



HHS Public Access

Author manuscript

J Exp Soc Psychol. Author manuscript; available in PMC 2018 September 01.

Published in final edited form as:

J Exp Soc Psychol. 2017 September ; 72: 101–117. doi:10.1016/j.jesp.2017.04.006.

Value-Based Standards Guide Sexism Inferences for Self and Others

Chelsea Mitamura, Lynnsey Erickson, and Patricia G. Devine

University of Wisconsin-Madison

Abstract

People often disagree about what constitutes sexism, and these disagreements can be both socially and legally consequential. It is unclear, however, why or how people come to different conclusions about whether something or someone is sexist. Previous research on judgments about sexism has focused on the perceiver's gender and attitudes, but neither of these variables identifies comparative standards that people use to determine whether any given behavior (or person) is sexist. Extending Devine and colleagues' values framework (Devine, Monteith, Zuwerink, & Elliot, 1991; Plant & Devine, 1998), we argue that, when evaluating others' behavior, perceivers rely on the morally-prescriptive values that guide their own behavior toward women. In a series of 3 studies we demonstrate that (1) people's personal standards for sexism in their own and others' behavior are each related to their values regarding sexism, (2) these values predict how much behavioral evidence people need to infer sexism, and (3) people with stringent, but not lenient, value-based standards get angry and try to regulate a sexist perpetrator's behavior to reduce sexism. Furthermore, these personal values are related to all outcomes in the present work above and beyond other person characteristics previously used to predict sexism inferences. We discuss the implications of differing value-based standards for explaining and reconciling disputes over what constitutes sexist behavior.

Keywords

sexism; standards; inferences; values

Post: I'm not sexist but I really think I'm a better driver than every girl...

#sorrynotsorry #learnhowtodrive

Address correspondence for this project to Chelsea Mitamura at ChelseaMitamura@gmail.com. Chelsea Mitamura and Patricia G. Devine, Psychology Department, University of Wisconsin-Madison, 1202 West Johnson Street, Madison, WI, 53706.

Conceived research: Mitamura, Erickson, Devine; Designed research: Mitamura, Erickson, Devine; Coordinated data collection: Mitamura, Erickson; Analyzed data: Mitamura; Wrote paper: all authors; Revised paper: all authors.

Data and materials for this project can be found at: <https://osf.io/qgnzd/>

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Response: You're not sexist but I really think you are. #sorrynotsorry
#learnhowtothink (notsexistbut.tumblr.com)

Addressing sexism in our contemporary society can be both challenging and confusing. Although sexism has become increasingly socially unacceptable in recent years (Swim, Aikin, Hall, & Hunter, 1995), disputes rage on about what is and is not considered "sexist." The exchange above exemplifies the type of strident debates about sexism that occur frequently in daily life. Given that behavior labeled as sexist is often subject to social (e.g., Waxman, 2015) and even legal sanctions (e.g., *Hopkins v. Price Waterhouse*, 1985; *Wal-mart Stores, Inc. v. Dukes*, 2011), determining what does and does not qualify as sexism constitutes a critically important undertaking.

The scientific literature, which is replete with examples of disagreements about the extent to which any given behavior is sexist or innocuous, adds to the confusion (Brant, Mynatt, & Doherty, 1999; Smith, 1992; Swim, Mallett, Russo-Devosa, & Stangor, 2005; Trenholm & Manciallas, 1978). Indeed, the ambiguities surrounding "sexism" are reminiscent of the ambiguity surrounding the definition of "obscenity" when Supreme Court Justice Potter Stewart famously stated, "*I know it when I see it...*" regarding potential obscenity in Louis Malle's film *The Lovers* (*Jacobellis v. Ohio*, 1964). Rather than formalize a set of criteria to determine whether the film qualified as obscene, Justice Stewart, writing for the majority, relied on his personal perspective to conclude that the film was not obscene. The Court, however, was not unanimous and the justices who prepared the minority opinion arrived at the opposite conclusion. In considering contemporary struggles to define sexism, it appears that sexism, like obscenity, is effectively in the eye of the beholder. Though people clearly differ in their judgments about the same potentially sexist behaviors, it is unclear why people come to these disparate conclusions.

Sources of variability in judgments of sexism

Some researchers have attempted to explain the variability in judgments of sexism in others' behavior by exploring individual differences, most notably gender and attitudes, that may predispose some people to see sexism more readily than others. Several studies showed, for example, that women are more likely than men to view the same behaviors as sexist and people lower in self-reported levels of sexism are more likely to conclude that specific behaviors are sexist (Blodorn, O'Brien, & Kordys, 2012; Inman & Baron, 1996; Smith, 1992; Swim et al., 2005). Although examining sexism inferences as a function of gender and attitudes describes *who* is more or less likely to infer sexism, these approaches do not fully account for *why* these people see more sexism in others' behavior and they do not directly identify the psychological processes involved in reaching a conclusion of sexism.

According to classic inference models, inferences are made through a comparison process, which requires identifying a standard against which to compare a person's behavior or qualities to determine if he or she possesses the characteristic of interest (Mussweiler, 2003; Trope, 1986). Thus, understanding judgments of sexism requires identifying the standards people use to draw their conclusions. Because neither gender nor attitudes prescribe specific standards for what constitutes sexism, it is necessary to look beyond these characteristics.

Swim and colleagues (2005) speculated that perceivers might use similar criteria to judge other people's behavior as they do when they judge their own behavior (i.e., a behavior viewed as unacceptable for the self would similarly be viewed as unacceptable for others). Though highly plausible, this analysis neither identifies the standards that guide people's inferences for what is appropriate for the self or others, nor what gives rise to observed variability in these standards across people. We argue that it is necessary to address both issues to understand the source of the oft contentious disputes regarding what is and is not considered sexist behavior.

To address these issues, we turned to work by Devine and colleagues that demonstrated that people's personal standards for appropriate behavior for themselves in the prejudice domain are derived from their values concerning how to treat stigmatized others (Devine, Monteith, Zuwerink, & Elliot, 1991; Klonis, Plant & Devine, 2005; Plant & Devine, 1998). Whereas some people have values that prohibit prejudice, others have values that are more accepting of prejudice. Plant and Devine's (1998; Klonis et al., 2005) Internal Motivation to Respond without Prejudice Scale (IMS) captures variability in the extent to which people have internalized the values of being nonprejudiced (or nonsexist) and use these values to set standards for their own behavior.¹ Devine and colleagues' work draws our attention to two aspects of a standard – its location on an acceptability continuum and its importance to the self-concept. The location of a standard (more or less stringent) determines whether behaviors are deemed acceptable or unacceptable. The personal importance of the standard, as reflected in the degree to which the standard is internalized, self-defining and accessible, in turn, influences how people affectively and behaviorally respond to transgressions of the standard.

Devine and colleagues have argued that both the type of affect one experiences and whether self-regulatory efforts to realign one's behavior with one's standards occur, follow from evaluating the morality of the behavior relative to these value-based standards (Devine et al., 1991; Plant & Devine, 1998; see also Higgins, 1987). An abundance of evidence demonstrates that people who score higher in IMS have internalized, chronically accessible egalitarian values that lead them to establish stringent, personally important, nonsexist standards. Thus, when they fail to adhere to these standards, they feel a form of agitated distress, specifically anger directed toward the self (i.e., guilt) and engage in regulatory efforts to reduce the discrepancy between their standards and their behavior (Amodio, Devine, & Harmon-Jones, 2007; Monteith, 1993; Plant & Devine, 1998). People who score lower in IMS, on the other hand, set less stringent standards, allowing for more bias in their behavior. Moreover, their standards are not strongly internalized or highly accessible; even though they often behave with bias exceeding even their more permissive standards, these transgressions are not self-threatening and thus do not elicit guilt or efforts to realign their behavior with their standards.

To date, Devine and colleagues have focused exclusively on the affective and regulatory significance of personal standards for judgments about *one's own* behavior. We argue,

¹Although some of the relevant evidence was demonstrated in the racial prejudice domain, we will use the terms "sexist" and "nonsexist" for simplicity and consistency throughout the present work.

however, that extending this framework to the interpersonal arena may give us leverage to test Swim and colleagues' (2005) speculation that a similar evaluation process occurs when judging others' behavior as when judging one's own behavior. Because people derive the standards they set for their own behavior from their morally-prescriptive values, we reason that they derive the standards they set for other people's behavior from the same set of values. To the extent that this is true, standards for acceptable behavior in the sexism domain should be similarly stringent for the self and others; further, violations of the important, nonsexist standards imposed on others should trigger agitation-related affective reactions, as they do for the self (Higgins, 1987). Specifically, Higgins (1987) argued that when a target violates a perceiver's value-based, moral (ought) standard for others, the perceiver would likely feel anger and resentment directed toward the target. Following Devine and colleagues' work on self-regulation in the prejudice domain, we suspect that to the extent that a target violates perceivers' personally important value-based standard, they will also make efforts to regulate the target's behavior to bring it in line with the standards they set for others (e.g., Monteith, 1993; Plant & Devine, 1998). We test both affective and regulatory hypotheses in the present set of studies.

Before moving on to an overview of the present work, we wish to note that Devine and colleagues' work (Klonis et al., 2005; Plant & Devine, 1998) recognizes the possibility that the standards people impose on others could be derived from social norms, which discourage expressions of sexism and prejudice (Barreto & Ellemers, 2013; Blanchard, Lily, & Vaughn, 1991; Monteith, Deenan & Tooman, 1996; Swim et al., 2005). If people derive the standards they set for others from social norms, we would expect these standards to be stringent and to display little variability. However, it is also true that people vary in their sensitivity to the pressure to respond consistently with nonsexist norms. Klonis et al.'s (2005) External Motivation to Respond without Sexism Scale (EMS) captures variability in the extent to which people are reactive to the pressure imposed by others to respond without sexism. Whereas high EMS people are sensitive to pressure imposed by others to respond without sexism, low EMS people are not. Although EMS has not been related to the location of people's personal standards in prior work (Plant & Devine, 1998), we included it in the present studies to explore its potential role in setting standards for others and predicting people's reactions to a target's sexist behaviors.

The Present Work

The primary goal of the present work was to explore the utility of extending Devine and colleagues' values analysis of the affective and regulatory significance of personal standards for judgments about *one's own* behavior to judgments about *others'* behavior. Testing the validity of this analysis in the interpersonal context requires first that we establish that IMS, an indicator of people's personal values regarding sexism, predicts the standards people set for sexism in others. Beyond setting standards, the values framework suggests that IMS should predict the inferential, affective, and regulatory consequences of encountering potentially sexist behaviors in others. We explore these issues in a series of three studies.

In all three studies, we manipulated the extremity of the sexist behaviors participants responded to, evaluating participants' standards for, inferences from, and reactions to either

moderate or extremely sexist behaviors. In past work (Brant et al., 1999; Swim et al., 2005), people more readily drew inferences of sexism from more extremely sexist behaviors than from moderately sexist behaviors. This pattern could indicate that one's personal standards may be more important when people evaluate moderately sexist behaviors, for which there is more room for interpretational ambiguity. However, even when taking behavior type into account, there was unexplained variability in the extent to which these separate categories of behaviors were viewed as sexist (Brant et al., 1999; Swim et al., 2005), which suggests that it is prudent to examine the impact of IMS with regard to both types of behaviors.

In Study 1, we examined how IMS predicts the location of the personal standards regarding sexism that people set for themselves and the standards people set for others. We anticipated that participants' reports of standards for the self would replicate the pattern reported in previous studies such that high IMS participants would report more stringent standards than their low IMS counterparts (Plant & Devine, 1998). We also expected that people would report more stringent standards for extremely than for moderately sexist behaviors. Finally, if, as we propose, people derive the standards they set for the self and the standards they set for others from the same overarching values, then the stringency of the standards people set for others should also be predicted by IMS and the stringency of the two sets of standards should not differ.

In Study 2, we move beyond reports of the location of standards for sexist behavior to examine how people use their standards to judge sexism in another person. To this end, we employed a social inference task in which participants evaluated sequentially presented extremely or moderately sexist behaviors and were asked to indicate when they had enough evidence to infer that a target was sexist (Biernat, Ma, & Nario-Redmond, 2008). We hypothesized that high IMS people, whose standards are more stringent, would require fewer behaviors to conclude the target was sexist than low IMS people regardless of behavior type. We reasoned, however, that it was possible that IMS could interact with behavior type, such that everyone would require a similarly low number of extremely sexist behaviors to infer sexism but that people high in IMS would require fewer moderately sexist behaviors to infer sexism than would their low IMS counterparts (Monteith et al., 1993).

