

Bias and accuracy in judging sexism in mixed-gender social interactions

Group Processes & Intergroup Relations

2017, Vol. 20(6) 850–866

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1368430216638530

journals.sagepub.com/home/gpi



Jin X. Goh,¹ Aria Rad,¹ and Judith A. Hall¹

Abstract

This research examined bias and accuracy in judging hostile and benevolent sexism during mixed-gender interactions. Bias is defined as underestimation or overestimation of a partner's sexism. Accuracy is defined as covariation in two different ways, as (a) the strength of the association between a dyad member's judgment and their partner's sexism, across dyads, and (b) the ability to differentiate sexism between multiple targets. In Studies 1 and 2, members of mixed-gender dyads rated their own and their partners' benevolent and hostile sexism. Overall, there was no covariation, across dyads, between perceptions and the partner's self-reported sexism. However, women overestimated men's hostile sexism; there was no evidence of biases for women judging men's benevolent sexism. Men underestimated women's hostile sexism and overestimated benevolent sexism. In Study 3, participants watched brief videos of male or female students (targets) from Study 1 and completed benevolent or hostile sexism items for each target as they thought the target would fill them out. Accuracy for detecting sexism across multiple targets (using sensitivity correlations) was significantly above chance for both forms of sexism. Male and female participants were more accurate at detecting hostile sexism in male targets than female targets. Participants were more accurate at detecting benevolent sexism of same-gender targets than opposite-gender targets. When judging targets of opposite gender, women's judgments were significantly above chance for both forms of sexism, but men were not accurate for either forms of sexism. These studies suggest that there is bias and accuracy in first impression judgments of sexism.

Keywords

accuracy, benevolent sexism, bias, hostile sexism, social interactions

Paper received 19 January 2015; revised version accepted 17 February 2016.

Upon meeting a person for the first time, one usually forms impressions (e.g., demographic characteristics, personality traits, emotional states, and attitudes) of the person relatively quickly. These first impressions are important for a host of reasons and chief among them is determining whether a person would pose a potential threat to oneself (e.g., Zebrowitz & Collins, 1997). However,

these first impressions may or may not be accurate. While there is considerable research dedicated to

¹Northeastern University, USA

Corresponding author:

Jin X. Goh, Department of Psychology, Northeastern University, 360 Huntington Ave., Boston, MA 02215, USA.
Email: goh.j@husky.neu.edu

understanding accuracy of first impression in judging personality traits (Connelly & Ones, 2010), emotions (Matsumoto et al., 2000), and demographic characteristics such as sexual orientation (Tskhay & Rule, 2013), relatively little research has been dedicated to understanding first impressions of social attitudes despite the obvious importance of such ability. The present research is concerned with the judgment of sexism in target persons engaged in casual conversation. Such accuracy, if it exists, may lead to successful confrontation or resolution, or may enhance a person's ability to avoid others who could cause damaging outcomes or who are simply offensive.

The current research examined both accuracy and bias in judging sexist attitudes toward women during live mixed-gender social interactions and from watching video clips of those interactions. We explicitly differentiated bias from accuracy when analyzing dyad partners' hostile and benevolent sexism, and thus we are able to contribute to existing research both methodologically and theoretically.

Hostile and Benevolent Sexism

Glick and Fiske (1996, 2001) proposed that the nature of sexism is ambivalent: Sexism can hold both hostile and benevolent properties. Hostile sexism is an antipathy toward women and it upholds male dominance through the degradation and defamation of women; it is rooted in fear and aversion of women who defy nontraditional gender roles and the status quo. Benevolent sexism, in contrast, posits that women are wonderful but weak, and therefore they require men's adoration and protection. Because of its seemingly positive nature, benevolent sexism is an insidious ideology. Unlike simple politeness, benevolent sexism reinforces the idea that men are more competent than women. Benevolent sexism is rooted in paternalistic affection rather than outright domination to reinforce gender inequality.

Due to its negative nature, hostile sexism often incurs resistance and thus is less likely to be accepted by women. Across 19 countries, Glick et al. (2000) found that men, compared to women,

consistently expressed stronger endorsement of hostile sexism; in contrast, few gender differences were found in the endorsement of benevolent sexism, and in certain countries such as Cuba and Nigeria, women even endorsed benevolent sexism to a higher degree than men did. Moreover, research has found that women generally rated benevolent sexist men more favorably than hostile sexist men, and people were less likely to perceive or acknowledge benevolent sexism as "sexism" (Good & Rudman, 2010; Kilianski & Rudman, 1998).

According to Glick and Fiske (2001), benevolent sexism can be attractive for women because it is difficult to resist sexism in such a form. Indeed, research has shown that, behaviorally, men with more benevolent sexist attitudes behaved friendlier nonverbally (e.g., smiled more) and were more patient while interacting with women whom they had just met (Goh & Hall, 2015). Within intimate romantic relationships, men with more benevolent sexism were more likely to help their female partners by directly providing plans and solutions; while this dependency-oriented support may seem positive on the surface, it actually diminished their female partners' feelings of competence and positive regard (Hammond & Overall, 2015).

Research has also shown that exposure to benevolent sexism can have negative effects. Women who read statements supporting benevolent sexism (e.g., "Men are incomplete without women"), as opposed to statements supporting hostile sexism (e.g., "Women are too easily offended"), were subsequently more willing to support the status quo and less motivated to participate in collective action against it (Becker & Wright, 2011; Jost & Kay, 2005). By incentivizing women with benevolence and affection, men can assert their power without much resistance as women may become more complacent in matters of gender inequality.

Despite their surface-level differences, benevolent sexism and hostile sexism exist as complementary rather than opposing elements. Hostile and benevolent sexism both contribute to inequality in contemporary society and an individual

can endorse both attitudes simultaneously and to different degrees. Skills in recognizing these two forms of sexism in another person would carry certain benefits such as determining the suitability of a potential romantic partner or friend. However, the accuracy of first impression in judging sexism has not been examined.

Defining and Measuring Bias and Accuracy

Imagine a mixed-gender interaction between a man named Barney and a woman named Lily. Within this interaction, Lily's perception and judgment of Barney (and vice versa) will influence the progress of the interaction. For instance, Lily needs to gauge how much of a hostile sexist Barney is in order to adjust her behavior (e.g., her judgment may help her decide to end the interaction early). Lily may overestimate Barney's hostile sexism or she may underestimate it, thinking he is less of a hostile sexist than he actually is. This over- or underestimation is what we refer to as bias and it might occur for just one target person or with a whole group of target people. Bias is also referred to as directional bias (West & Kenny, 2011) and mean-level bias (Fletcher & Kerr, 2010). Measuring bias in perceivers' judgments involves calculating arithmetic discrepancy (i.e., mean-comparison) between the judgment and criterion. Using the previous example, the criterion is Barney's actual sexism as operationally defined, while the judgment is Lily's estimation.

