. *Cognition and Emotion*, *25*, 676–690.

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology*, *88*, 498–509.

Loersch, C., & Payne, B. K. (2011). The situated inference model an integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science*, *6*, 234–252.

Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *Journal of Neuroscience*, *21*, 1–6.

Moors, A., & De Houwer, J. (2006). Automaticity: A conceptual and theoretical analysis. *Psychological Bulletin*, *132*, 297–326.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101–115.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Kesebir, S. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*, 10593–10597.

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the Affect Misattribution Procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin*, *39*, 375–386.

Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277–293.

Payne, B. K., Hall, D. L., Cameron, C. D., & Bishara, A. J. (2010). A process model of affect misattribution. *Personality and Social Psychology Bulletin*, *36*, 1397–1408.

Payne, B. K., & Lundberg, K. (2014). The affect misattribution procedure. *Social and Personality Psychology Compass*, *8*, 672–686.

Sava, F. A., MaricuToiu, L. P., Rusu, S., Macsinga, I., Vîrgă, D., Cheng, C. M., & Payne, B. K. (2012). An inkblot for the implicit assessment of personality: The semantic misattribution procedure. *European Journal of Personality*, *26*, 613–628.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, *102*, 947–963.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*, 199–214.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to Implicit Association Tests and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York: Guilford Press.

Teige-Mocigemba, S., Penzl, B., Becker, M., Henn, L., & Klauer, K. C. (2016). Controlling the "uncontrollable": Faking effects on the affect misattribution procedure. *Cognition and Emotion*, *8*, 1470–1484.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 95–116). New York: Guilford Press.

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles*, *55*, 259–266.

Zitelny, H., Shalom, M., & Bar-Anan, Y. (2017). What is the implicit gender-science stereotype? Exploring correlations between the gender-science IAT and self-report measures. *Social Psychological and Personality Science*. Advance online publication. https://doi.org/10.1177/1948550616683017