The tasks used in Studies 1 and 2 require participants to explicitly consult their values, making values highly salient. Of central importance for the values' approach, however, is how people spontaneously react to sexist behavior in others as a function of IMS, when the context does not demand bringing personal values to bear. In Study 3, we created a task in which we situated extremely and moderately sexist behaviors in a more ecologically valid context, and gave participants greater leeway to respond to the sexist behaviors, or not, as they saw fit. In this study, we also varied the number of sexist behaviors the target engaged in to examine how the amount of evidence participants are exposed to affects participants' spontaneous reactions. Devine and colleagues have shown that among high IMS people for whom nonsexist values are chronically salient, even a single response that violates personal standards is sufficient to spontaneously trigger guilt and self-regulatory efforts (Amodio et al., 2007; Plant & Devine, 1998; 2009). As such, we expected that whether the target engaged in one or more moderately or extremely sexist behaviors, high IMS people would infer the target was sexist, experience anger directed toward the target, and make efforts to

regulate the target's behavior to reduce sexism. Low IMS people, whose sexist values are not personally important or chronically salient, would be unlikely to spontaneously infer sexism, become angry or try to regulate the target's behavior when there was minimal, moderately sexist evidence but it seemed possible that they could do so when the evidence of sexism in the target's behavior increased.

We argue that IMS is more closely linked to personal standards than gender, attitudes, or EMS. To test this assertion, in each study, we evaluate the extent to which the predicted IMS effects hold when including gender, self-reported sexist attitudes, and EMS, as well as these variables' relationships with experimentally manipulated variables in our analyses.

Study 1

In Study 1, we assess the relationships among participants' level of IMS, the stringency of the standards they set for sexist behavior in themselves, and the stringency of the standards they set for sexist behavior in others. To this end, we adapted a method Devine and colleagues (Devine et al., 1991; Plant & Devine, 1998) previously used to assess standards for the self, in which people report how they *should* behave in various situations. Extremity of the behaviors was varied such that some participants reported their standards for self and others in response to extremely sexist behaviors, whereas others reported their standards in response to moderately sexist behaviors. We expected that, overall, people's standards for extremely sexist behaviors would be more stringent than those for moderately sexist behaviors (Brand et al., 1999; Swim et al., 2005). More importantly, we expected that the stringency of the standards for self and others would be predicted by IMS and that the stringency of the two sets of standards would not differ.

Method

Participants and design—Introductory psychology students (175 females, 52 males) participated in this study online in exchange for course extra credit. The study had a 2 (Standard Focus: self vs. others) \times 2 (Behavior Type: extreme vs. moderate) \times 2 (Standard Order: self first vs. others first) \times IMS, mixed-model factorial design. Behavior type and standard order were between-subject variables. Standard focus was a within-subjects variable. Participants were randomly assigned to the between-subjects conditions. In this study and all subsequent studies, we evaluate the effects of gender, MSS and EMS and whether they account for variance beyond the effects of IMS. We note that gender is particularly important to evaluate in this context because Klonis et al. (2005) validated the IMS sexism measure with only male participants. Thus, in the present work it will be critical to examine (1) the impact of gender on personal standards for self (as well as for others) and (2) whether the relationship between personal standards for self (and for others) and IMS is similar for male and female participants.

For each study, all measures, manipulations, and exclusions are disclosed. To determine target sample size for Studies 1 and 2, as a rule of thumb, we doubled the sample size of Devine and colleagues' previous work on IMS and value-based standards and rounded up to the nearest hundred. In Study 3, we increased our target sample to 350 because of its larger design (i.e., adding an additional manipulated variable). In each study, we stopped data

collection at the end of the week during which we attained our target sample size, resulting in final samples slightly higher than our targets. Data were analyzed only after data collection ended.

Procedure—In all studies, participants completed the IMS scale (5 items; e.g., “Being nonsexist toward women is important to my self-concept”), the EMS scale (5 items; e.g., “I try to act in nonsexist ways because of pressure from others”), and Swim et al.’s (1995) Modern Sexism Scale (MSS; 8 items; e.g., “Discrimination against women is no longer a problem in the United States”) in an online testing session early in the semester. IMS and EMS responses ranged from 1 (*strongly disagree*) to 9 (*strongly agree*). MSS responses ranged from 1 (*strongly disagree*) to 5 (*strongly agree*).

Participants were told that they would be completing a series of brief tasks. Participants rated the extent to which *they should* engage in 11 behaviors and the extent to which *other people should* engage in the same behaviors. Participants reported their should ratings in response to extreme or moderate behaviors (see pretest below) on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants were assured that their responses would be kept confidential and instructions emphasized the importance of being honest. Following the standards questionnaires, participants completed demographic information and were debriefed.

Personal standards for self and others. The instructions for reporting one’s personal standards for self and personal standards for others (**in bold**) read as follows.

We often set up standards or guidelines for evaluating our own (**others’**) behavior. We usually phrase these guidelines in terms of how we believe we (**other people**) SHOULD respond or behave in various situations. Based on your own personal standards for how you (**other people**) SHOULD respond, consider each statement on the following page. Please give your response according to the scale provided.

For example, one of the statements read as follows:

In general, *I should* (**other people should**) feel uncomfortable having a female boss.

Before reporting on the second set of standards, participants were told they would consider the same items but with a different standard in mind. It was explained that the second set of standards may or may not match the first set of standards they reported. Participants’ total standards for self score and standards for others score were formed by averaging their ratings across the eleven extremely or moderately sexist behaviors, after reverse scoring when necessary.

Behavior type pretest—We conducted a pre-test to identify the extremely and moderately sexist behaviors to be used in the studies (See Appendix A). The 22 behaviors used in the present study were chosen based on pre-testing in which a sample of 200 undergraduates (105 females, 95 males), none of whom participated in the main studies, rated the extent to which 83 behaviors were sexist on a scale ranging from 1 (not at all sexist) to 7 (extremely sexist). We selected 11 items that had the highest ratings of sexism (extreme: $M = 6.04$, $SD = 1.20$) and 11 items that were judged in the middle of the scale

(moderate: $M = 5.10$, $SD = 1.52$). We tested whether the means of the 11 extremely sexist behaviors and the 11 moderately sexist behaviors differed using a t-test; they did, $t(385) = 10.33$, $p < .001$. Though high IMS people, in general, rated behaviors as more sexist than did low IMS people $B = .25$, $SE = 0.03$, $t(198) = 7.48$, $p < .001$, all people rated the extreme behaviors as more sexist than the moderate behaviors. As such, despite the natural rating confound, the extremely and moderately sexist behaviors chosen for the present studies are well suited to test our hypotheses regarding how behavioral standards and the use of those standards to infer and react to sexism differ as a function of IMS.

Results

Overview of analyses—In all studies, unless otherwise noted, we conducted analyses using General Linear Models (GLMs). Using GLMs, we regressed all mean-centered predictor variables on our outcome variable simultaneously, as recommended by Judd, McClelland, and Ryan (2011). Behavior type was coded so that 0 = moderately sexist behaviors and 1 = extremely sexist behaviors, order was coded so that 0 = standards for others first and 1 = standards for self first, and gender was coded so 0 = female and 1 = male. Prior to examining interactions, all dichotomous variables were recoded into $-.5$ and $.5$ and all continuous variables were mean-centered. In the present study, standards for self ratings and standards for others ratings were scored such that lower ratings represent more stringent *should* standards than higher ratings.²

Across studies for each outcome variable, we fit two models. First, in the primary model we examined the outcome variable as a function of IMS, behavior type, and their interaction. If we observed a significant effect of IMS or the $IMS \times$ behavior type interaction in Model 1, in Model 2 we tested whether this effect held when including MSS, gender, EMS, and each of these variable's interaction with behavior type in the model. See Table 1 for descriptive statistics and correlations among the study variables and Table 2 for the summary of the regression analyses.

Standards for the self—In Model 1, as expected, a main effect of behavior type emerged such that participants' standards for extremely sexist behaviors were more stringent than their standards for moderately sexist behaviors. Further, replicating Plant and Devine (1998), IMS was related to standards for the self, such that the higher participants were in IMS, the more stringent their standards for self. In Model 2, the only other effect to emerge was a main effect of EMS, but the theoretically predicted main effect of IMS remained significant.

Standards for others—Consistent with our analysis on standards for self, participants reported more stringent standards for others when viewing extremely sexist behaviors than when viewing moderately sexist behaviors. IMS was also related to standards for others; the higher people were in IMS, the more stringent their standards for others. In Model 2, a main

²Preliminary analyses revealed that standard order was associated with a significant effect on only one variable (standards for the self), $B = .29$, $t(225) = 2.52$, $p < .05$. Participants who reported standards for others first reported permitting less sexism in their own behavior than those who reported standards for self first. The relationship among predictor variables on standards for self remained the same when controlling for order. We thus report analyses collapsed across order for all variables.

effect of EMS emerged, but again the theoretically predicted main effect of IMS remained significant.

Self–Other standard comparison—To ensure that people do not differ in the stringency (i.e. location) of their self and other standards as a function of IMS, we created a difference score between standards for the self and standards for others. No significant effects emerged on difference scores in Model 1, $|B|$'s = 0.12, t 's = 1.38, p 's > 0.17 or in Model 2, $|B|$'s = 0.16, t 's = 1.09, p 's > 0.28, suggesting that the difference between people's personal standards for self and for others does not vary as a function of IMS, behavior type, or their interaction nor self-reported sexism, gender, EMS or any of these variable's interactions with behavior type.

Discussion

The primary goal of Study 1 was to explore the relation between IMS and the location of people's personal standards for their own and others' behavior. Replicating past research, higher levels of IMS were associated with more stringent standards for the self (Plant & Devine, 1998). Supporting our novel predictions, IMS was also related to the location of the standards people set for others, and the stringency of standards set for the self and for others did not differ. These findings represent the first extension of Devine and colleagues' values framework into the interpersonal arena.

It is important to note that the relationships between IMS and each set of standards held when including self-reported sexism, gender, EMS and their interactions with behavior type in the model. It is especially noteworthy that no effects of gender emerged despite the prototypic nature of the sexist behaviors in the present work (i.e., directed from males to females; Inman & Baron, 1996). These findings suggest that, once IMS is accounted for, males and females do not differ in the stringency of the standards they set for sexism in the self or others. Although a main effect of EMS emerged on both standards for self and standards for others, EMS did not account for as much variance as IMS. We will evaluate the extent to which EMS and the other person variables emerge in consistent ways across the studies. Overall, these findings highlight the importance of considering person characteristics that are theoretically linked to the standards people set for their own or others' behavior.

Our findings also suggest that extremely sexist behaviors are viewed as less permissible than moderately sexist behaviors. Though IMS and behavior type did not interact to predict the location of standards set for the self or those set for others, it is possible that high and low IMS people may differ in how they *use* extremely vs. moderately sexist behaviors during the inference process. We will explore this issue in Study 2.

Study 2

Study 1 demonstrated that people's personal values for sexist behavior are related to the standards they set for other people. Study 1 did not, however, provide information on how evidence is used to make judgments about others. To address this issue, in Study 2 we explored whether people's values predict the type and amount of evidence required to infer

sexism when participants are explicitly asked to make a sexism judgment. Adapting an inference paradigm used by Biernat et al. (2008), participants were sequentially presented with a set of extremely or moderately sexist behaviors and asked to indicate when they had enough evidence to conclude that a target was sexist. We hypothesized that high IMS people would require fewer behaviors to conclude the target was sexist than low IMS people regardless of behavior type. Because past work showed more variability in inferences made from moderately compared to extremely sexist behavior, (Brant et al., 1999; Swim et al., 2005), we also recognize the possibility that personal values may play a larger role in inferences when people are asked to evaluate moderately compared to extremely sexist behavior. In this case, we would expect IMS to interact with behavior type such that everyone would require similarly few extremely sexist behaviors to infer sexism, but high IMS people would require fewer moderately sexist behaviors to infer sexism than would low IMS people.

Method

Participants and design—Introductory psychology students (119 females, 92 males) participated in this study online in exchange for course extra credit. The design was a 2 (Behavior Type: moderately sexist vs. extremely sexist) X 2 (Order: moderately sexist first vs. extremely sexist first) × IMS mixed-model factorial³. Order was a between-subjects and behavior type a within-subjects variable. Participants were randomly assigned to the order condition. Of the 211 participants, 12 were eliminated from analyses because of task inattention (i.e., they continued to request items after the 15 unique behaviors began to recycle). We note, however, that including these 12 participants' data in the analyses (with scores 16 and over coded as 15) did not change the pattern or significance of the reported effects.

Materials and procedure—The study was conducted online and took 15–20 minutes to complete. Participants were told that we were interested in how people confirm that a person is sexist based on behavior. In the task, a single behavior appeared on the screen (e.g., “is disappointed when a woman joins his softball team.”), and participants were asked to indicate when they had enough evidence to confirm that target was sexist. The behaviors were those from Study 1, changed in format to reflect an action.