For just the one dyad—Lily and Barney—there is no corresponding measure of accuracy, which we define as a covariation between judgments and criterion. However, across a sample of dyads, one person's ratings can be correlated with the partner's self-ratings (criterion), yielding a correlation that describes the group's covariation: if higher ratings go with higher partner self-ratings, we would say the group has some accuracy. This *group-level* (Hall, Bernieri, & Carney, 2005) calculation of accuracy based on a set of dyads is also referred to as the truth force (West & Kenny, 2011) and tracking accuracy (Fletcher & Kerr, 2010). This operational definition of accuracy

was used in Studies 1 and 2, which were both dyadic studies. However, with this method, higher accuracy cannot be attributed uniquely to people's perceptual skills because the accuracy might equally be due to the partners' degree of expressivity of the quality being judged (Hall, Rosip, Smith LeBeau, Horgan, & Carter, 2006; Hall, Schmid Mast, & Latu, 2015; Snodgrass, Hecht, & Ploutz-Snyder, 1998); in an extreme case, all of the accuracy could be due to how much the partners revealed the quality being judged (e.g., their sexism), and not due to judges' ability to pick up their partners' sexism. Therefore, accuracy calculated in this way has some limitations as an index of perception accuracy.

In Study 3, we used a less ambiguous and more statistically powerful indicator of perception accuracy by adding more targets to be rated. The analogue in the real-life situation would be if more male friends join Barney and Lily, and Lily could then make comparative judgments. This was accomplished in Study 3 by asking new perceivers to rate the sexism of multiple target persons. Now a positive correlation between ratings and criterion values can be attributed to individual perceivers' accuracy (because all perceivers rate the same group of targets), with a further advantage being that accuracy scores (correlations) can be calculated for individual perceivers, allowing more options in data analysis. This way of calculating accuracy has been called sensitivity correlation (Judd & Park, 1993) and trait-based accuracy (Back & Nestler, 2016).

As stated before, we differentiate two forms of accuracy: one derived from judgments in dyadic interactions and the other derived from judgments across multiple targets. Correlations between the judgments and the criteria capture covariations and strength of the relationship. Correlations, because they standardize the variables being correlated, are not subject to mean-level response differences and therefore accuracy measured using correlations offers a very different window into interpersonal perception than a mean-comparison (Judd & Park, 1993).

Bias and accuracy are statistically and theoretically independent processes, and one can be

present with or without the other (Fletcher, 2015; Fletcher & Kerr, 2010; Hall, Stein, Roter, & Rieser, 1999; Judd & Park, 1993; West & Kenny, 2011). In the past, researchers have focused on bias or accuracy, and have rarely considered both in conjunction. Furthermore, researchers have also used the term accuracy interchangeably with bias even though these two constructs are statistically independent. With recent methodological advances, both bias and the group-level kind of accuracy can now be calculated simultaneously across dyadic interactions. For instance, the truth and bias model (Stern, West, & Schoenthaler, 2013; West & Kenny, 2011) allows for simultaneous calculation of both group-level accuracy and bias within one statistical model. Bias is the mean discrepancy between the criterion and the judgment with positive values indicating overestimation of the criterion and negative values indicating underestimation. Accuracy is the strength of the criterion on the judgment, which is calculated using regression with the judgment as the outcome variable and the truth criterion as the predictor variable. In the regression model, the intercept represents the bias and the main effect represents the accuracy.

Bias and Accuracy in Intergroup Relations

People are biased when judging the characteristics of outgroup members such that they tend to overestimate outgroup members' likelihood of possessing stereotypical characteristics (Judd & Park, 1993). Diekmann, Eagly, and Kulesa (2002) examined male and female participants' estimates of attitudes concerning various social and political issues held by men and women (using the General Social Survey, a nationally representative survey, as the criterion). Participants consistently underestimated men's support for social and political issues that favored women, thus suggesting that people are negatively biased in estimating men's gender attitudes. Edmonds, Cahoon, and Shipman (1991) measured male and female students' estimates of a typical member of the opposite gender's attitude on acceptance of

interpersonal violence, attitudes toward sex role stereotyping, adversarial sexual beliefs, sexual conservatism, and rape myth acceptance. Male and female participants' mean-level responses were used as the criterion for the other gender to measure bias in estimation of gender attitudes. The authors found that while men were not biased in estimating women's attitudes across these various domains (i.e., no significant differences between men's estimates and women's self-reports), women consistently overestimated men's self-reported gender attitudes on these domains. These findings suggest that biases in estimation exist regarding attitudinal judgments.

While the literature is rich in understanding accuracy in judgments of emotional states and personality traits (e.g., Ambady, Hallahan, & Rosenthal, 1995; Carney, Colvin, & Hall, 2007; Funder & Colvin, 1988; Matsumoto et al., 2000), few studies have examined accuracy in domains relating to prejudice. This literature suggests people may be accurate to a certain degree. Black perceivers, relative to White perceivers, were able to detect racism at above-chance levels in White people engaging in interracial interactions by observing 20-s silent videos of their interaction behavior (Richeson & Shelton, 2005). Accuracy in detecting racial prejudice in White men has also been found using static facial photographs alone (Hehman, Leitner, Deegan, & Gaertner, 2013). This ability to detect racism may in part be driven by minorities' suspicion of White people's motives during interracial interactions (LaCrosse et al., 2015). In terms of gender dynamics, Diekmann et al. (2002) showed that male and female participants' judgments correlated positively with men's and women's responses on the General Social Survey, demonstrating that these judgments can also be accurate in the correlational sense.

Most relevant to the current research, Rudman and Fetterolf (2014) examined bias in participants who judged the opposite gender's hostile and benevolent sexism level. Male and female participants completed the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996) to measure their own endorsement of hostile sexism and

benevolent sexism. They then completed a second ASI as they thought a typical member of the opposite gender would respond, thus allowing the researchers to measure men's and women's estimates of the other gender's sexist attitudes using the group-level endorsement as the criterion. Across three different populations (Caucasian Americans, Asian Americans, and African Americans), women overestimated men's hostile sexism but they underestimated men's benevolent sexism. Men, on the other hand, overestimated women's benevolent sexism but underestimated women's hostile sexism. These discrepancies led the authors to conclude that both men and women missed the mark in judging the other gender's sexist attitudes. But as noted by previous research (Judd & Park, 1993; West & Kenny, 2011), discrepancy in terms of biased estimation, which Rudman and Fetterolf used, does not mean accuracy in the covariation sense. While Rudman and Fetterolf's research offered evidence that people were *biased* in estimating sexism, their *accuracy* in discerning who was more sexist than whom was not measured.

Functional Theories

According to functional theories (e.g., Zebrowitz & Collins, 1997), accuracy serves an adaptive function, and characteristics that are specifically relevant to an individual are judged quicker and more accurately. Richeson and Shelton (2005) argued that Black people were more accurate than White people at detecting racial bias of White targets because accuracy in such a context serves adaptive purposes for Black people more so than it would for White people. Similarly, LaCosse et al. (2015) and Hehman et al. (2013) found that motivated minority members were more accurate at detecting racial prejudice in White targets.