The exact instructions read:

We are interested in how people decide whether or not a person is sexist based on the person's behaviors. Specifically, we are interested in the total number of behaviors that are necessary to confirm that a person is sexist. In this task, you will read statements about a guy named Josh, whom you've never met. The statements describe Josh's behaviors and thoughts. Your task is to read the statements one by one and tell us when you do not need to see any more statements to be certain that Josh is sexist. You can read through as many statements as needed to confirm that Josh is sexist. The statements describing Josh's behaviors and thoughts will appear

³In the present study participants were asked to either suspect or confirm when the actor was sexist. All relevant results remained significant in both conditions and thus in this paper we collapsed across conditions.

on the screen one at a time. After reading each statement, decide if you have enough information to confirm that Josh is sexist. If you feel you need to see another statement to confirm that Josh is sexist, please press 'N' for 'Next.' When you feel satisfied that you have seen enough statements to confirm that Josh is sexist, please press 'S' for 'Stop.'

Participants completed the task with moderately sexist and extremely sexist behaviors (order counterbalanced and different actors were used). Participants then read a debriefing document.

The dependent variable was number of behaviors needed to infer the target was sexist.

Results

Overview of analyses—Across Study 2 outcomes, we fit the same two models as in Study 1. Preliminary analyses revealed no effect of order and we do not discuss it further. See Table 3 for descriptive statistics and correlations among study variables and Table 4 for the summary of the regression analyses.

Inference analyses—Model 1 revealed a main effect of behavior type, suggesting that participants required 1.99 more moderately ($M = 4.07$, $SD = 2.89$) than extremely ($M = 2.08$, $SD = 1.60$) sexist behavior to infer the target was sexist. An IMS main effect also emerged, such that the higher participants were in IMS, the fewer behaviors they needed to infer sexism. These main effects were qualified by an $IMS \times$ Behavior Type interaction (see Figure 1). Descriptively, in the extremely sexist behavior condition, IMS was unrelated to the number of behaviors required to infer sexism, $B = -.09$, $SE = 0.07$, $t(197) = -1.22$, $p = .22$. However, in the moderately sexist behavior condition, participants higher in IMS required fewer behaviors to infer sexism, $B = -.48$, $SE = 0.13$, $t(197) = -3.75$, $p < .001$.

In Model 2, main effects of gender and MSS emerged, but there was no main effect of EMS and none of these variables interacted with behavior type. Moreover, with all these terms in the model, the IMS main effect and the $IMS \times$ Behavior Type interaction remained significant.

Discussion

A major goal of Study 2 was to determine whether IMS related to the use of behavioral evidence that varied in extremity to infer sexism in a task in which people accumulated evidence over time. Our findings revealed that when behaviors were extremely sexist, there was little variance in the number of behaviors participants required to infer sexism. When behaviors were more moderate, however, higher IMS people required fewer behaviors than their lower IMS counterparts to reach the conclusion that a target was sexist. As in Study 1, though other main effects emerged, our theoretically relevant IMS effects held when including self-reported sexist attitudes, gender, EMS, and each of these variables' interactions with behavior type in the model.

Although Study 2 tells us that people at varying levels of IMS *can* use behavioral evidence to make sexism inferences, specifically when instructed to do so, it has some limitations that

restrict its ability to test the utility of Devine and colleagues' values framework in the interpersonal arena (Klonis et al., 2005; Plant & Devine 1998). First, rarely do perceivers encounter one sexist behavior after another, as was the case in Study 2. More likely, a person engages in sexist behaviors interspersed with other, irrelevant behaviors. This reality is important because irrelevant information can dilute the influence of relevant information when making a judgment (Nisbett, Zukier, & Lemley, 1981). Second, precisely because Study 2 specified a judgement criterion (i.e., participants were required to make a sexism inference), the context may not reflect natural situations in which people would have to bring their value based standards to bear and spontaneously make sexism inferences. In short, behaviors that lead to sexist inferences in the task employed in Study 2 may not do so when sexist behaviors are interspersed with irrelevant behaviors and judgments of sexism are not explicitly required.

Finally, the rather artificial nature of the imagined paradigm used in Study 2 is also limited with regard to testing hypotheses about perceivers' affective reactions to a target's sexist behavior and whether perceivers would make efforts to regulate a target's sexist behaviors. For a perceiver to become angered, he or she must believe that the sexist behavior is consequential for the target. And, for one to act to address the sexist behavior, the situation must be one in which there is a possibility that one's efforts could help reduce the perpetrator's sexist behavior in the future. In Study 2, we neither measured anger nor provided participants with an opportunity to take corrective action. Study 3 was designed to address these limitations.

Study 3

The primary goals of Study 3 were to examine whether the spontaneous inferential, affective, and regulatory consequences of encountering a target engaging in sexist behavior vary as a function of IMS. To pursue these goals, in Study 3 we created a naturalistic judgment context in which participants would evaluate a target's sexist behavior as it unfolds over time, interspersed with irrelevant behavior and with no explicit goal to judge the target's sexism. To test whether people higher in IMS are more sensitive to sexist behaviors in others than people lower in IMS, we manipulated behavior extremity but also added a manipulation of the number of the target's sexist behaviors.

Specifically, participants in Study 3 were led to believe that they would provide feedback on a new system for evaluating teaching assistants (TAs). They were told they would read two transcripts in which the TA to be evaluated (1) discussed the Introductory Physics class and teaching strategies with other Physics TAs and (2) interacted with a student during office hours. Embedded into the transcripts, the TA made 1 or 3 extremely or moderately sexist statements related to women being less interested in or capable of understanding the course material (Shapiro & Williams, 2012). After reading the transcripts, participants reported their affective responses and had the opportunity to leave feedback in the form of open-ended comments for the TA and the TA's supervisor. No explicit judgments of sexism were requested. The open-ended comments provided the opportunity to confront the TA by labeling him or his behavior as sexist and to offer suggestions for reducing sexist behavior.

We anticipated that, because high IMS people's values are chronically accessible, they would bring their standards to bear and infer sexism in a target regardless of the amount or extremity of the sexist behaviors. Low IMS people, however, would be less likely to spontaneously bring their standards to bear and would thus be less likely to infer sexism in a target, particularly when the evidence is minimal and the behaviors are moderate. Furthermore, building on Devine and colleagues' research examining the intrapersonal affective and self-regulatory consequences of violating self-imposed standards, we expected that if, as anticipated, high IMS people were more likely to infer sexism than their low IMS counterparts (Plant & Devine, 1998), high IMS people would also be more likely to experience anger directed toward the TA and attempt to regulate his behavior.

Method

Participants and design—Introductory Psychology students (215 females, 134 males) completed the study in lab in exchange for course extra credit. The design was a 2 (Behavior Type: extremely sexist vs. moderately sexist) \times 2 (Behavior Number: one vs. three) \times IMS, mixed-model factorial. Participants were randomly assigned to behavior type and behavior number conditions.

Pretest—Research assistants, who were given a description of the TA evaluation, generated sexist statements that a TA could plausibly say in a TA meeting or in office hours. Undergraduates (136 female, 92 male) rated the extent to which each of 47 statements was sexist (1 = *not at all sexist* to 7 = *extremely sexist*). Twenty-seven of the statements were described as occurring during a meeting in which TAs discussed how to help students who had struggled on an exam. The other twenty statements were described as occurring during a male TA's office hours in response to a female student saying that she is a math major. We identified three extreme statements, three moderate statements, and two neutral statements to be incorporated into the transcripts. In the three behavior condition, two sexist statements appeared in the TA meeting transcript and one sexist statement appeared in the office-hours transcript. In the one behavior condition, one sexist statement appeared in the transcript of the TA meeting; neutral statements replaced the other sexist statements.

In the TA meeting transcript, the two extremely sexist statements were: "Having girls in the class seems to slow us down." ($M = 6.29$, $SD = 1.27$) and "Since a lot of the girls are struggling, maybe we should incorporate more examples using topics girls are interested in, like cooking and fashion." ($M = 6.00$, $SD = 1.50$). The moderately sexist statements were: "The girls just seem less engaged by the material." ($M = 4.21$, $SD = 1.76$) and "It just seems that a lot of the girls just don't enjoy the class, so I think we need to try and engage them more." ($M = 3.90$, $SD = 1.79$). The neutral statement: "We might want to hold an extra review session for any students who feel that they are struggling." ($M = 1.43$, $SD = 1.21$). For the office-hours transcript, in response to the student reporting that she was a math major, the extremely sexist response was: "Oh, that's surprising. Women's math and spatial abilities simply aren't as good as men's." ($M = 6.44$, $SD = 1.17$). The moderately sexist response was: "I guess that explains why you're really excelling compared to other girls in this class." ($M = 3.52$, $SD = 1.82$). The neutral response was: "The math department at this school is really great." ($M = 1.16$, $SD = .74$). A t-test demonstrated that the means of the 3

extremely sexist behaviors and the 3 moderately sexist behaviors were significantly different, $t(228) = 16.11, p < .001$.

Materials and procedure—Participants, in groups of 1 to 15, were seated at computers separated by partitions. They were told that they would be helping to test a new TA evaluation procedure for the university and would be evaluating Mike, a TA in Introductory Physics. They learned they would read one transcript of a TA in a meeting with other TAs and a transcript of the TA in office hours with a student (See Appendix B for full transcripts). Participants were asked to pay attention to instances in which they thought the TA performed well or poorly and were told that they would be given an opportunity to evaluate and provide feedback for the TA.

Immediately after reading the two transcripts, participants completed an affect measure. Participants were then given a free response section to provide feedback for the TA that would be viewed by both the TA and his supervisor; they were assured that their responses would be anonymous. As an attention check, participants were asked to report the female student's major. Finally, participants filled out demographic information and were debriefed.

Measures

Affect indices: Participants reported their affect on 32 affect items (see Devine et al., 1991; Plant & Devine, 1998) ranging from 1 (*does not apply at all*) to 7 (*applies very much*). The scree test from a principal components analysis suggested a three-factor solution, accounting for 59.9% of the variance. We extracted three factors using an oblimin rotation and retained items that had unique loadings above .60. To construct the affect indices, we averaged the items that loaded on each factor, reverse coding when necessary. The first factor, anger at others, accounted for 36.5% of the variance and included: *angry at others, uneasy, bothered, uncomfortable, frustrated, irritated with others, and disgusted at others*. The second factor, positive affect, accounted for 16.9% of the variance and included: *friendly, consistent, happy, energetic, optimistic, good, and content*. The third factor, negative self-directed affect, accounted for 6.4% of the variance and included: *self-critical, angry at myself, guilty, annoyed at myself, regretful, disappointed in myself, and disgusted with myself*.

Free response TA evaluation: Two coders, blind to participants' scores on all the individual difference measures, coded the free response feedback that participants wrote to the TA and his supervisor. Inter-rater reliability was evaluated with a Kappa statistic. Comments were coded for *inferences of sexism* ($\kappa = .82$) and *regulatory efforts* ($\kappa = .75$). Inferences of sexism included statements that labelled the TA's comments as inappropriate or made clear dispositional claims about the TA being sexist (e.g., "You could have valuable knowledge or be good at your job, but I can't even evaluate it because I am unable to see past how ignorant, sexist, and rude you are..."; "The comment about girls slowing down the pace of the class was off color and unhelpful."). Regulatory effort comments included explicit advice or suggestions for the TA on how to be less sexist or inappropriate in the future (e.g., "Don't stereotype all girls...", "He needs to improve on listening and being respectful of people, regardless of their gender...").

Results

Overview of analyses—To analyze the dichotomous outcomes in this study (inferences of sexism and regulatory efforts), we fit a Generalized Linear Model using the binomial family with a logit link. Gender and behavior type were coded as in Studies 1 and 2. Behavior Number was coded such that 0 = one behavior and 1 = three behaviors. For each outcome variable, we fit the same two models as in Studies 1 and 2, but with the addition of Behavior Number as an independent variable. Suggesting that participants paid attention to the transcript, of the 349 participants, all but one correctly reported the female student's major. Analyses with and without this participant's data revealed that all effects were unchanged and, because the participant had clearly engaged in all other aspects of the paradigm (e.g., including lengthy and thoughtful feedback for the TA), the participant's data were included in reported analyses. See Table 5 for descriptive statistics and correlations among Study 3 variables.

Affect analyses

Anger at others: See Table 6 for the summary of the regression analyses. Model 1 revealed an effect of behavior number such that people experienced more anger at others after reading that the TA engaged in three sexist behaviors than after reading that he engaged in one sexist behavior. Further, the behavior type effect suggested that people reported more anger after reading extremely sexist than moderately sexist behaviors. Finally, the IMS effect showed that the higher people were in IMS, the more anger at others they reported.