Functional theories can potentially explain accuracy and bias in judging prejudice in others, but evidence to date is limited to racism and generalizability to sexism has not been fully examined. It is unclear whether people are accurate or inaccurate, or biased or not, in judging sexist attitudes in actual people whose behavior they have

observed. This is further complicated by the fact that accuracy and bias are independent and both can be found in judging other people's attitudes (Diekmann et al., 2002). The present research addresses this important gap and examines bias and accuracy in judging hostile and benevolent sexism. The truth and bias model was utilized in Studies 1 and 2 to analyze bias and group-level accuracy in judging a partner's sexist attitude during mixed-gender interactions. Individual perceivers' sensitivity correlations were used to understand accuracy in judging sexism across multiple targets in Study 3.

Current Research

Rudman and Fetterolf (2014) based their findings on each gender's estimation of the other gender's sexist attitudes *as a group*; specifically, participants imagined how a typical member of the opposite gender would respond. In the present Studies 1 and 2, participants judged the hostile and benevolent sexism of their own interaction partners in mixed-gender interactions. In Study 3, participants judged hostile and benevolent sexism from excerpts of Study 1's videos. Thus, the present research moves from stereotype at an *intergroup* level into the realm of bias and accuracy at an *interpersonal* level. The four bias hypotheses, consistent with the findings by Rudman and Fetterolf (2014), are as follows:

Hypothesis 1: Women will overestimate men's hostile sexism.

Hypothesis 2: Men will underestimate women's hostile sexism.

Hypothesis 3: Women will underestimate men's benevolent sexism.

Hypothesis 4: Men will overestimate women's benevolent sexism.

Also, although we measured accuracy as previously defined, we did not make a strong prediction regarding accuracy because we did not know whether it would be possible to detect sexism at the dyadic level or from short exposures to

multiple targets' behavior. However, past research offers reasons why accuracy might indeed exist. First, previous analysis of the behavior of the targets whose sexism was being judged showed that there were behavioral cues associated with both benevolent and hostile sexism (Goh & Hall, 2015). Second, accuracy for judging negative intergroup attitudes has been shown in past research (e.g., Hehman et al., 2013; LaCosse et al., 2015; Richeson & Shelton, 2005) though no one has investigated accuracy of detecting sexism.

Study 1

Method

Participants. Participants were 30 mixed-gender dyads ($M_{\text{age}} = 19.37$) from Northeastern University who participated for partial course credit. Originally there were 33 dyads, but three dyads were removed because they either knew one another prior to the study or one member of the dyad did not speak English fluently.

Procedure. Male and female students were recruited separately through the Psychology Department's participant pool. After a pair of male and female students arrived, they were informed that this was a study on social interaction but not that we were measuring sexism. Participants first engaged in a structured social interaction in which they selected trivia questions and took turns asking one another the questions; they then engaged in a 3-min free-structured interaction in which they were allowed to discuss anything they wished (most conversations centered on getting acquainted, the trivia game, and their classes). The whole interaction was videotaped with participants' knowledge. After this, participants completed a battery of surveys measuring their personality, self-esteem, and finally their own sexism levels as well as judgment of their partner's sexism level. The personality and self-esteem measures were included as distractors and were not analyzed for this study. Participants were then debriefed and asked to sign a waiver releasing their videos for analysis (for more information regarding the nonverbal and

verbal analyses of the videos, see Goh & Hall, 2015).

Ambivalent Sexism Inventory (ASI). The ASI (Glick & Fiske, 1996) contains 22 items that measure both benevolent and hostile sexism on a 6-point rating scale (0 = *disagree strongly*; 5 = *agree strongly*). Examples of hostile sexism items are "Women exaggerate problems they have at work" and "Most women interpret innocent remarks or acts as being sexist." Examples of benevolent sexism items are "A good woman should be set on a pedestal by her man" and "In a disaster, women ought not necessarily to be rescued before men" (reverse-scored). Participants first filled out the ASI for themselves, and then they completed a second ASI as they thought their partner would fill it out.

Results and Discussion

We tested the four hypotheses using the truth and bias model, comparing each dyad member's judgment of the partner's sexism scores with the partner's own self-reported sexism scores. In the truth and bias model, the judgment of the partner's sexism is the outcome variable (centered on the grand mean of participants' self-reported sexism), and the participants' self-reported sexism is treated as the predictor (grand mean centered). Bias is the intercept of the model with a positive value indicating overestimation and a negative value indicating underestimation. Accuracy is the main effect of the model. Gender of the participants (male = -1; female = 1) was entered as a moderator. All analyses were conducted using the MIXED procedure in SPSS 22 with hostile sexism and benevolent sexism as separate models. See Table 1 for descriptive statistics and Table 2 for estimates in the truth and bias model.

Across all participants' judgments of their partners' hostile sexism, neither the intercept (indicating bias), $b = .07$, $t(21.34) = 0.73$, $p = .474$, nor the main effect (indicating accuracy) was significant, $b = .01$, $t(48.66) = 0.01$, $p = .930$. This suggests that in general, people were not biased

Table 1. Means and standard deviations of opposite-gender target ratings in Studies 1 and 2.

Sexism ratings	Women's judgment	Partners' self-report	Men's judgment	Partners' self-report
Study 1				
Hostile sexism	3.15 (0.77)	2.36 (0.85)	1.39 (0.72)	2.02 (1.14)
Benevolent sexism	2.24 (0.65)	2.66 (0.78)	3.04 (0.65)	2.24 (1.01)
Study 2				
Hostile sexism	3.49 (0.48)	3.21 (0.69)	2.60 (0.52)	3.00 (0.71)
Benevolent sexism	3.61 (0.54)	3.61 (0.72)	3.92 (0.48)	3.34 (0.81)

Note. In Study 1, hostile and benevolent sexism were measured using a 0 (*strongly disagree*) to 5 (*strongly agree*) scale. In Study 2, hostile and benevolent sexism were measured using a 1 (*strongly disagree*) to 6 (*strongly agree*) scale.

Table 2. Bias and accuracy estimates of hostile and benevolent sexism in Study 1.

Bias and accuracy	Overall judgment				Women's judgment			Men's judgment		
	<i>b</i>	<i>SE</i>	<i>t</i>	Gender diff. <i>t</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>b</i>	<i>SE</i>	<i>t</i>
Hostile sexism										
Bias	.07	0.10	0.73	8.53***	.95	0.15	6.46***	-.81	0.13	-6.02***
Accuracy	.01	0.10	0.09	0.51	.06	0.17	0.36	-.04	0.12	-0.37
Benevolent sexism										
Bias	.22	0.09	2.49*	-5.01***	-.19	0.12	-1.56	.63	0.12	5.45***
Accuracy	.07	0.10	0.71	-1.50	-.08	0.16	-0.49	.21	0.11	1.87 ⁺

Note. Gender diff. *t* = Gender differences in bias and accuracy with bias showing the main effect of gender and accuracy showing the interaction effect.