Behavior number and behavior type interacted to predict anger at others. Descriptively, in the one behavior condition, anger did not vary across behavior type, $B = .16$, $t(341) = .79$, $p = .42$. In the three-behavior condition, however, people who read extremely sexist behaviors experienced more anger than people who read moderately sexist behaviors, $B = 1.47$, $SE = .22$, $t(341) = 6.87$, $p < .001$, $R^2 = .12$. The effects of behavior number and behavior type serve as sensible manipulation checks on the efficacy of our paradigm. More interestingly, as can be seen in Figure 2, the analysis yielded an interaction between IMS and behavior number. In the one behavior condition, IMS was unrelated to the amount of anger reported, $B = -.03$, $t(341) = -0.51$, $p = .61$. In the three behavior condition, however, the higher people were in IMS, the more anger they reported, $B = .22$, $SE = .06$, $t(341) = 3.40$, $p < .001$, $R^2 = .03$.

In Model 2, the $IMS \times Behavior\ Number$ interaction remained significant; no other effects emerged.

Negative self-directed affect and positive affect analyses: There were no effects of IMS, behavior number, or their interaction on negative self-directed affect, $|B|$'s $.19$, t 's 1.27 , p 's $.16$. However, people who saw extremely sexist behaviors experienced less positive affect than people who saw moderately sexist behaviors, $B = -.33$, $t(341) = -2.27$, $p = .02$, 95% CI = $[-0.60, -0.04]$, $R^2 = .01$, and people who saw three behaviors experienced less positive affect than people who saw one behavior, $B = -.67$, $t(341) = -4.65$, $p < .001$, 95% CI = $[-0.95, -0.38]$, $R^2 = .06$. The behavior number and behavior type main effects suggest that, in general, the more total and the more extreme sexist behaviors people encounter, the

less positive affect they experienced. There was also a main effect of IMS such that the higher people were in IMS, the less positive affect they experienced after reading the transcripts, $B = -.10$, $SE = .04$, $t(341) = 2.22$, $p = .03$, 95% CI = [-0.18, -0.01], $R^2 = .01$.

Free response TA evaluations—See Table 7 for the summary of the regression analyses. In examining *inferences of sexism*, Model 1 revealed a main effect of behavior number, such that people in the three behavior condition were more likely to infer sexism than people in the one behavior condition. There was also a main effect of behavior type, such that participants were more likely to infer sexism in the extremely sexist condition than the moderately sexist condition. Furthermore, the analysis revealed a main effect of IMS, such that the higher people were in IMS, the more likely they were to infer sexism. Beyond these main effects, the analysis also yielded an interaction between IMS and behavior number (See Figure 3). Descriptively, in the one behavior condition, IMS was unrelated to likelihood of inferring sexism, $B = .07$, $\chi^2(1, N = 349) = .53$, $p = .47$. In the three behavior condition, on the other hand, the higher people were in IMS, the more likely they were to infer sexism, $B = .48$, $\chi^2(1, N = 349) = 12.76$, $p < .001$.

In Model 2 the only other effect to emerge was a main effect of MSS and the IMS \times behavior number effect remained significant.

The pattern of results was the same for *regulatory efforts*, revealing main effects of both behavior number and behavior type, as well as the IMS \times Behavior Number interaction (See Figure 3). In the one behavior condition, IMS was unrelated to probability of making regulatory efforts, $B = -.08$, $\chi^2(1, N = 349) = .39$, $p = .53$, but, in the three behavior condition, the higher people were in IMS, the more likely they were to make regulatory efforts, $B = .27$, $\chi^2(1, N = 349) = 7.52$, $p = .006$.

In Model 2, main effects for MSS and EMS emerged and there were two higher order interactions involving behavior number. Even with the emergence of these additional effects, the IMS \times Behavior Number interaction remained significant.

Relationship among IMS, behavior number, negative affect, and regulatory efforts—A primary hypothesis in the present study is that when others violate the standards people set for others, people will experience anger and, consequently, try to regulate the behavior of the target. Because behavior number moderated the effect of IMS on anger at others, as well as the effect of IMS on regulatory efforts, we were interested in whether behavior number would also moderate the effect of IMS on regulatory efforts when controlling for anger at others and the IMS \times anger at others interaction. Results of this analysis revealed that the magnitude of the IMS \times Behavior Number interaction on regulatory efforts was reduced from 0.35 when tested alone, $\chi^2(1, N = 349) = 5.33$, $p = .02$, to 0.26 when controlling for anger and the interaction between IMS and anger, $\chi^2(1, N = 349) = 2.81$, $p = .09$. In contrast, there was a strong relationship between anger at others and regulatory efforts such that people who reported more anger at others were more likely to make efforts to regulate the TA, $B = .31$, $\chi^2(1, N = 349) = 13.49$, $p < .001$. These results suggest that viewing three sexist behaviors causes high IMS people to experience anger at others as well as causes high IMS people to engage in corrective action.

People high in IMS only experienced anger at others and chose to make efforts to regulate the TA's behavior when they saw three sexist behaviors. And, there was a strong relationship between anger at others and regulatory efforts, even controlling for the interaction between IMS and number of sexist behaviors. Together, these findings suggest that there was an indirect effect of IMS on regulatory efforts through anger that was only present when participants viewed three behaviors (i.e., a moderated mediation model, Muller, Judd, & Yzerbyt, 2005; see Figure 4). We investigated this hypothesis directly using quasi-Bayesian estimation, as implemented in the mediation package in R (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). The results of this analysis suggest that behavior number does indeed moderate the indirect effect of IMS on regulatory efforts, $B = .01$, 95% CI = [0.028, .002]. For people who viewed only one sexist behavior, the indirect effect of IMS on regulatory efforts via anger at others was approximately 0, $B = -.001$, 95% CI = [-0.007, .005], suggesting that people who are relatively high in IMS did not experience anger at others and attempt to regulate the TA's behavior. However, when they saw three behaviors, the indirect effect of IMS on regulatory efforts was positive, $B = .01$, 95% CI = [0.003, .026]. This effect suggests that, when people high in IMS saw three behaviors, they became angry. This anger may in turn have motivated them to put forth effort to bring the TA's behavior in line with their nonsexist standards for others. Given that anger was not directly manipulated, we cannot definitively conclude that anger caused the participants to attempt to regulate the TA's behavior. Nonetheless, it seems clear that IMS is related to both anger and corrective action, at least when people see three sexist behaviors, and that people who experience more anger exhibit more corrective behavior, regardless of IMS.

Discussion

Just as people's personal values regarding sexism guide their reactions to bias in their own behaviors, our findings suggest that those same personal values guide their reactions to potential bias in the behavior of others. In the present study, it is notable that although the number of potentially sexist behaviors encountered influenced high but not low IMS people's reactions, the extremity of the sexist behaviors did not influence anyone's reactions. Encountering a single moderately or extremely sexist behavior did not lead anyone to infer the target was sexist or to affective or regulatory responses. Encountering three potentially sexist behaviors, on the other hand, regardless of behavior extremity, led high but not low IMS people to infer the TA was sexist, experience anger, and make efforts to regulate the TA's behavior by providing feedback about how to respond in less sexist ways. Across the outcome measures, effects emerged with the other person variables, but these effects were not consistent, and importantly, all theoretically relevant IMS effects held when including sexist attitudes, gender, EMS and each of these variables' interactions with the manipulated variables in the models.

Though we hypothesized that high IMS people's chronically accessible values would lead them to infer sexism even with minimal evidence, as they do when judging the self, this was not the case. We reason, however, that in judging the self, people are highly confident when they violate their standards, and can thus hold themselves immediately accountable. When judging others, however, because people lack access to personal history of sexist behavior, it may be more difficult to be certain that the other is sexist. In addition, because accusations

of sexism are serious and consequential to both the target and the perceiver (Czopp & Monteith, 2003; Dodd, Giuliano, Boutell, & Moran, 2002; Fitzgerald, Swan, & Fischer, 1995; Kaiser & Miller, 2001), we reason that all people are judicious when making what could be viewed as a harsh accusation. Although high IMS people may be reluctant to make a sexism inference based on a single behavior, this behavior may sensitize them to the possibility that the target is sexist. In this way, we suspect that the TA's first sexist behavior, whether extreme or moderate, led participants high in IMS to form a hypothesis that the TA was sexist that was then tested with the subsequently presented behaviors (Darley & Gross, 1983). Once a pattern was detected with a greater number of behaviors, high IMS people perhaps felt warranted in drawing a sexism inference.

But what of the low IMS people who clearly can draw inferences of sexism from moderately and extremely sexist behaviors, as in Study 2, but do not in Study 3? We suggest that because their values regarding sexism are neither personally important nor chronically accessible, it is possible that in Study 3 low IMS participants did not attend to the potentially sexist behaviors among the other, non-sexist behaviors. Alternatively, they may have noticed the behaviors but not have viewed the behaviors as unacceptable or provocative enough to report an unprompted inference or regulatory response in open-ended comments. Finally, even if low IMS people attended to the potentially sexist behaviors and found them unacceptable, the extraneous information in Study 3 may have diluted the impact of the behaviors (Nisbett et al., 1981), making them less salient or influential in their evaluation of the TA.

Our findings cannot account for why low IMS people failed to respond to sexist behaviors. However, teasing apart the cognitive processes that do and do not lead to sexism inferences in future work could be beneficial, particularly in developing interventions to reduce sexism. That is, identifying points in the inference process that facilitate or deter people from reacting to sexism may provide leverage points at which to intervene to encourage the detection of sexism (e.g., increase attention to sexist behaviors, increase concern about behaviors that disadvantage women) and, once detected, to encourage action to counter its effects.

General Discussion

In an effort to understand why people disagree about what is and is not sexist behavior, we explored the utility of extending Devine and colleagues' values analysis of the intrapersonal processes of detecting and reacting to sexism in the self, to the interpersonal arena in which people may detect and react to potential sexism in others (Devine et al., 1991; Monteith, 1993; Plant & Devine, 1998). To this end, in a program of three studies, we examined the standards regarding sexist behavior people set for others as compared to those set for the self and the inferential, affective and regulatory consequences of encountering targets behaving in sexist ways. The pattern of findings across studies strongly supports the validity of applying the values framework to the interpersonal context in which people observe and react to sexism in targets.

Study 1, for example, showed that the standards people set for others are derived from the same values regarding sexism as are the standards people set for themselves. That is, people high in IMS, who embrace and internalize non-sexist values, set more stringent standards for both the self and for others than people low in IMS, who do not embrace non-sexist values. And, our findings suggest that people appear to hold self and others to similarly strict standards. These findings are foundational for testing the applicability of Devine and colleagues' values analysis to the interpersonal context and support the theoretical significance of IMS as an individual difference characteristic that is relevant for examining how people react to sexism in others.

Having established the standards to which people hold others, Studies 2 and 3 explored how people *use* these value-based standards in interpreting and evaluating others' behaviors and brought into focus the importance of the extent to which people's values are salient when observing sexism in a target. Although high IMS people's values regarding sexism are important and chronically salient across situations, this is not the case for people low in IMS. In Study 2, values regarding sexism were made highly salient for all participants by explicitly requiring participants to report when they could make a sexism inference. When a target's behaviors were extremely sexist, participants required little evidence to conclude that the target was sexist. When the behaviors were moderate, however, the role of participant values was evident. Following the findings from Study 1, high IMS participants, whose standards for moderately sexist behaviors are more stringent than low IMS people's standards, required fewer behaviors than low IMS people to support a sexism inference.

Study 3 addressed circumstances in which participants had to spontaneously bring their standards to bear as evidence accumulated in a more ecologically valid context. This context is important because it reflects the type of everyday settings in which people may observe potentially sexist behaviors and have to decide whether the behavior(s) warrant an inference of sexism and what type of response, if any, should follow. As such, this type of setting affords the opportunity to test the affective and regulatory consequences that the values framework suggests should follow when one infers a target is sexist. When provided with minimal evidence, participants, regardless of their level of IMS, were reluctant to infer that the target was sexist; this was true whether the behaviors were extreme or moderate. This finding suggests that inferences of sexism, which can be quite harsh, are not advanced easily. The study also showed, however, that as moderate *or* extreme evidence accumulated, high IMS people were much more likely than their low IMS counterparts to generate inferences of sexism, express anger at the target and make efforts to reduce the target's sexism. Together, Studies 2 and 3 suggest that though high and low IMS people can infer sexism when explicitly asked to do so, only high IMS people, whose personally important values are chronically accessible across contexts, are likely to do so spontaneously.

Study 3's findings highlight how, in ordinary settings, people's values regarding sexism are at the heart of the disparate conclusions about what is and is not considered sexist behavior. Overall, these findings support Swim and colleagues' (2005) speculation that people use similar criteria and evaluative processes to judge their own and others' behavior. They also suggest, however, that it may be most prudent to examine inferences as a function of IMS, an individual difference variable that is theoretically connected to value-based standards for

sexism. Indeed, across studies, IMS provided consistent predictive utility and accounted for variance in outcomes above and beyond person variables previously advanced to explain sexism inferences. Although across studies, sexist attitudes, gender, and EMS each emerged as outcome predictors on occasion, none did so consistently and including these additional individual difference measures along with their interactions with manipulated variables did not eliminate the effects of IMS.

These findings strongly echo Plant and Devine's (1998) recommendations regarding the importance of looking beyond traditional attitude measures when exploring the processes involved in the intrapersonal and interpersonal regulation of prejudice. Often researchers have used traditional attitude measures as a proxy for motivation, assuming that lower prejudice people would be more motivated to respond without prejudice than higher prejudice people. Though attitudes covary with IMS such that high IMS people are less prejudiced and sexist (Klonis et al, 2005; Plant & Devine, 1998; the current studies), time and again, IMS proves to be a more sensitive predictor of people's reaction to their own manifestations of prejudice and sexism and, in the present studies, their reactions to sexism in others. We have argued that IMS is important precisely because it is linked to standards for what is appropriate in the prejudice and sexism domains, whereas attitudes are not (Plant & Devine, 1998). Similarly, though IMS is related to gender, when equating people for IMS, gender does not have unique effects.

It is important to note that the present work suggests that when people bring their divergent, value-based standards to bear, disagreements may inevitably arise. Failing to recognize the source of these disagreements, however, has led researchers, institutions, organizations, and even the legal system astray in how they attempt to moderate these often heated misunderstandings in their respective domains. Consider, for instance, a researcher examining different variations of humor. If the researcher were to ask participants to indicate how often they hear sexist jokes, the present work suggests that low and high IMS people may interpret the question very differently. Whereas high IMS people may believe the item encompasses all humor poking fun at women (e.g., about PMS), low IMS people may believe the item only refers to blatantly derogatory jokes (e.g., about women's lack of intelligence), and thus research findings would be confounded, unbeknownst to the researcher.

Indeed, without a consensus regarding the behaviors that people deem acceptable or unacceptable, addressing sexism disputes or concretely enforcing any behavioral regulations to decrease sexist behavior becomes difficult. By recognizing the subjectivity in interpretations of sexist behavior stemming from personal values, however, institutions can work to bridge the gaps between differing conceptions of sexism and to discourage sexist behavior in ways that make sense to people who vary in their nonsexist values (i.e., differ in level of IMS).

One leverage point to create consensus surrounding what people consider sexism could be change in personal values themselves. In this way, theoretically, researchers could intervene to guide low IMS people to endorse more egalitarian values (i.e., those of high IMS people). This could, in turn, cause low IMS people to broaden the scope of the behaviors they

consider unacceptable and to decrease their own biased behavior. Because personal values are so entrenched and central to the self-concept, however, changing these values is both difficult and time-consuming (Rokeach, 1973). At least in the short-term, therefore, other strategies to reconcile differences in sexism beliefs may be more fruitful.

Working to intervene at the institutional level, for example, may have more practical and widespread utility. We suggest that institutions that wish to eliminate the potential ambiguity surrounding sexism may have to establish their own explicit, *concrete* guidelines for which behaviors they deem inappropriate. It should be noted that simply endorsing egalitarian, nondiscriminatory policies, which many organizations already do, would be inadequate as a strategy. Such policies leave the decision of what is and is not considered sexist effectively in the eye of the beholder. And, people on campuses who endorse such broad anti-sexism policies may assume that the relatively narrow or broad scope of behaviors *they* consider sexist are, likewise, impermissible to the institution. This can easily lead to misunderstandings and conflict when a student is, for example, under disciplinary review for making a potentially sexist remark and the review board is made up of people who vary in level of IMS.

To the extent that clear guidelines could be developed and understood by members of the organization, however, these guidelines could mitigate differences in how low and high IMS people make sense of behavior by providing a communal lens, or standard, with which to interpret and regulate their own and others' behavior. And, regardless of whether people personally endorse the institutional standard, they would presumably want to comply with the standard to succeed within the institution. This behavioral shift could, in turn, create a more egalitarian and welcoming environment for all people.

One caveat to this approach, however, is that although EMS does not play a large role in predicting inferences when the self is judging another, its role is likely increased when the self is being judged by others. That is, when people who are predominantly high in EMS (i.e., those who behave with the goal to avoid punishment), feel pressured to adhere to others' standards, they often become angry and behave in ways that run directly counter to external mandate when freed from scrutiny, presumably to reassert their autonomy (Plant & Devine, 2001; see also Legault, Gutsell, & Inzlicht, 2011). Thus, our suggestion for explicit guidelines for sexism should be implemented with caution. Because people with more lenient standards naturally find more behaviors acceptable, to them, the "politically correct" restrictions set by institutions may seem overly sensitive and unwarranted. And, because these "politically correct" restrictions are often ambiguous within institutions, it is not surprising that these people become frustrated or confused when they are admonished for behaviors they personally deem acceptable. As we move forward, therefore, it will be imperative to strike a balance between delineating sexism in explicit terms, allowing people to feel autonomous in their behavior, and giving people ample reason for why a behavior should be considered sexist and unacceptable (Legault et al., 2011).

One potential way to avoid abstract, subjective definitions of "sexism" and to prevent backlash from those who find external sanctions regarding sexism unjustified and threatening would be to anchor guidelines for sexism to concrete behaviors that create

disadvantage. Rather than argue about *whose* standards may be the correct standards, institutions can point to objective evidence that links specific types of behaviors to outcomes that perpetuate disadvantage for a target group. In this way, regardless of what people personally believe or how stringent their personal standards may be, these guidelines could provide a tangible, explainable reason for why certain behaviors are off limits. By making these guidelines widely known, they could actively decrease biased behavior and, over time, even encourage more individual egalitarian values.

Additionally, though the patterns of observed findings throughout the present work strongly support the hypotheses advanced, it would be interesting and important to expand on the current results to evaluate their generality across contexts and judgment domains. Having established the importance of value-based standards for sexism inferences in our university classroom context, for example, it would be critical to extend the present work into other contexts in which the issues surrounding sexism are also significant (e.g., employment, legal). Though we cannot generate any principled reason to anticipate that our conceptual approach would play out differently in different contexts, it seems possible that there may be moderators of the inference processes that we are not aware of in these contexts.

Furthermore, findings from the present work may have implications for interpersonal judgments beyond those regarding sexism. The present considerations likely generalize, for example, to the race domain in which people similarly disagree about what constitutes racism (Sommers & Norton, 2006). Indeed, Devine and colleagues have demonstrated the utility of the values framework in predicting self-evaluation and regulation across bias domains (e.g., race, gender, sexuality). We would anticipate, therefore, that the framework would similarly generalize across domains when used to predict other-oriented judgements. With that said, because people often see racism as a more severe transgression than sexism (Czopp & Monteith, 2003; Rodin, Price, Bryson & Sanchez, 1990), we note that racially biased behavior could incite more extreme other-oriented judgements and responses than those surrounding sexism. On the other hand, because inferences of racism may be more socially charged than inferences of sexism, people may be more reluctant to label behaviors as racist, leading people to need more evidence of racial bias to infer racism. Future work should thus explore the use of value-based standards for judging others in different morally ambiguous judgment domains.

Finally, though our evidence suggests that high IMS people are more willing to confront a sexist perpetrator (either by explicitly labeling a behavior as sexist or unacceptable or by making recommendations for behavioral change) than low IMS people, these confrontations were safe. Participants in our study did not have to confront the perpetrator in person, behavior which is often associated with a host of negative psychological (e.g., fear of humiliation) and interpersonal (e.g., decreased favorability from others, being labelled a complainer; Czopp & Monteith, 2003; Dodd et al., 2002; Fitzgerald et al., 1995) consequences. Indeed, people often overestimate the extent to which they would confront sexism in hypothetical situations relative to real-time situations (Woodzicka & LaFrance, 2001). Because confrontation has the potential to decrease biased behavior in the perpetrator (Czopp, Monteith, & Mark, 2006; Mallett & Wagner, 2011), however, it is worth testing whether patterns of confrontation in the present work would play out similarly in in-person

interactions in which personal costs may be more salient. Given the importance nonsexist values, we suspect that high IMS people would still be more likely to confront sexism than their low IMS counterparts.

Conclusion

The present work represents a significant advance in our understanding of how people judge the behavior of others as well as clarifies the processes that lead to frequent disagreements about sexism in everyday life. By extending Devine and colleagues' theoretical values framework into the interpersonal domain, we can partially account for *why* people differ in their interpretations of the same behavioral evidence. People set their standards for sexist behavior in others based on their personal values regarding sexism. And, when people who vary in the extent to which they endorse nonsexist values judge potentially sexist behaviors, their differential interpretations lead to misunderstandings, and, often, anger. Yet, we suggest that to make progress toward true equality, what is considered sexist needs to be taken out of the eye of the beholder. We propose that by tying explicit guidelines for sexism to the inequality that results from biased behavior, we may create reasonable, consensual, concrete boundaries for acceptable behavior. And, by removing the ambiguity surrounding sexist behavior, we can focus our energies on creating a broader understanding of the negative effects of sexism, reducing expressions of sexism, and, ultimately, promoting equality.

Acknowledgments

Funding: Preparation of this article was supported by NIH grant 5R01GM111002-02, by a Wisconsin Alumni Research Foundation Professorship awarded to the last author, and by a Hilldale Undergraduate/Faculty Research Fellowship.

We thank William T. L. Cox, Emily L. Dix, Patrick S. Forscher, and Katharine Scott for their helpful comments on previous versions of this paper. We also thank the Devine Lab RAs for their help running the studies presented in this paper.

References

- Amodio DM, Devine PG, Harmon-Jones E. A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*. 2007; 18:524–530. [PubMed: 17576266]
- Barreto M, Ellemers N. Sexism in contemporary societies: How it is expressed, perceived, confirmed, and resisted. *The SAGE handbook of gender and psychology*. 2013:288–305.
- Biernat M, Ma JE, Nario-Redmond MR. Standards to suspect and diagnose stereotypical traits. *Social Cognition*. 2008; 26:288–313.
- Blanchard FA, Lilly T, Vaughn LA. Reducing the expression of racial prejudice. *Psychological Science*. 1991; 2:101–105.
- Blodorn A, O'Brien LT, Kordys J. Responding to sex-based discrimination: Gender differences in perceived discrimination and implications for legal decision making. *Group Processes & Intergroup Relations*. 2012; 15:409–424.
- Brant CR, Mynatt CR, Doherty ME. Judgments about sexism: A policy capturing approach. *Sex Roles*. 1999; 41:347–374.
- Czopp AM, Monteith MJ. Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*. 2003; 29:532–544. [PubMed: 15273006]
- Czopp AM, Monteith MJ, Mark AY. Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*. 2006; 90:784–803. [PubMed: 16737373]

- Darley JM, Gross PH. A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*. 1983; 44:20.
- Devine PG, Monteith MJ, Zuwerink JR, Elliot AJ. Prejudice with and without compunction. *Journal of Personality and Social Psychology*. 1991; 60:817–830.
- Dodd EH, Guiliano T, Boutell J, Moran BE. Respected or rejected: Perceptions of women who confront sexist remarks. *Sex Roles*. 2002; 45:7–8.
- Fitzgerald LF, Swan S, Fischer K. Why didn't she just report him? The psychological and legal implications of women's responses to sexual harassment. *Journal of Social Issues*. 1995; 51:117–138.
- Higgins ET. Self-discrepancy: A theory relating self and affect. *Psychological Review*. 1987; 94:319–340. [PubMed: 3615707]
- Hopkins v. Price Waterhouse, 490 U.S. 228 (1985)
- Inman ML, Baron RS. Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology*. 1996; 70:727–739. [PubMed: 8636895]
- Jacobellis v. Ohio, 378 U.S. 184 (1964).
- Judd, CM., McClelland, GH., Ryan, CS. *Data analysis: A model comparison approach*. Routledge: 2011.
- Kaiser CR, Miller CT. Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*. 2001; 27:254–263.
- Klonis SC, Plant EA, Devine PG. Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin*. 2005; 31:1237–1249. [PubMed: 16055643]
- Legault L, Gutsell JN, Inzlicht M. Ironic effects of antiprejudice messages how motivational interventions can reduce (but also increase) prejudice. *Psychological Science*. 2011; 22:1472–1477. [PubMed: 22123778]
- Mallett RK, Wagner DE. The unexpectedly positive consequences of confronting sexism. *Journal of Experimental Social Psychology*. 2011; 47:215–220.
- Monteith MJ. Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*. 1993; 65:469–485.
- Monteith MJ, Deneen NE, Tooman GD. The effect of social norm activation on the expression of opinions concerning gay men and Blacks. *Basic and Applied Social Psychology*. 1996; 18:267–288.
- Muller D, Judd CM, Yzerbyt VY. When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*. 2005; 89:852–863. [PubMed: 16393020]
- Mussweiler T. Comparison processes in social judgment: mechanisms and consequences. *Psychological Review*. 2003; 110:472. [PubMed: 12885111]
- Nisbett RE, Zukier H, Lemley RE. The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*. 1981; 13:248–277.
- Plant EA, Devine PG. Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*. 1998; 75:811–832.
- Plant EA, Devine PG. Responses to other-imposed pro-black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology*. 2001; 36:486–501.
- Plant EA, Devine PG. The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*. 2009; 96:640–652. [PubMed: 19254109]
- Rodin MJ, Price JM, Bryson JB, Sanchez FJ. Asymmetry in prejudice attribution. *Journal of Experimental Social Psychology*. 1990; 26:481–504.
- Rokeach, M. *The nature of human values*. New York: The Free Press; 1973.
- Shapiro JR, Williams AM. The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*. 2012; 66:175–183.
- Smith ER. Differential perceptions of sexism. *Women & Therapy*. 1992; 12:187–200.
- Sommers SR, Norton MI. Lay theories about white racists: What constitutes racism (and what doesn't). *Group Processes & Intergroup Relations*. 2006; 9:117–138.
- Swim JK, Aikin KJ, Hall WS, Hunter BA. Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*. 1995; 68:199–214.