⁺*p* < .10. **p* < .05. ****p* < .001.

or accurate in judging their partners' hostile sexism. Importantly, and relevant to the hypotheses, there was a main effect of gender, $b = .88$, $t(21.34) = 8.53$, $p < .001$. A Gender x Accuracy interaction was not significant, $b = .05$, $t(49.57) = 0.51$, $p = .615$.

To examine Hypotheses 1 and 2, the effect of gender on judgment was disentangled using recommendations by Aiken and West (1991) in which the group of interest was recoded as zero in the analyses; thus, all other effects in the model would refer to that group. For women, the intercept (bias) was significant in the positive direction, $b = .95$, $t(27.84) = 6.46$, $p < .001$, but it was significant in the negative direction for men, $b = -.81$, $t(28.00) = -6.02$, $p < .001$. In other words, women overestimated their male partners' hostile sexism while men underestimated their female partners' hostile sexism, confirming Hypotheses 1 and 2 respectively.

As for benevolent sexism, participants showed a bias for overestimation, $b = .22$, $t(27.22) = 2.49$, $p = .019$, but accuracy was not obtained, $b = .07$, $t(49.93) = 0.71$, $p = .480$. There was a main effect of gender, $b = -.41$, $t(27.08) = -5.01$, $p < .001$. There was not a significant interaction between gender and accuracy, $b = -.15$, $t(51.25) = -1.50$, $p = .141$.

The effect of gender on the nature of bias was disentangled using the same procedure as before. Contrary to Hypothesis 3, the intercept was not significant for women, $b = -.19$, $t(27.98) = -1.56$, $p = .130$, but the direction suggested that women were indeed underestimating men's benevolent sexism. Finally, Hypothesis 4 was supported such that men overestimated women's benevolent sexism, $b = .63$, $t(28.02) = 5.45$, $p < .001$.

Using mixed-gender partner ratings, we were able to replicate Rudman and Fetterolf's (2014) findings for the most part. What Rudman and

Fetterolf found for intergroup judgment (i.e., judging groups in the abstract) also held when based on actual interpersonal behavior. Specifically, women overestimated their male partners' hostile sexism but underestimated (although not significantly) their benevolent sexism. Men, on the other hand, underestimated their female partners' hostile sexism and overestimated their benevolent sexism. Accuracy at the group level was not observed in general and the interaction between gender and accuracy was not significant for either form of sexism.

Study 2

Study 1 provided a conceptual replication of Rudman and Fetterolf (2014), except Hypothesis 3 was not supported such that women were not underestimating men's benevolent sexism. Furthermore, accuracy was not found at the group level. The lack of accuracy may have been due to lack of contextual information as the conversation centered on a very mundane, getting-acquainted task and perceivers had no other targets to compare the partner to. Previous research on accuracy of judging racial attitudes used more racially salient topics such as racial profiling (Richeson & Shelton, 2005). Therefore, Study 2 differed from Study 1 and Rudman and Fetterolf (2014) by using a more gender-salient conversational topic: distribution of housework. This topic was chosen because it was not overtly about sexism (which might make judgment of sexism too easy) yet would be likely to invite a discussion of traditional gender roles, which might bring out sexist attitudes if they exist. We predicted that when given more contextual information, participants would be more accurate and less biased in judging their partners' sexism.

Method

Participants. Participants were 29 mixed-gender dyads ($M_{\text{age}} = 18.49$) who participated as part of a larger study on a different topic. All of the participants were Northeastern University undergraduates who completed the study in exchange for

partial course credit. Originally there were 31 dyads but two were excluded because they knew one another prior to the study.

Procedure. Male and female students were again recruited separately through the participant pool. After a pair of male and female students arrived, they were told that this was a study on accuracy in the judgment of personality. After consenting to the study, participants first completed a battery of questionnaires individually that contained the ASI (1 = *strongly disagree*; 6 = *strongly agree*). Participants then engaged in a free-structured 3-min conversation in which they introduced themselves and discussed any topic they wished. Afterwards, participants were instructed to discuss gender roles at home for another 3 min; they were asked to share their opinions on how their parents distributed housework such as cleaning, cooking, childcare, yard work, or taking out the trash; they were also asked how they would distribute housework if they were to have a family in the future. They were not, however, asked directly to describe their own gender attitudes, as that would make the judgment task too simple. After the second interaction, participants filled out the ASI as they thought their partner would complete it. Unlike Study 1, interactions were not video recorded.

Results and Discussion

The analytic procedure was the same as Study 1. See Table 1 for descriptive statistics and Table 3 for the estimates in the truth and bias model.

In judging partners' hostile sexism, participants did not exhibit bias, $b = -.07$, $t(21.04) = -0.97$, $p = .342$. Interestingly, accuracy was significant in the negative direction, $b = -.25$, $t(52.40) = -2.73$, $p = .009$, implying that people were perceiving hostile sexist partners as not sexist (or perceiving low hostile sexist partners as overly sexist). There was a main effect of gender, $b = .47$, $t(20.78) = 7.69$, $p < .001$, without a Gender x Accuracy interaction, $b = .04$, $t(53.64) = 0.41$, $p = .682$.

Hypotheses 1 and 2 tested women and men's judgments of their partners' hostile sexism; the

Table 3. Bias and accuracy estimates of hostile and benevolent sexism in Study 2.

Bias and accuracy	Overall judgment				Women's judgment			Men's judgment		
	<i>b</i>	<i>SE</i>	<i>t</i>	Gender diff. <i>t</i>	<i>b</i>	<i>SE</i>	<i>t</i>	<i>b</i>	<i>SE</i>	<i>t</i>
Hostile sexism										
Bias	-.07	0.07	-0.97	7.69***	.40	0.09	4.59***	-.53	0.09	-5.76***
Accuracy	-.25	0.09	-2.73**	0.41	-.21	0.13	-1.66	-.29	0.13	-2.19*
Benevolent sexism										
Bias	.29	0.07	4.08***	-2.11*	.16	0.10	1.51	.43	0.09	4.76***
Accuracy	-.10	0.09	-1.11	-0.26	-.12	0.14	-0.87	-.08	0.11	-0.67

Note. Gender diff. *t* = Gender differences in bias and accuracy with bias showing the main effect of gender and accuracy showing the interaction effect.

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

main effect of gender was again disentangled using recommendations proposed by Aiken and West (1991). Women overestimated men's hostile sexism, $b = .40$, $t(26.79) = 4.59$, $p < .001$. In contrast, men underestimated women's hostile sexism, $b = -.53$, $t(27.00) = -5.76$, $p < .001$. Hypotheses 1 and 2 were again supported.

As for benevolent sexism, participants again overestimated their partners' self-reported sexism level, $b = .29$, $t(23.71) = 4.08$, $p < .001$. Accuracy was not significant overall, $b = -.10$, $t(49.21) = -1.11$, $p = .274$. The main effect of gender was found, $b = -.14$, $t(23.66) = -2.11$, $p = .046$, but it did not interact with accuracy, $b = -.02$, $t(51.31) = -0.26$, $p = .793$.