- Swim JK, Mallett R, Russo-Devosa Y, Stangor C. Judgments of sexism: A comparison of the subtlety of sexism measures and sources of variability in judgments of sexism. *Psychology of Women Quarterly*. 2005; 29:406–411.
- Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. *Journal of Statistical Software*. 2014; 59:1–38. [PubMed: 26917999]
- Trenholm S, Todd de Mancillas WR. Student perceptions of sexism. *Quarterly Journal of Speech*. 1978; 64:267–283.
- Trope Y. Identification and inferential processes in dispositional attribution. *Psychological Review*. 1986; 93:239–257.
- Wal-Mart Stores v. Dukes, 564 U.S. (2011)
- Waxman, OB. Avengers: Age of Ultron stars under fire for calling Scarlett Johansson’s character a ‘slut’ and ‘whore’. 2015 Apr 23. Retrieved from <http://time.com/3832841/chris-evans-jeremy-renner-scarlett-johansson-black-widow/>
- Woodzicka JA, LaFrance M. Real versus imagined gender harassment. *Journal of Social Issues*. 2001; 57:15–30.

Appendix A

Pretested Behaviors used in Study 1 and Study 2

The present statements are those used directly in Study 2. In Study 1, statements were changed to reflect self or other Should statements (e.g., “Pays a woman a lower salary than a man” changed to “I should pay a woman a lower salary than a man”).

Mean	SD	
6.36	1.10	Pays a woman a lower salary than a man.
6.27	1.34	Says that a woman deserved to be raped because of what she was wearing.
6.13	1.07	Believes that women shouldn’t hold positions of power because they are too emotional.
6.11	1.13	Thinks a high-powered businesswoman is only there because she is attractive.
6.07	1.23	Thinks women should not be in politics.
6.04	1.06	Favors men over women when assigning lead projects or assignments in the workplace.
5.98	1.27	Thinks choice of college is not as important for women as for men.
5.94	1.15	Assumes a female lab partner would not be as good as a male lab partner in chemistry class.
5.93	1.38	Believes a wife is subordinate to a husband in the family structure.
5.93	1.26	Thinks a woman should be careful not to appear smarter than the man she is dating.
5.72	1.22	Assumes a female is at an interview for a secretary position rather than a business associate position.
6.04	1.20	MEAN EXTREME
5.35	1.43	Is disappointed when a woman joins his softball team.
5.33	1.44	Feels uncomfortable having a female boss
5.31	1.66	Laughs when a friend says to his wife, “make me a sandwich, woman!”
5.31	1.41	Believes keeping track of a child’s activities should be mostly the mother’s task.
5.28	1.66	Calls a woman a “slut” when she has casual sex.
5.07	1.39	Is surprised that an airplane pilot is female.
5.03	1.53	Thinks that a man burping is funny but a woman burping is unacceptable.
4.99	1.52	Says that a woman must be “PMS-ing” if she is in a bad mood.
4.91	1.61	Laughs at a joke based on how foolish women are.

Mean	SD	
4.78	1.55	Teaches his son to fix a car but does not offer to teach his daughter.
4.69	1.48	Assumes the author of a political book or article is male when sex is unspecified.
5.10	1.52	MEAN MODERATE

Appendix B

Exact wording of personal should statements in Study 1. Statements were changed in format to reflect standards for others (e.g., Other people should pay a woman a lower salary than a man.)

Extreme Statements

- I should pay a woman a lower salary than a man.
- I should say that a woman deserved to be raped because of what she was wearing.
- I should think a high-powered business woman is only there because she is attractive.
- I should believe that women shouldn't hold positions of power because they are too emotional.
- I should think women should not be in politics.
- I should favor men over women when assigning lead projects or assignments in the workplace.
- I should think choice of college is not as important for women as for men.
- I should assume a female lab partner would not be as good as a male lab partner in chemistry class.
- I should believe a wife is subordinate to a husband in the family structure.
- I should think a woman should be careful not to appear smarter than the man she is dating.
- I should think a female is at an interview for a secretary position rather than a business associate position.

Moderate Statements

- I should be disappointed when a woman joins my sports team.
- I should feel uncomfortable having a female boss.
- I should laugh when a friend says to his wife, 'make me a sandwich, woman!' I should believe keeping track of a child's activities should be mostly the mother's task.
- I should call a woman a 'slut' when she has casual sex.
- I should be surprised when an airplane pilot is female

I should think a man burping is funny but a woman burping is unacceptable.

I should say that a woman must be ‘PMS-ing’ if she is in a bad mood.

I should offer to teach my son to fix a car but not offer to teach my daughter.

I should assume the author of a political book or article is male when sex is unspecified.

I should laugh at a joke based on how foolish women are.

Appendix C

Presented below are the three, extremely sexist behavior condition transcripts. Extremely sexist statements made by the target TA are bolded, and italicized. In the one behavior condition, participants read only the first sexist statement (starred below) and the second two statements were replaced with pretested, neutral statements. All statements and their pretested means and standard deviations follow the transcripts. Instructions read as follows:

In this task, you will be reading the transcript of a randomly selected TA during a TA meeting and during office hours with a student. You have been selected to evaluate Mike, a TA in the Physics Department. After reading the transcripts, you will then answer a few questions on the computer.

While you are reading, please note instances where the TA did things well, poorly, how you think he or she might have acted differently, and any recommendations you may have for the TA. Your perspective will help us in our goal to provide quality teaching for our undergraduate students by ensuring the continued improvement of our TAs.

Transcript of meeting of TA's

- **Male TA #1:** Hi everyone, so let's get started. We just wanted to bring everyone together to check in on how the semester has been going so far in your discussion sections for intro. Would anyone like to start?
- **Male TA #2:** Sure, I'll go. So I haven't had too many issues in my discussion sections, the students seem involved with the topics and ask pretty good questions. My main concern is students' performance on the last exam. Since the last exam was equation-heavy, I heard that some students struggled with trying to memorize everything and know when to use each equation. Has anyone else heard this from students or have any ideas on how we could better prepare them?
- **Mike:** Yeah, I heard the same complaint from a few students too. It just seems like some students get it and some don't. That's just how it is. I also noticed a pattern in which students are succeeding in class versus those who aren't.
**Having girls in the class seems to slow us down.*
- **Male TA #2:** I think if we just focus on reviewing more material in discussion sections, maybe paying particular attention to the equations and possible difficulties we think the students might have when using them, that would help.

- **Mike:** True. I think if we can all collaborate and maybe make a review sheet for the students with all the equations listed, that might be helpful. I think we should also stress in our sections that it's not about memorizing the equations, but knowing when to use them.
- **Male TA #1:** Right. So I think the review sheet is a good idea. We can also work on integrating more of the assignments into examples that we use in class. I've also tried to get students to ask more questions, so I give everyone a notecard in the beginning of class to write down a question about the material and then about 20 minutes before class ends, I'll go through the cards and read them out loud. As a class, we then try to answer the questions. I think that helps students who are afraid to ask a question in class get their questions answered and also help everyone to work through the material.
- **Mike:** That's a good idea. I think our biggest concern should be making sure all students understand the material and all questions are answered. There are going to be some students who just understand it and some that will need more time, and I think we need to try and identify those students who need more help and address their problems. *Since a lot of the girls are struggling, maybe we should incorporate more examples using topics girls are interested in, like cooking and fashion.*
- **Male TA #1:** So let's plan to schedule one or two extra review sessions, maybe next week, and we can go from there. Also, think about ways to increase your availability to answering students' questions, like using notecards or even increasing your office hours. We'll plan to meet at our regular time next week to check in. Good meeting everybody, and we'll keep in touch with questions.

Transcript of meeting between TA and female student

<knock on the door>

- **TA (Mike):** Come in!
- **Student (Sarah):** Hi, I'm here for office hours. Do you have a couple minutes so we could go over my last exam?
- **Mike:** Sure, take a seat. Let me just pull out your exam, what's the last four digits of your student ID number?
- **Sarah:** 5776.
- **Mike:** <shuffles some papers> Ok, here we go. So, Sarah, you got a B on this exam. What do you want to talk about?
- **Sarah:** Well I was just wondering if we could talk about how I could better prepare and study for the next exam. I think my main problem was knowing what equations to apply to each problem. So like on number 5 on the exam, I wasn't sure what equation to use since there's kind of two different problems going on.

- **Mike:** Sure, so what I've been telling other students is to write out a list of all the equations we've gone over in class, pretty much everything that we've covered, and then make separate columns and write in the column next to the equation what types of problems we've used that equation for. So it might be best to do this after each lecture we have, just so you have a running list and you don't have to write down everything right before the exam. Also make sure you know what the variables stand for in each equation and know the units for each, just in case you can cancel out variables. Does that make sense?
- **Sarah:** Yeah, I think so. For the last exam I just tried memorizing all the equations without really thinking about how to apply them. I just got really overwhelmed with how many equations could've been on the exam.
- **Mike:** Yeah, it's understandable. Also, I think the next exam will be more concerned with conceptual understanding of the material, so you might want to focus more on understanding the reasoning behind the equations and what they mean. In lecture next week, we'll be focusing more on theory, so it might also help to actually read the chapters in the book after lecture, since the book can be a little dense. And I know we did this during introductions the first day of discussion, but what's your major again?
- **Sarah:** Oh, I'm a math major.
- **Mike:** *Oh, that's surprising. Women's math and spatial abilities simply aren't as good as men's.*
- **Sarah:** Ummm, yeah. I guess. So back to the exam, I also have some trouble when there are problems that compare two scenarios that use the same equation. I get a little caught up with what variables I can cancel out and the relationship between everything.
- **Mike:** Yeah, so it's important to understand when you manipulate an equation how it affects all the other variables. I'll use an easier example. So in the velocity equation, velocity equals distance divided by time. If you have a problem that starts out with a given velocity and a given distance, and then it asks if you double the velocity and keep the same distance, what needs to happen with the time. You should know to cut the time in half. Even though this is a simple equation, you get the idea. Just make sure you know the relationships between the variables.
- **Sarah:** Ok, that makes sense. I think I just need some more practice and I should be ok. I think that's all I had for you today, thanks for your help.
- **Mike:** Sure, no problem. Feel free to email me if anything else comes up. See you tomorrow in class!

Pretested statements made during the TA meeting	Mean	SD
Extremely sexist * Having girls in the class seems to slow us down.	6.29	1.27

Pretested statements made during the TA meeting		Mean	SD
	Since a lot of girls are struggling, maybe we should incorporate more examples using topics girls are interested in, like cooking a fashion	6.00	1.50
Moderately sexist	*The girls just seem less engaged by the material.	4.21	1.76
	It just seems that a lot of the girls just don't enjoy the class, so I think we need to try and engage them more.	3.90	1.79
Neutral	We might want to hold an extra review session for any students who feel that they are struggling.	1.43	1.21

*The statement used in the one behavior condition, followed by neutral statements.