To disentangle the main effect of gender, Hypotheses 3 and 4 tested women and men's judgments of their partners' benevolent sexism. Contrary to Hypothesis 3, women did not show bias in their judgments of partners' benevolent sexism, $b = .16$, $t(26.88) = 1.51$, $p = .143$, and the direction hinted at a possible overestimation. Hypothesis 4 was supported such that men overestimated their female partners' benevolent sexism, $b = .43$, $t(27.01) = 4.76$, $p < .001$.

Through a discussion paradigm that was more gender-salient, we were able to provide a conceptual replication of Rudman and Fetterolf's (2014) findings and those of Study 1. Women overestimated their male partners' hostile sexism but bias in estimation of their partners' benevolent sexism was not significant. Men underestimated their

female partners' hostile sexism and overestimated their benevolent sexism. It should be noted that the effects were generally smaller than in Study 1, and this may be in part due to more information obtained through the discussion as participants may have been able to glean more information regarding their partners' gender attitudes, leading to less bias. Indeed, the focus of the conversation was on division of housework, a particularly relevant property of benevolent sexism (Glick, Diebold, Bailey-Werner, & Zhu, 1997), and the direction of bias was reversed from Study 1 such that women seemed to be overestimating men's benevolent sexism. However, given that Hypothesis 3 was not supported in Study 1 or Study 2, interpretation should be made with caution.

Study 3

One major limitation of the first two studies and that of Rudman and Fetterolf (2014) is the lack of same-gender judgments. It is unclear whether accuracy exists when men judge other men's sexism or when women judge other women's sexism. Thus, a goal of Study 3 was to include these comparisons. Another limitation of the first two studies was the inability to measure accuracy except at the dyadic level. Accuracy at the dyadic level cannot be unambiguously attributed to perceivers because of the confounding with partner expressivity as discussed before (and elsewhere, e.g., Hall et al., 2015).

Therefore, we recruited new participants in Study 3 to measure accuracy of detecting sexism based on watching multiple targets' behaviors as this method circumvents the confounding and also allows for individual perceivers to be scored for their accuracy. Participants watched either male or female targets from Study 1 and completed either benevolent sexism or hostile sexism items for each target as they thought the person in the video would fill them out. This was a 2 (participant gender: male vs. female) \times 2 (target video gender: male vs. female) \times 2 (sexism judgment condition: benevolent vs. hostile) between-subjects design.

Data were analyzed using sensitivity correlation to examine accuracy in detecting who is more sexist than whom across a number of male or female targets. It should be noted that the truth and bias model cannot be applied herein because the model specifies that the truth or the criterion cannot be a constant across perceivers (Stern et al., 2013; West & Kenny, 2011).

Method

Participants and procedure. Participants were 165 undergraduates (50.30% female; $M_{\text{age}} = 18.90$) who participated for partial course credit. Four participants were previously excluded for being underage at the time of participation without parental consent or for circling the same response for all ratings.

Participants came in groups of one to four. Each group watched videos of 27 male targets or 27 female targets from Study 1. Within each group, participants were randomly assigned to rate the targets' benevolent sexism or hostile sexism. After signing consent forms, participants were given a packet of sexism ratings. They were instructed to rate the target on the sexism scales provided after each video clip, as they thought the target would fill them out. The experimenter paused at the end of each video to allow time for ratings and then played the next video. Once participants had finished rating all targets, they completed a feminism scale and demographic information, and were debriefed.

Videos. Videos from Study 1 were edited so that only the male or the female targets were visible

and consisted of only the middle 30 s of the 3-min unstructured social interaction. Because of camera malfunction and not all targets agreed to release their videos, 27 rather than 30 dyads were on the tape. The videos were presented according to the order the targets had originally come into the lab. As mentioned earlier, conversations centered on getting acquainted, classes, and the trivia game they had played earlier in the interaction. None of the targets mentioned or discussed topics concerning sexism.

Sexism ratings. Given the number of targets, a shortened ASI was used for sexism ratings consisting of six items for benevolent sexism and six items for hostile sexism. Previous research demonstrated that these items had strong reliability and correlated highly with the full 22-item ASI (Hammond & Overall, 2013). Items were rated on a 6-point scale as in Study 1 (0 = *disagree strongly*; 5 = *agree strongly*). After appropriate reverse-scoring, participants' ratings were averaged to form a benevolent or a hostile sexism score for each target.

Feminism scale. Participants completed the 10-item short form of Liberal Feminist Attitude and Ideology Scale (Morgan, 1996). Example items include "A woman should have the same job opportunities as a man" and "Women should be considered as seriously as men as candidates for the Presidency of the United States." Two additional items were included to measure identification with being a feminist (Breen & Karpinski, 2008): "I consider myself a feminist" and "My friends and family consider me a feminist." All items were measured on a 7-point scale (1 = *disagree strongly*; 7 = *agree strongly*), and higher averaged scores indicate greater endorsement of feminist attitude and feminist identity after appropriate reversals. Neither feminist attitude nor feminist identity correlated with accuracy in detecting hostile and benevolent sexism ($r_s < .18$; $p_s > .11$).

Calculation of Accuracy Scores

Each participant either watched 27 videos of male targets or 27 videos of female targets from

Table 4. Means and standard deviations of sexism detection accuracy scores in Study 3.

Participant gender	Target gender	Hostile sexism		Benevolent sexism	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Female	Male	0.10*	0.20	0.11**	0.17
Male	Female	0.00	0.17	0.03	0.22
Female	Female	-0.01	0.21	0.17***	0.21
Male	Male	0.13**	0.20	0.23***	0.20

Note. Significance as shown by the asterisks represents one-sample *t* test against chance (zero).

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

the dyads in Study 1, and each participant either rated each of the 27 targets on the Benevolent Sexism Scale or the Hostile Sexism Scale. Each participant's ratings of the targets' sexism scores were correlated with those targets' self-reported sexism scores, across targets, yielding a Pearson correlation for each participant. This constitutes that participant's accuracy in identifying different degrees of sexism across the targets. Each Pearson correlation was Fisher-*z* transformed for normalization before analysis and converted back to the Pearson correlation metric for presentation. By calculating a correlation score for each participant, individual variation in accuracy can be obtained. A positive correlation indicates that a person has some ability to distinguish between the targets. A correlation of zero means there is no covariation and that the judge must have been guessing. This method is commonly used in interpersonal accuracy research and has been applied in stereotype accuracy research (e.g., Dickman et al., 2002; Judd & Park, 1993).

Results

Overall accuracy for hostile and benevolent sexism. To find out whether, in general, participants were accurate about the targets' sexism, one-sample *t* tests against the null (no accuracy) value of zero were used. Averaging over gender of participants and targets, both kinds of sexism were judged at levels significantly above chance: for benevolent sexism ($M = 0.13$; $SD = 0.21$), $t(82) = 5.75$, $p < .001$; for hostile sexism ($M = 0.05$; $SD = 0.20$), $t(81) = 2.46$, $p = .016$. An independent samples *t* test was used to test whether the two kinds of

sexism were judged with equal accuracy. Benevolent sexism was more accurately judged than hostile sexism, $t(163) = 2.44$, $p = .016$.

Next, analyses were performed to examine how the gender of the participant and the target influenced accuracy separately for hostile and benevolent sexism.

Hostile sexism: Target and participant gender comparisons. Women were significantly accurate at judging male targets' hostile sexism, $t(20) = 2.21$, $p = .039$. Men, on the other hand, were not accurate above chance at judging female targets' hostile sexism, $t(20) = 0.00$, *ns*. Turning to same-gender ratings, women were not significantly accurate at judging female targets' hostile sexism, $t(19) = 0.11$, *ns*, while men were significantly accurate at judging male targets' hostile sexism, $t(19) = 2.82$, $p = .011$ (see Table 4).

When accuracy scores were analyzed in a 2 (participant gender) \times 2 (target gender) ANOVA, a main effect of target gender emerged, $F(1, 78) = 7.03$, $p = .010$, such that participants were more accurate at judging male targets' hostile sexism ($M = 0.11$; $SD = 0.20$) than they were at judging female targets' hostile sexism ($M = 0.00$; $SD = 0.19$). No other effects were significant, $p_s > .67$.

Benevolent sexism: Target and participant gender comparisons. Women were significantly accurate (i.e., above guessing) at judging male targets' benevolent sexism, $t(21) = 2.94$, $p = .008$. Men, in contrast, were not accurate above chance at judging female targets' benevolent sexism, $t(20) = 0.61$, *ns*. As for same-gender ratings, women were significantly accurate at judging female targets'

benevolent sexism, $t(19) = 3.76, p = .001$. Men were also accurate at detecting male targets' benevolent sexism, $t(19) = 5.13, p < .001$ (see Table 4).

The ANOVA showed there was a significant interaction of participant gender and target gender, $F(1, 79) = 9.19, p = .003$. Pairwise comparisons showed that for judging male targets, men ($M = 0.23; SD = 0.20$) were more accurate than women ($M = 0.11; SD = 0.17$), $t(40) = 2.16, p = .037$. For female targets, women ($M = 0.17; SD = 0.21$) were more accurate than men ($M = 0.03; SD = 0.22$), $t(39) = 2.14, p = .039$. The main effects were not significant, $ps > .13$.

Discussion

In general, participants were significantly accurate at detecting hostile and benevolent sexism from 30-s video clips. However, participant and target gender both play important roles as well as did the form of sexism being judged.

For hostile sexism, men and women had significantly above-chance accuracy only for judging male targets. This corresponds to the factorial analyses as there was a main effect of target gender such that both male and female participants were able to detect male targets' hostile sexism better than female targets' hostile sexism. Although women do hold sexist beliefs regarding their own gender, people may not necessarily know what or how to look for women's hostile sexist beliefs toward other women. If people do not know what hostile sexism held by women should look like, then they would most likely be guessing. Indeed, accuracy scores were around zero (guessing level) for men and women when judging female targets' hostile sexism.

When judging benevolent sexism, participants were more accurate when judging targets of the same gender than opposite gender. In other words, there was an ingroup advantage to accurately detecting benevolent sexism. This ingroup advantage is consistent with several findings in the interpersonal accuracy literature, which have shown that people are more accurate at judging targets from the same cultural group (e.g.,

Elfenbein & Ambady, 2002; Rogers & Biesanz, 2014). Because males and females often socialize within their own gender group, this could create an ingroup advantage if people reveal their attitudes more freely within their ingroup, allowing for more learning of how those attitudes are expressed. Why there was ingroup advantage for benevolent but not hostile sexism is unclear.

Importantly, one sample t tests revealed that when participants were judging sexism in opposite-gender targets, women judged male targets significantly above chance for both hostile and benevolent sexism. This effect was not observed in men, whose judgments of female targets were not significantly different from chance for either form of sexism. The ability of women to detect sexism in men (but not vice versa) lends credence to functional theories (Richeson & Shelton, 2005), which suggest that social minority members become more accurate because accuracy serves adaptive behavioral responses.

General Discussion

The present studies added to the literature on the perception of sexism in several ways. Both bias and accuracy were measured, and accuracy was measured using two different methodologies, and we used dyadic interaction and videos of real targets rather than having participants imagine the attitudes of the opposite gender, thus extending the generalizability of Rudman and Fetterolf's (2014) findings to a more real-life context.

Studies 1 and 2 employed the truth and bias model to examine bias in judging partner's sexism in mixed-gender dyadic interactions. We largely replicated and extended Rudman and Fetterolf's findings. Both men and women were biased in judging their partners' sexism level. Specifically, women overestimated men's hostile sexism. Men, in contrast, underestimated women's hostile sexism but overestimated their benevolent sexism. However, women's underestimation of benevolent sexism was not replicated. This may in part be due to low power as there were only approximately 30 dyads in both studies, while Rudman and Fetterolf (2014) had a larger sample size.

Accuracy at the group (not individual) level was also examined in Studies 1 and 2 using the truth and bias model. Significant accuracy was not found in either study and there was indication of significant inaccuracy in Study 2. Future research could delve into the question of group-level accuracy deeper and one improvement on the current research design would be asking participants to judge their partners at multiple time points (e.g., beginning, middle, and end of study). Multiple time points allow for more statistical power and one can track accuracy of judgments across time points (e.g., Overall, Fletcher, & Kenny, 2012; West, Dovidio, & Pearson, 2014). Another improvement is using multiple targets, which is what we did in Study 3.

Study 3 used a more sensitive and individual-level method for calculating accuracy by having participants judge multiple targets. There was significant accuracy overall for judging hostile and benevolent sexism, but there were important differences according to combinations of participant and target gender and which form of sexism was being judged. Both male and female participants were more accurate at detecting hostile sexism in male targets than female targets. As for benevolent sexism, male participants were more accurate than female participants in detecting male targets' benevolent sexism, but female participants were more accurate than male participants in detecting female targets' sexism. When participants were judging sexism in opposite-gender targets, women were able to judge male targets significantly above chance for both hostile and benevolent sexism, but men's judgments of female targets did not differ significantly from chance for either form of sexism. However, accuracy was not dramatically high in an absolute sense, with accuracy means of .13 for benevolent sexism and .05 for hostile sexism when theoretically they could range up to 1.00. Nevertheless, accuracy is typically low when judged from brief excerpts of strangers' spontaneous (i.e., not posed) behavior—for example, judging racism (Richeson & Shelton, 2005), intelligence (Murphy, Hall, & Colvin, 2003), personality traits (Connelly & Ones, 2010), thoughts and feelings (Gesn &

Ickes, 1999; Hall & Schmid Mast, 2007), preferences for pictures (North, Todorov, & Osherson, 2012), and physical pain (Ruben & Hall, 2013).