Pretested statements made during office hours		Mean	SD
Extremely sexist	Oh, that's surprising. Women's math and spatial abilities simply aren't as good as men's.	6.44	1.17
Moderately sexist	I guess that explains why you're really excelling compared to other girls in this class.	3.52	1.82
Neutral	The math department at this school is really great.	1.16	0.74

Highlights

We offer a framework to explain the source of disputes about what constitutes sexism

People set standards for sexism in themselves and others based on their values

These values predict how much behavioral evidence people require to infer sexism

When others are sexist, people with nonsexist values get angry and try to intervene

To prevent disputes concerning sexist behavior, explicit guidelines are needed

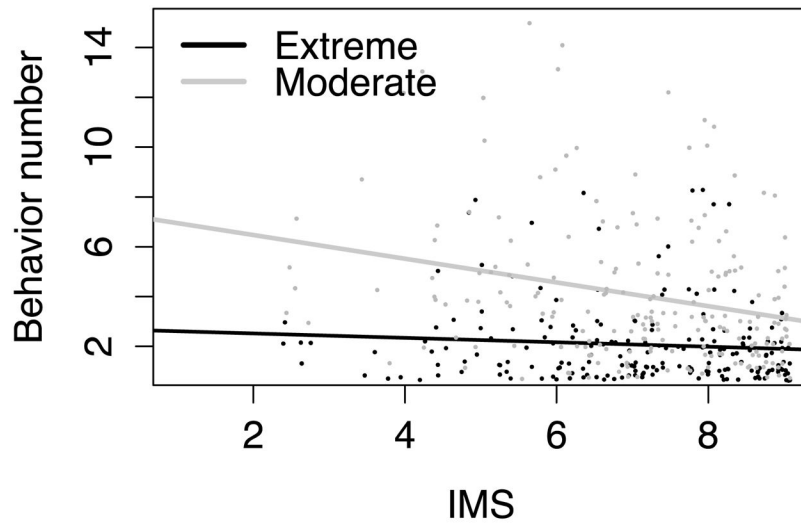


Figure 1. Number of behaviors needed to infer sexism as a function of behavior type and IMS.
Note: Points are jittered in this and all other graphs to avoid over-plotting; IMS = Internal Motivation to Respond Without Sexism.

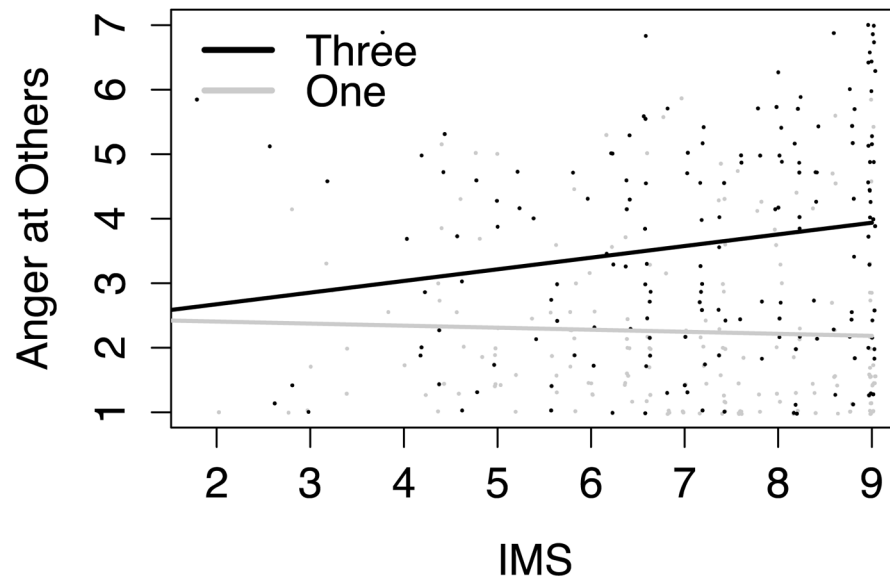
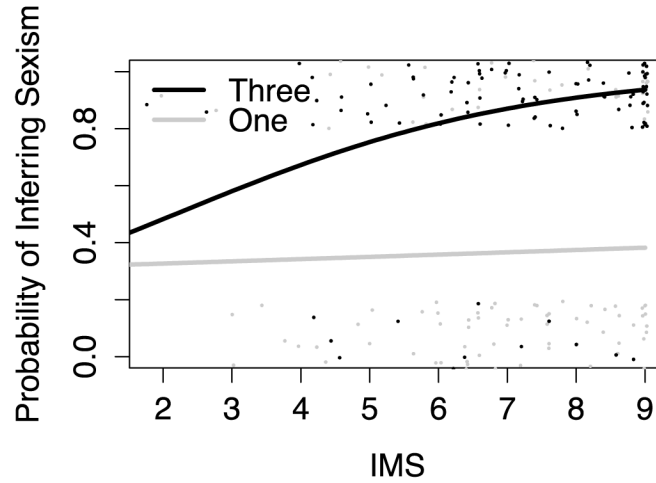


Figure 2.

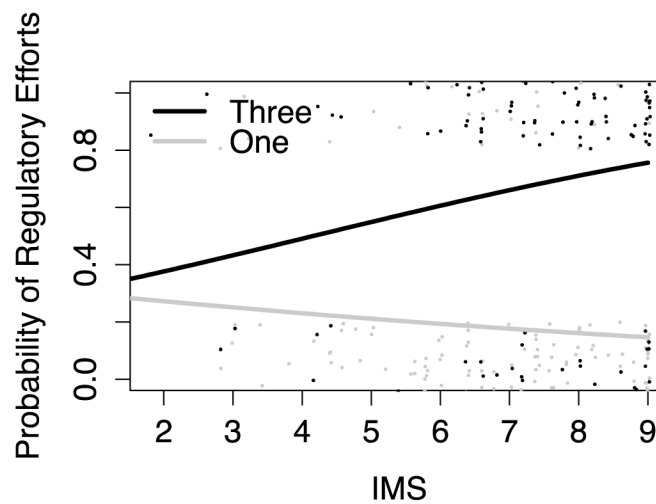
Anger at others as a function of behavior number and IMS.

Note: IMS = Internal Motivation to Respond Without Sexism Scale.

Panel A



Panel B

**Figure 3.**

The probability of inferring sexism (Panel A) or regulatory efforts (Panel B) as a function of behavior number and IMS.

Note: Dependent variables shown include *probability of inferences of sexism* (Panel A) in which participants labeled the TA as sexist or that his behavior was inappropriate and *probability of regulatory efforts* (Panel B) in which participants gave the TA suggestions to reduce sexist or unacceptable behavior. IMS = Internal Motivation to Respond Without Sexism Scale.

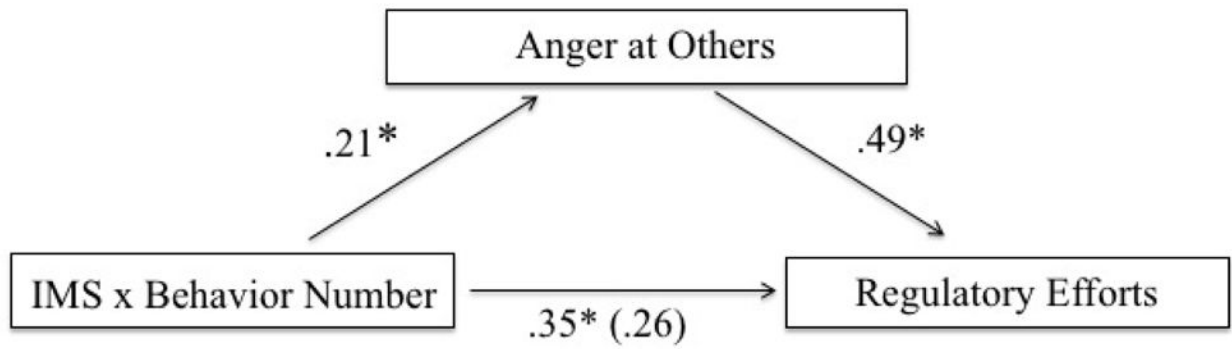


Figure 4.

Standardized regression coefficients for the moderated mediation relationship between IMS \times Behavior Number and regulatory efforts through anger at others.

Note: The standardized regression coefficient between IMS \times behavior number and regulatory efforts, controlling for anger at others, is in parentheses. $*p < .05$.

Table 1

Descriptive Statistics and Correlations for Study 1

	Standards Others	Standards Self	IMS	EMS	MSS	Gender
Standards Others	0.92					
Standards Self	0.66	0.88				
IMS	-0.39**	-0.47**	.89			
EMS	0.18**	0.14*	0.00	.84		
MSS	0.17*	0.20**	-0.32**	0.11	.83	
Gender	0.06	0.14*	-0.20**	0.05	0.29**	
<i>Mean</i>	2.06	1.90	7.52	4.36	2.44	0.23
<i>SD</i>	1.04	0.86	1.22	1.98	0.67	0.42
<i>Skew</i>	1.09	0.87	-0.67	0.25	0.29	1.28

Note. N = 199; Where appropriate, Cronbach's alpha is shown in the diagonal; IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale;

* p < .05,

** p < .01.

Table 2
Comparing Personal Standards for Self and Personal Standards for Others in Model 1 v. Model 2

Panel A (Personal Standards for Self):												
Model 1						Model 2						
	B	Std Er	t	p	R ²	Conf. Int.	B	Std Er	t	p	Conf. Int.	
IMS	-0.30	0.04	-7.47	<.001	***	-0.38 -0.21	-0.27	0.04	-6.46	<.001	***	0.13 -0.35
Behavior Type	-0.50	0.10	-5.15	<.001	***	-0.69 -0.31	-0.53	0.10	-5.47	<.001	***	0.09 -0.72
Gender	-	-	-	-	-	-	0.03	0.12	0.26	0.79	-	0.00 -0.20
MSS	-	-	-	-	-	-	0.10	0.08	1.78	0.08	-	0.00 -0.06
EMS	-	-	-	-	-	-	0.07	0.02	2.72	0.01	**	0.02 0.02
IMS × Behavior Type	0.13	0.08	1.62	0.11	0.01	-0.03 0.29	0.15	0.08	1.78	0.08	-	0.01 -0.02
Behavior Type × Gender	-	-	-	-	-	-	-0.34	0.24	-1.44	0.15	-	0.01 -0.81
Behavior Type × MSS	-	-	-	-	-	-	0.10	0.16	0.65	0.52	-	0.00 -0.21
Behavior Type × EMS	-	-	-	-	-	-	0.04	0.05	0.89	0.37	-	0.00 -0.05

Panel B (Personal Standards for Others):												
Model 1						Model 2						
	B	Std Er	t	p	R ²	Conf. Int.	B	Std Er	t	p	Conf. Int.	
IMS	-0.31	0.05	-5.94	<.001	***	-0.41 -0.21	-0.30	0.06	-5.37	<.001	***	0.10 -0.41
Behavior Type	-0.40	0.13	-3.16	.001	**	-0.65 -0.15	-0.43	0.13	-3.41	<.001	***	0.04 -0.68
Gender	-	-	-	-	-	-	-0.12	0.16	-0.07	0.46	-	0.00 -0.42
MSS	-	-	-	-	-	-	0.11	0.10	1.03	0.31	-	0.00 -0.10
EMS	-	-	-	-	-	-	0.10	0.03	2.98	0.003	**	0.03 0.03
IMS × Behavior Type	0.01	0.10	0.07	0.95	0.00	-0.20 0.21	0.06	0.11	0.52	0.60	-	0.00 -0.16
Behavior Type × Gender	-	-	-	-	-	-	-0.18	0.31	-0.57	0.57	-	0.00 -0.79
Behavior Type × MSS	-	-	-	-	-	-	0.14	0.21	0.67	0.50	-	0.00 -0.27
Behavior Type × EMS	-	-	-	-	-	-	0.00	0.06	0.02	0.98	-	0.00 -0.13

Note. Panel A: Personal Standards for the Self, Model 1 df = 233, Model 2 df = 217; Panel B: Personal Standards for Others, Model 1 df = 233, Model 2 df = 217; IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale; Conf Int. = 95% Confidence intervals. All variables were mean-centered prior to interactions. Behavior type is coded such that moderately sexist behaviors = -.5, extremely sexist behaviors = .5. Gender is coded such that females = -.5, males = .5.