Rudman and Fetterolf (2014) concluded that there was no accuracy in their paper, but this conclusion should be modified to pertain only to the bias question and not to accuracy in the correlational sense. Both approaches have value but likely in different contexts. The bias question may be especially relevant when thinking about social groups and when the context does not allow the opportunity to make target-specific ratings, as was the case in Rudman and Fetterolf's study. While group-level accuracy was not significant in Studies 1 and 2, accuracy defined as the ability to distinguish sexism between targets was significant in Study 3. After all, the methodologies for calculating accuracy are quite different with the former (group-level accuracy) being one coefficient calculated for the group while the latter is an individual-differences measure. The former measures covariation between judgments and criteria across dyads (expressed as a regression coefficient), but the latter measures whether a judge can discern one target from another in terms of their sexism. Furthermore, there is an expressivity-perceptivity confound inherent with accuracy obtained from dyadic interactions (e.g., Hall et al., 2006, 2015; Snodgrass et al., 1998). The lack of accuracy may reflect a person's inability to detect sexism in others, or it may be due to the interaction partner's expressivity; if the partner is not expressing enough cues that could point to his or her sexism level, then making accurate judgments would be a very difficult task. Study 3 circumvented the expressivity-perceptivity confound and contributed novel results to this literature by measuring accuracy in terms of distinguishing between targets and comparing same-gender and opposite-gender pairings. Overall, accuracy was above chance even though the stimuli were only 30 s long.

What could have contributed to these biased and accurate judgments of sexism? Functional theories suggest that accuracy and bias serve adaptive function for the perceivers (e.g., Fletcher, 2015; Richeson & Shelton, 2005; Zebrowitz & Collins, 1997). It is advantageous for women to

overestimate men's hostile sexism in order to be on guard, particularly against men whom they have just met. They may figure it is better to be safe than sorry when estimating male strangers' hostile sexism, which is a more overt and aggressive form of sexism than benevolent sexism. But, even while overestimating in this way, it is still adaptive for women to be able to accurately distinguish which men are more sexist than other men. Indeed, Study 3 found that women were able to accurately detect both hostile and benevolent sexism in men significantly above chance, but men were not accurate in judging women. Men do not necessarily need to be accurate when judging women's sexism as such judgments do not serve an obvious adaptive function for men because men typically hold more power and status in society than women and do not necessarily need to protect themselves against women's sexist attitudes against women. The same logic would suggest that if one were to measure sexism against *men* (e.g., using the Ambivalence Toward Men Inventory; Glick & Fiske, 1999), then men would demonstrate accuracy in judging female targets' sexism.

This research is, however, limited within the context of first impressions between men and women who did not know one another. This was particularly the case in Study 3 where participants had only half a minute of video on which to judge the sexism of a particular target person. In contexts where dyads have more familiarity with one another, differences may arise. For instance, accuracy in personality judgments increases when judgments are made by familiar others such as friends and family compared to strangers (for a meta-analytic review, see Connelly & Ones, 2010). One could imagine how bias and accuracy in judging hostile and benevolent sexism may differ in more intimate contexts such as romantic relationships, considering that intimacy can be deeply entwined with both forms of sexism (Rudman & Glick, 2008). In romantic relationships, men and women have more information and opportunity to appraise their partners' sexism. Hammond and Overall (2013) found that men's hostile sexism predicted overestimation of their romantic partners' negative behaviors (i.e., thinking the partners

were behaving more negatively than the partners' self-report), which subsequently led to poorer relationship quality and more negative behavior on the men's part. While they did not specifically examine accuracy and bias in judging partners' sexism, they demonstrated that biased perceptions of partners' behavior can have adverse consequences in an intimate context, especially among hostile sexist men. Research examining the different contexts whereby sexism judgments can differ is important and merits serious consideration. Nonetheless, the context of first impressions is equally important as the beginning of any intimate romantic relationships relies on first impressions.

The field of interpersonal accuracy is growing, but in order for the field to continue to proliferate, more explorations into bias and accuracy are needed. Both bias and accuracy play tremendous roles in our everyday lives as we regularly interact, perceive, and judge people, whether they are intimate or not. Prejudice is one domain that is beginning to examine both bias and accuracy, and the current research adds to this budding field by showing that people can still be accurate in detecting sexism even through a limited 30-s window of social interactions between two people conversing on mundane topics. Although research has extensively studied sexism at an intergroup level, the interpersonal processes of sexism in mixed-gender interactions are seldom examined. And as we continue to make strides in gender equality in our society, care should be given to address the everyday, ordinary social interactions between men and women where sexism still lurks.

Acknowledgements

We thank Tessa West for her generous assistance with data analyses and revision of the paper. We also thank Amanda Falcon for her assistance in conducting the study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Study 1 was supported by the Clara Mayo Grant from the Society for the Psychological Study of Social Issues to the first author.

References

- Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*, 518–529. doi:10.1037/0022-3514.69.3.518
- Back, M. D., & Nestler, S. (2016). Judging personality. In J. A. Hall, M. Schmid Mast & T. V. West (Eds.), *The social psychology of perceiving others accurately*. Cambridge, UK: Cambridge University Press.
- Becker, J. C., & Wright, S. C. (2011). Yet another dark side of chivalry: Benevolent sexism undermines and hostile sexism motivates collective action for social change. *Journal of Personality and Social Psychology, 101*, 62–77. doi:10.1037/a0022615
- Breen, A. B., & Karpinski, A. (2008). What's in a name? Two approaches to evaluating the label feminist. *Sex Roles, 58*, 299–310. doi:10.1007/s11199-007-9317-y
- Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality, 41*, 1054–1072. doi:10.1016/j.jrp.2007.01.004
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychology Bulletin, 136*, 1092–1122. doi:10.1037/a0021212
- Diekmann, A. B., Eagly, A. H., & Kulesa, P. (2002). Accuracy and bias in stereotypes about the social and political attitudes of women and men. *Journal of Experimental Social Psychology, 38*, 268–282. doi:10.1006/jesp.2001.1511
- Edmonds, E. M., Cahoon, D. D., & Shipman, M. (1991). Predictions of opposite-sex attitudes concerning gender-related social issues. *Bulletin of the Psychonomic Society, 29*, 295–296. doi:10.3758/BF03333923
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203–235. doi:10.1037/0033-2909.128.2.203
- Fletcher, G. J. O. (2015). Accuracy and bias of judgments in romantic relationships. *Current Directions in Psychological Science, 24*, 292–297. doi:10.1177/0963721415571664
- Fletcher, G. J. O., & Kerr, P. S. G. (2010). Through the eyes of love: Reality and illusion in intimate relationships. *Psychological Bulletin, 136*, 627–658. doi:10.1037/a0019792
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149–158. doi:10.1037/0022-3514.55.1.149
- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology, 77*, 746–761. doi:10.1037/0022-3514.77.4.746
- Glick, P., Diebold, J., Bailey-Werner, B., & Zhu, L. (1997). The two faces of Adam: Ambivalent sexism and polarized attitudes toward women. *Personality and Social Psychology Bulletin, 23*, 1323–1334. doi:10.1177/01461672972312009
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*, 491–512. doi:10.1037/0022-3514.70.3.491
- Glick, P., & Fiske, S. T. (1999). The Ambivalence Toward Men Inventory: Differentiating hostile and benevolent beliefs about men. *Psychology of Women Quarterly, 23*, 519–536. doi:10.1111/j.1471-6402.1999.tb00379.x
- Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist, 56*, 109–118. doi:10.1037/0003-066X.56.2.109
- Glick, P., Fiske, S. T., Mladinic, A., Saiz, J., Abrams, D., Masser, B., . . . López, W. L. (2000). Beyond prejudice as simple antipathy: Hostile and benevolent sexism across cultures. *Journal of Personality and Social Psychology, 79*, 763–775. doi:10.1037/0022-3514.79.5.763
- Goh, J. X., & Hall, J. A. (2015). Nonverbal and verbal expressions of men's sexism in mixed-gender interactions. *Sex Roles, 72*, 252–261. doi:10.1007/s11199-015-0451-7
- Good, J. J., & Rudman, L. A. (2010). When female applicants meet sexist interviewers: The costs of being a target of benevolent sexism. *Sex Roles, 62*, 481–493. doi:10.1007/s11199-009-9685-6
- Hall, J. A., Bernieri, F. J., & Carney, D. R. (2005). Nonverbal behavior and interpersonal sensitivity. In J. A. Harrigan, R. Rosenthal & K. R. Scherer (Eds.), *The new handbook of methods in nonverbal behavior*

- research (pp. 237–281). Oxford, UK: Oxford University Press.
- Hall, J. A., Rosip, J. C., Smith LeBeau, L., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology, 42*, 18–27. doi:10.1016/j.jesp.2005.01.005
- Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion, 7*, 438–446. doi:10.1037/1528-3542.7.2.438
- Hall, J. A., Schmid Mast, M., & Latu, I. (2015). The vertical dimension of social relations and accurate interpersonal perception: A meta-analysis. *Journal of Nonverbal Behavior, 39*, 131–163. doi:10.1007/s10919-014-0205-1
- Hall, J. A., Stein, T. S., Roter, D. L., & Rieser, N. (1999). Inaccuracies in physicians' perceptions of their patients. *Medical Care, 37*, 1164–1168. doi:10.1097/00005650-199911000-00008
- Hammond, M. D., & Overall, N. C. (2013). Men's hostile sexism and biased perceptions of intimate partners: Fostering dissatisfaction and negative behavior in close relationships. *Personality and Social Psychology Bulletin, 39*, 1585–1599. doi:10.1177/0146167213499026
- Hammond, M. D., & Overall, N. C. (2015). Benevolent sexism and support of romantic partner's goals: Undermining women's competence while fulfilling men's intimacy needs. *Personality and Social Psychology Bulletin, 41*, 1180–1194. doi:10.1177/0146167215593492
- Hehman, R., Leitner, J. B., Deegan, M. P., & Gaertner, S. L. (2013). Facial structure is indicative of explicit support for prejudicial beliefs. *Psychological Science, 24*, 289–296. doi:10.1177/0956797612451467
- 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. doi:10.1037/0022-3514.88.3.498
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100*, 109–128. doi:10.1037/0033-295X.100.1.109
- Kilianski, S. E., & Rudman, L. A. (1998). Wanting it both ways: Do women approve of benevolent sexism? *Sex Roles, 39*, 333–352. doi:10.1023/A:1018814924402
- LaCosse, J., Tuscherer, T., Kunstman, J. W., Plant, E. A., Trawalter, S., & Major, B. (2015). Suspicion of White people's motives relates to relative accuracy in detecting external motivation to respond without prejudice. *Journal of Experimental Social Psychology, 61*, 1–4. doi:10.1016/j.jesp.2015.06.003
- Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kooken, K., Ekman, P., . . . Goh, A. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman's Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior, 24*, 179–209. doi:10.1023/A:1006668120583
- Morgan, B. L. (1996). Putting the feminism into feminism scales: Introduction of a Liberal Feminist Attitude and Ideology Scale (LFAIS). *Sex Roles, 34*, 359–390. doi:10.1007/BF01547807
- Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality, 71*, 465–493. doi:10.1111/1467-6494.7103008
- North, M. S., Todorov, A., & Osherson, D. N. (2012). Accuracy of inferring self- and other-preferences from spontaneous facial expressions. *Journal of Nonverbal Behavior, 36*, 227–233. doi:10.1007/s10919-012-0137-6
- Overall, N. C., Fletcher, G. J., & Kenny, D. A. (2012). When bias and insecurity promote accuracy: Mean-level bias and tracking accuracy in couples' conflict discussions. *Personality and Social Psychology Bulletin, 38*, 642–655. doi:10.1177/0146167211432764
- Richeson, J. A., & Shelton, J. N. (2005). Brief report: Thin slices of racial bias. *Journal of Nonverbal Behavior, 29*, 75–86. doi:10.1007/s10919-004-0890-2
- Rogers, K. H., & Biesanz, J. C. (2014). The accuracy and bias of interpersonal perceptions in intergroup interactions. *Social Psychological and Personality Science, 5*, 918–926. doi:10.1177/1948550614537307
- Ruben, M. A., & Hall, J. A. (2013). "I know your pain": Proximal and distal predictors of pain detection accuracy. *Personality and Social Psychology Bulletin, 39*, 1346–1358. doi:10.1177/0146167213493188
- Rudman, L. A., & Fetterolf, J. C. (2014). How accurate are metaperceptions of sexism? Evidence for the illusion of antagonism between hostile and benevolent sexism. *Group Processes & Intergroup Relations, 17*, 275–285. doi:10.1177/1368430213517272
- Rudman, L. A., & Glick, P. (2008). *The social psychology of gender: How power and intimacy shape gender relations*. New York, NY: Guilford.

- Snodgrass, S. E., Hecht, M. A., & Ploutz-Snyder, R. (1998). Interpersonal sensitivity: Expressivity or perceptivity? *Journal of Personality and Social Psychology, 74*, 238–249. doi:10.1037/0022-3514.74.1.238
- Stern, C., West, T. V., & Schoenthaler, A. (2013). The dynamic relationship between accuracy and bias in social perception research. *Social and Personality Psychology Compass, 7*, 303–314. doi:10.1111/spc3.12024
- Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups: A review and meta-analysis. *Personality and Social Psychology Review, 17*, 72–86. doi:10.1177/1088868312461308
- West, T. V., Dovidio, J. F., & Pearson, A. R. (2014). Accuracy and bias in perceptions of relationship interest for intergroup and intragroup roommates. *Social Psychological and Personality Science, 5*, 235–242. doi:10.1177/1948550613490966
- West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review, 118*, 357–378. doi:10.1037/a0022936
- Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review, 1*, 204–223. doi:10.1207/s15327957pspr0103_2