Table 3

Descriptive Statistics and Correlations for Study 2

	Behaviors to Infer Sexism	IMS	EMS	MSS	Gender
Behaviors to Infer					
Sexism					
IMS	-0.24**	0.87			
EMS	0.02	0.10	0.88		
MSS	0.18*	-0.38**	0.03	0.82	
Gender	-0.05	-0.27**	0.18*	0.39**	
<i>Mean</i>	3.07	7.05	4.74	2.46	0.42
<i>SD</i>	1.82	1.57	2.01	0.68	0.50
<i>Skew</i>	1.22	-0.83	0.03	0.05	0.31

Note. N = 199; Where appropriate, Cronbach's alpha is shown in the diagonal; IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale;

* p < .05,

** p < .01.

Table 4
Comparing Number of Behaviors Needed to Make a Sexism Inference in Model 1 v. Model 2

	Model 1					Model 2							
	B	Std Er	t	p	R ²	Conf. Int.	B	Std Er	t	p	R ²	Conf. Int.	
IMS	-0.28	0.08	-3.52	<.001	***	-0.44	-0.12	0.09	-3.19	0.002	**	-0.45	-0.10
Behavior Type	-1.99	0.20	-9.75	<.001	***	-2.39	-1.59	1.55	-2.69	<.001	**	-7.24	-1.11
Gender	-	-	-	-	-	-	-0.69	0.28	-2.46	0.01	*	-1.25	-0.14
MSS	-	-	-	-	-	-	0.43	0.21	2.06	0.04	*	0.02	0.84
EMS	-	-	-	-	-	-	0.07	0.06	1.11	0.27	-	0.01	-0.05
IMS × Behavior Type	0.39	0.13	2.96	0.003	**	0.12	0.65	0.41	2.82	0.01	**	0.12	0.70
Behavior Type × Gender	-	-	-	-	-	-	0.66	0.47	1.42	0.16	-	-1.58	0.26
Behavior Type × MSS	-	-	-	-	-	-	-0.26	0.35	-0.76	0.45	-	0.00	-0.42
Behavior Type × EMS	-	-	-	-	-	-	-0.07	0.11	-0.66	0.51	-	0.00	-0.14

Note. Model 1 df = 197, Model 2 df = 194. IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale; Conf. Int. = 95% Confidence intervals. The *R*² is left blank for Behavior Type as this is a within-subjects variable and *R*, thus, does not apply. All variables were mean-centered prior to interactions. Behavior type is coded such that moderately sexist behaviors = -5, extremely sexist behaviors = .5. Gender is coded such that females = -.5, males = .5.

Table 5

Descriptive Statistics and Correlations for Study 3

Study 3

	Anger at Others	Negative Self-Affect	Positive Affect	Inference	Suggestion	IMS	EMS	MSS	Gender
Anger at Others	0.94								
Negative Self-Affect	0.38**	0.88							
Positive Affect	-0.33**	0.12*	0.90						
Inference	0.40**	-0.08**	-0.28**						
Regulatory	0.38**	-0.09**	-0.30**	0.66**					
IMS	0.10	-0.08	-0.13*	0.14**	0.10	0.88			
EMS	-0.09	0.03**	0.10**	-0.06	-0.13	0.00	0.85		
MSS	-0.14*	0.06	0.15**	-0.18**	-0.16**	-0.35**	0.07	0.81	
Gender	-0.15	0.04	0.09	-0.16**	-0.05	-0.17**	0.03	0.36**	
<i>Mean</i>	2.91	1.65	3.32	0.61	0.41	7.02	4.31	3.29	0.38
<i>SD</i>	1.63	0.93	1.39	0.49	0.49	1.69	1.92	1.05	0.49
<i>Skew</i>	0.62	1.66	0.06	-0.43	0.34	-0.72	0.25	0.10	0.48

Note. N = 349; Where appropriate, Cronbach's alpha is shown in the diagonal; IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale.

* p < .05,

** p < .01.

Table 6

Comparing Anger at Others in Model 1 vs. Model 2

	Model 1						Model 2							
	B	Std Er	t	p	R ²	Conf. Int.	B	Std Er	t	p	R ²	Conf. Int.		
IMS	0.09	0.04	2.11	0.04	*	0.01	0.01	0.18	1.16	0.25	0.00	-0.04	0.15	
Behavior Type	0.82	0.15	5.49	<.001	***	0.06	0.53	1.12	1.30	0.20	0.00	-0.40	1.97	
Behavior Number	1.32	0.15	8.83	<.001	***	0.16	1.03	1.62	3.47	<.001	***	0.02	0.91	3.28
Gender	-	-	-	-	-	-	-	-	-1.87	0.06	0.01	-0.64	0.02	
MSS	-	-	-	-	-	-	-	-	-1.44	0.15	0.00	-0.28	0.04	
EMS	-	-	-	-	-	-	-	-	-1.29	0.20	0.00	-0.13	0.03	
IMS × Behavior Type	0.02	0.09	0.20	0.84		0.00	-0.16	0.19	0.24	0.81	0.00	-0.16	0.21	
IMS × Behavior Number	0.25	0.09	2.82	0.01	**	0.02	0.08	0.43	2.16	0.03	*	0.01	0.02	0.39
Behavior Type × Behavior Number	1.31	0.30	4.38	<.001	***	0.04	0.72	1.90	-0.29	0.77	0.00	-2.72	2.03	
Behavior Type × Gender	-	-	-	-	-	-	-	-	-1.01	0.31	0.00	-1.00	0.32	
Behavior Number × Gender	-	-	-	-	-	-	-	-	-1.66	0.10	0.01	-1.22	0.10	
Behavior Type × MSS	-	-	-	-	-	-	-	-	0.09	0.93	0.00	-0.31	0.34	
Behavior Number × MSS	-	-	-	-	-	-	-	-	-0.73	0.47	0.00	-0.44	0.20	
Behavior Type × EMS	-	-	-	-	-	-	-	-	0.41	0.68	0.00	-0.12	0.19	
Behavior Number × EMS	-	-	-	-	-	-	-	-	-0.56	0.58	0.00	-0.20	0.11	
IMS × Behavior Type × Behavior Number	-0.01	0.18	-0.05	0.96		0.00	-0.36	0.34	0.42	0.68	0.00	-0.29	0.45	
Behavior Type × Behavior Number × Gender	-	-	-	-	-	-	-	-	-0.76	0.45	0.00	-1.83	0.81	
Behavior Type × Behavior Number × MSS	-	-	-	-	-	-	-	-	0.98	0.33	0.00	-0.32	0.97	
Behavior Type × Behavior Number × EMS	-	-	-	-	-	-	-	-	1.19	0.24	0.00	-0.12	0.49	

Note. Model 1 df = 339, Model 2 df = 327. IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale; Conf. Int. = 95% Confidence intervals. All variables were mean-centered prior to interactions. Behavior type is coded such that moderately sexist behaviors = -.5, extremely sexist behaviors = .5. Behavior number is coded such that one behavior = -.5, three behaviors = .5. Gender is coded such that females = -.5, males = .5.

Table 7

Comparing Inferences of Sexism in Model 1 vs. Model 2

	Model 1					Model 2						
	B	Std Er	t	ChiSq	p	Conf. Int.	B	Std Er	t	ChiSq	p	Conf. Int.
IMS	0.28	0.09	3.22	10.86	0.001	***	0.01	0.18	0.21	4.70	0.03	*
Behavior Type	1.29	0.33	3.86	17.53	<.001	***	0.53	1.12	1.39	14.65	<.001	***
Behavior Number	2.67	0.33	8.00	96.95	<.001	***	1.03	1.62	2.90	91.70	<.001	***
Gender	-	-	-	-	-	-	-0.30	0.36	-0.82	0.66	0.42	-0.64
MSS	-	-	-	-	-	-	-0.46	0.18	-2.47	6.42	0.01	*
EMS	-	-	-	-	-	-	-0.09	0.09	-1.01	1.03	0.31	-0.13
IMS × Behavior Type	-0.15	0.17	-0.85	0.72	0.40	-	-0.16	0.19	-0.33	0.11	0.74	-0.16
IMS × Behavior Number	0.40	0.17	2.34	5.55	0.02	*	0.08	0.43	0.45	5.48	0.02	*
Behavior Type × Behavior Number	0.20	0.67	0.30	0.09	0.77	-	0.72	1.90	0.39	0.25	0.62	-2.72
Behavior Type × Gender	-	-	-	-	-	-	0.49	0.72	0.68	0.47	0.49	-1.00
Behavior Number × Gender	-	-	-	-	-	-	0.00	0.72	0.00	0.00	1.00	-1.22
Behavior Type × MSS	-	-	-	-	-	-	0.37	0.37	1.00	1.00	0.32	-0.31
Behavior Number × MSS	-	-	-	-	-	-	-0.46	0.37	-1.35	1.62	0.20	-0.44
Behavior Type × EMS	-	-	-	-	-	-	-0.07	0.18	-0.43	0.18	0.67	-0.12
Behavior Number × EMS	-	-	-	-	-	-	-0.14	0.18	-0.80	0.64	0.42	-0.20
IMS × Behavior Type × Behavior Number	0.23	0.34	0.67	0.45	0.50	-	-0.36	0.34	0.35	0.39	0.36	-0.29
Behavior Type × Behavior Number × Gender	-	-	-	-	-	-	1.18	1.44	0.82	0.68	0.41	-1.83
Behavior Type × Behavior Number × MSS	-	-	-	-	-	-	0.13	0.74	0.18	0.03	0.86	-0.32
Behavior Type × Behavior Number × EMS	-	-	-	-	-	-	-0.39	0.35	-1.11	1.25	0.26	-0.12

Note. Model 1 df = 339, Model 2 df = 327. IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale; Conf. Int. = 95% Confidence intervals. All variables were mean-centered prior to interactions. Behavior type is coded such that moderately sexist behaviors = -.5, extremely sexist behaviors = .5. Behavior number is coded such that one behavior = -.5, three behaviors = .5. Gender is coded such that females = -.5, males = .5.

Table 8

Comparing Regulatory Effort in Model 1 vs. Model 2

	Model 1					Model 2							
	B	Std Er	t	ChiSq	p	Conf. Int	B	Std Er	t	ChiSq	p	Conf. Int.	
IMS	0.10	0.08	1.18	1.42	0.23	-0.06	0.26	0.01	0.14	0.02	0.89	-0.17	0.20
Behavior Type	0.94	0.28	3.41	12.31	<.001	***	0.41	1.49	1.19	3.38	12.98	<.001	***
Behavior Number	2.30	0.28	8.37	85.18	<.001	***	1.78	2.86	2.82	8.03	92.83	<.001	***
Gender	-	-	-	-	-	-	0.36	0.35	1.05	1.09	0.30	-0.32	1.05
MSS	-	-	-	-	-	-	-0.49	0.17	-2.92	8.78	0.003	**	-0.82
EMS	-	-	-	-	-	-	-0.21	0.08	-2.45	6.44	0.01	*	-0.38
IMS × Behavior Type	-0.14	0.16	-0.86	0.75	0.39	-0.47	0.18	-0.10	-0.53	0.28	0.60	-0.46	0.26
IMS × Behavior Number	0.36	0.16	2.18	4.56	0.03	*	0.03	0.67	0.48	6.50	0.01	*	0.11
Behavior Type × Behavior Number	-0.36	0.55	-0.65	0.43	0.51	-1.46	0.71	-0.16	-0.23	0.05	0.82	-1.60	1.20
Behavior Type × Gender	-	-	-	-	-	-	0.23	0.69	0.34	0.12	0.73	-1.11	1.63
Behavior Number × Gender	-	-	-	-	-	-	1.78	0.69	2.57	7.04	0.01	**	0.46
Behavior Type × MSS	-	-	-	-	-	-	0.00	0.33	-0.01	0.00	0.99	-0.66	0.65
Behavior Number × MSS	-	-	-	-	-	-	-0.37	0.33	-1.12	1.26	0.26	-1.04	0.28
Behavior Type × EMS	-	-	-	-	-	-	0.06	0.17	0.37	0.14	0.71	-0.26	0.40
Behavior Number × EMS	-	-	-	-	-	-	-0.14	0.17	-0.83	0.67	0.41	-0.47	0.20
IMS × Behavior Type × Behavior Number	0.02	0.33	0.05	0.00	0.96	-0.62	0.67	0.17	0.46	0.22	0.64	-0.55	0.91
Behavior Type × Behavior Number × Gender	-	-	-	-	-	-	1.68	1.39	1.22	1.48	0.22	-1.04	4.43
Behavior Type × Behavior Number × MSS	-	-	-	-	-	-	0.31	0.67	0.46	0.32	0.64	-1.00	1.62
Behavior Type × Behavior Number × EMS	-	-	-	-	-	-	-0.94	0.34	-2.79	8.49	0.00	**	-1.64

Note. Model 1 df = 339, Model 2 df = 327. IMS = Internal Motivation to Respond Without Sexism Scale; EMS = External Motivation to Respond Without Sexism Scale; MSS = Modern Sexism Scale; Conf. Int. = 95% Confidence intervals. All variables were mean-centered prior to interactions. Behavior type is coded such that moderately sexist behaviors = -.5, extremely sexist behaviors = .5. Behavior number is coded such that one behavior = -.5, three behaviors = .5. Gender is coded such that females = -.5, males = .5.